

ARE THERE INTERNAL REASONS?

PHAN SHIAW HWA JASON

(B. Communication Studies. (Hons.), NTU)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ARTS

DEPARTMENT OF PHILOSOPHY

NATIONAL UNIVERSITY OF SINGAPORE

2008

Acknowledgements

While working on a philosophical thesis is worthwhile in itself, it is far better that this work is greatly assisted by help from various individuals. I am indebted to Kyle Swan for his meticulous comments on several drafts and for his generosity in helping to limit the damage caused by my weakness of will. My thanks also to my anonymous examiners, as well as my friends — Pei-En, Joanna and Clement — for their valuable responses to parts of my thesis. I have also benefited significantly from the fun and support provided by several of my fellow philosophy graduate students. My greatest debt and thanks go to my parents and my God for their love.

Table of Contents

Acknowledgements	1
Table of Contents	2
Summary	4
Introduction	6
Internal and external reasons	6
Part 1: Is Williams's distinction viable?	
1.1. Two requirements of reason	11
1.2. A dilemma for the internalist	15
1.3. Response 1	
1.3.1. <i>Is</i> and <i>ought</i>	18
1.3.2. A reductive account	23
1.4. Response 2	37
1.4. Response 3	
1.4.1. Motivation by desire	40
1.4.2. Finlay's account of means-ends motivation	42
1.4.3. Motivation towards ends and means	46
1.4.4. Ends and means	49
1.4.5. Evaluating Finlay's account	52
1.5. Conclusion	60
Part 2: Can there be external reasons?	
2.1. Williams's argument against external reasons	62
2.2. The <i>Direction-of-fit</i> argument	67

2.3. Is the <i>Direction-of-fit</i> argument sound?	71
2.4. The <i>Desire-out, Desire-in</i> argument	73
2.5. Is the <i>Desire-out, Desire-in</i> argument sound?	78
2.6. The <i>Unimportance</i> argument	86
2.7. Is the <i>Unimportance</i> argument sound?	93
Part 3: Reasons Subjectivism and Objectivism	
3.1. Taking stock	101
3.2. Two views of reasons	103
3.3. Skepticism about Motivational arguments	108
3.4. Content arguments	114
Bibliography	118

Summary

Bernard Williams thinks that a necessary condition of reasons for action is that those reasons are related to the agent's subjective motivational set¹ by a sound deliberative route. All reasons for action which satisfy this condition are, in his terms, *internal* reasons. Williams goes on to argue that there are no external reasons, i.e. reasons which are independent of the agent's subjective motivational set. As such, internalism is a reasons-tracking thesis: it is the claim that reasons for action are tracked by the agent's sound deliberative limits, as determined by his subjective motivational set.

My essay has three parts. In Part 1, I argue that Williams's distinction between internalism and externalism is of limited use because internalism is clearly false. I explain that internalists face a dilemma: internalism entails highly-counterintuitive outcomes which threatens its plausibility, such as making it merely a *contingent* truth that an agent who acts out of false beliefs is thereby acting against his best reasons. Yet, the only way for internalism to avoid such outcomes is for some form of externalism to be true. But this renders internalism false, since it denies the truth of externalism. I consider three possible ways for the internalist to dissolve that dilemma but find them wanting.

In Part 2, I consider Williams's argument against the conceptual soundness of external reasons. The force of this argument lies in its charge that externalism faces a motivational problem: it cannot explain how we *can* be motivated by reasons, since motivation requires desires and externalism needs us to acquire desires the objects of which are unrelated to those of our actual desires. I argue that this charge is without adequate justification because (i) we have no reason to think that desires of the rational agent cannot be responsive to his beliefs or that (ii) the agent's motivating reasons must

¹ This is made up of the agent's desires, broadly construed. I shall explain this in detail later in the essay.

involve his antecedent desires. I then consider Stephen Finlay's interpretation of Williams's argument, which centers on showing how externalism faces a normative problem: it cannot account for why we *should* be motivated by reasons, since those reasons do not relate to our intrinsic cares and concerns. However, as I argue, externalism faces a normative problem only if normative reasons must have the relation he claims. Finlay has not provided compelling support for that claim and thus, it remains possible that we should be motivated by reasons which are unrelated to our intrinsic cares and concerns. We have no reason to think externalism faces a normative problem. I then conclude that Williams's argument has not shown reasons externalism to be conceptually unsound.

If my aforementioned arguments are successful, then we should accept that either reasons externalism is conceptually sound, or that there are no reasons for action. The issue therefore concerns whether there are reasons for action. However, this is far from the only crucial issue with regards to reasons for action, for there is an important division *within* reasons externalism. This division is engendered by the role of desires in explaining reasons, however it is a division *within* externalism. Even aside from concerns about the motivational ability and normativity of reasons, I explain that there remain many points of dispute between Subjectivists and Objectivists. The primary aim of my essay is to argue that it is more fruitful to focus on those disputes instead of motivational and normativity issues. To this end, it is helpful to shift our attention away from the internalism/externalism debate because that distinction centers on those very issues where there is, or should be, agreement.

Introduction

Bernard Williams thinks that a necessary condition of reasons for action is that those reasons are related to the agent's subjective motivational set by a sound deliberative route. All reasons for action which satisfy this condition are, in his terms, *internal* reasons. Williams goes on to argue that there are no external reasons, i.e. reasons which are independent of the agent's subjective motivational set. As such, internalism is a reasons-tracking thesis: it is the claim that reasons for action are tracked by the agent's sound deliberative limits, as determined by his subjective motivational set.

My essay has three parts. Firstly, I argue that Bernard Williams's distinction between internalism and externalism about reasons for action is of limited use because internalism, as he defines it, is clearly false. As such, arguments against the conceptual possibility of externalism, if sound, leave us with nihilism about reasons for action. Secondly, I examine variants of one such argument, based on motivational and queerness considerations, and conclude that they are unsuccessful. Thirdly, I explain how the internalism/externalism distinction can be replaced to capture a significant debate—an intramural debate among externalists. I argue that this debate centers on the content of external reason statements, rather than motivational and queerness issues.

Internal and external reasons

What is the relation between reason and desire? Bernard Williams distinguishes between two opposing views on that relation, specifically with regard to reasons for action. Williams thinks the difference between the two views lies in their claims concerning the truth conditions of a reason statement of the form "A has reason to Φ ",

where A refers to an agent and Φ an action. One view claims that the truth of a reason statement is contingent on the agent's motivational set, while the other denies that claim.

Williams takes the agent's motivational set to include "such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent".² He takes Reason Internalism (hereafter, RI) as the view that

(RI) A has a reason to Φ only if there is a sound deliberative route from A's subjective motivational set to A's Φ -ing.³

Reason Externalism (RE) is the claim that

(RE) Whether A has a reason to Φ is not contingent on whether there is a sound deliberative route from A's subjective motivational set to A's Φ -ing.

The internalist claims that the reasons which the agent has is not conditioned only on his *actual* motivational set, but also on his *hypothetical* motivational set, insofar as the latter is connected to the former by some sound deliberative route.⁴ Given the use of this

² "Internal and External Reasons," in *Moral Luck*, 105. Hereafter, I shall refer to it as *IER*. For convenience, in this essay I will take 'desires' to refer to the elements in the agent's subjective motivational set. Thus the term 'desire' should be understood broadly in this essay to refer to any of the range of motivational items which Williams counts as part of the motivational set.

³ "Postscript: Some Further Notes on Internal and External Reasons," *Varieties of Practical Reasoning*, 91. Hereafter, I shall refer to it as "Postscript".

⁴ Thus Williams writes that "the deliberative process can add new actions for which there are internal reasons, just as it can also add new internal reasons for given actions. The deliberative process can also subtract elements from S [the agent's subjective motivational set]... We should not, then, think of S as statically given. The processes of deliberation can have all sorts of effect on S, and this is a fact which a theory of internal reasons should be very happy to accommodate. (*IER*, 104-5)

limited hypothetical motivational set, Williams writes: “Indeed, the stance towards the agent that is implied by the internalist account can be usefully compared with that of an imaginative and informed advisor, who takes seriously the formula ‘If I were you...’”⁵

Williams considers such deliberation to be multi-faceted, ranging from means-ends reasoning to imagination. This allows for considerable changes to his actual motivational set, including the addition and subtraction of his ends. For instance, imagination may lead him to see in a clearer light what is involved in Φ -ing and as a result, he may acquire a desire to Φ . In that same way, deliberation makes possible considerable divergence between the agent’s actual and hypothetical motivational set: it may be true that if the agent properly exercises his imagination, he would come to desire to Φ , even if the antecedent does not ever obtain.

To illustrate his distinction, Williams applies it to the case of Owen Wingrave. Owen’s father claims there is a reason for Owen to join the army: family tradition. However, all of Owen’s desires lead in the opposite direction; “he hates everything about military life and what it means”.⁶ Suppose that there is no sound deliberative route from Owen’s motivational set to the desire to join the army. The internalist claims this falsifies the claim by Owen’s father that there is a reason for Owen to join the army. Family tradition is not a consideration which can secure a sound deliberative connection to Owen’s motivational set. However, externalists would not thereby accept that that reason statement is false. In the externalist’s view, it may *still* be the case that Owen has a reason to join the army.

⁵ “Postscript,” 92. Alternatively, we can understand Williams to be claiming that the agent’s reasons are conditioned on his actual motivational set, but taking into account its counterfactual properties.

⁶ *IER*, 106.

Williams goes on to argue that there are only internal reasons. Before I turn to that argument, I shall assess the viability of his distinction between internal and external reasons. As David Sobel puts it, Williams is making a claim about what *tracks* reasons for action, rather than offering a theory of reasons (i.e. a reasons-*making* thesis).⁷ It does not follow from (RI) that the agent has any reasons for actions: (RI) offers only the necessary condition for reasons and it is therefore compatible with nihilism about reasons at both the conceptual and substantive levels. It can be that the sufficient conditions for reasons are not met at either or both levels. However, we can assess the plausibility of Williams's reasons-tracking thesis by examining how well it copes with reason-claims which we consider highly plausible. This shall be my focus in Part 1. In Part 2, I shall consider Williams's argument for (RI).

In this essay, I shall be primarily considering Williams's explanation and defense of (RI). However, as noted by Stephen Finlay and Mark Schroeder in their Stanford Encyclopedia review article "Reasons for Action: Internal vs. External", there are many versions of reasons internalism—(RI) is but one version. They note that as a general reasons-tracking thesis, reasons internalism is the claim that "a consideration is a reason for an agent only if some motivational fact about that agent obtains".⁸ Since there can be many versions of reasons internalism, there can likewise be many versions of its denial, i.e. reasons externalism. Thus Finlay and Schroeder writes: "Philosophers generally describe their views as 'externalist' if they reject any thesis they consider to involve an interesting and controversial dependence of reasons on facts about motivation."⁹ Versions of reasons internalism take the following logical form: 'A has a reason to Φ only if C',

⁷ "Desires vs. The Reduction Basis of Desires," in *PEA Soup*.

⁸ "Reasons for Action: Internal vs. External," in *Stanford Encyclopedia of Philosophy*.

⁹ *Ibid.*

where C states the necessary condition(s) for the truth of reason statements. I think my arguments relate to all non-trivial variants of reasons internalism. I shall explain later in this essay how reasons internalism can be trivially true.¹⁰ But for all non-trivial variants, my arguments would apply to them through replacing C with the relevant terms.

¹⁰ See p. 76-8 of this essay.

Part 1: Is Williams's distinction viable?

1.1. Two requirements of reason

(RI) consists of two parts: (1) the statement of *what* A has reason to do, and (2) the necessary condition for the truth of (1). If (RI) is true, then for *whatever* we may take to be Φ , the truth of (1) is contingent on (2). Williams takes (2) to be desire-based¹¹ because he thinks that reason statements are true only when they can motivate the agent, but reason cannot motivate him except when conjoined with some already existing element in his motivational set. However, (RI) yields some highly counter-intuitive claims.

Consider Williams's petrol/gin case: the agent desires gin and believes that the bottle near him contains gin. He therefore drinks from that bottle. However, unbeknownst to him, the bottle contains petrol, not gin. The bottle which contains gin is in the next room. Let us term him the Mistaken Agent. Williams thinks this agent has acted against his best reasons in drinking from the bottle, even though his action was rational relative to his beliefs. This seems right. The agent has no reason to drink from the bottle near him; what he has reason to do is to drink from the bottle in the next room. (RI) may seem able to account for this since, in the present context, the only *sound* deliberative route from his motivational set is to his drinking from bottle in the next room. However, what is implied by that sound deliberative route? The agent desires to drink gin and the means to his drinking gin is to drink from the bottle in the next room. But why does the agent have a reason to drink gin *just because* he desires to drink gin? It may be said that doing

¹¹ For the purposes of this essay, I am using the term 'desire' broadly such that it refers to any element of the agent's motivational set.

so is the proper means to his end, and that is why he has a reason to do so. However, if we give that reply, then we seem to be committing to certain claims which conflict with (RI).

The first issue that arises is why the agent has any reason to pursue the intentional object of his desire, i.e. his drinking gin. Before delving into that, there are certain important qualifications to be made. Williams claims the agent has no reason to pursue the object of his desire if that desire is the outcome of false beliefs.¹² False beliefs can figure here in two ways. Firstly, the agent may come to desire something which does not exist. In that case, he has no reason to pursue the object of his desire. Secondly, he may come to desire something because he believes that would satisfy some desire of his. However, that belief may be false. In that case, he has no reason to pursue the object of a desire which arose (in part) out of a false means-ends belief. Apart from claiming that false beliefs do not provide the agent with reasons to act, Williams also claims true beliefs can provide the agent with reason to act even if he is not *actually* having those beliefs. For instance, Williams speaks of the role of imagination where the agent can come to have new desires upon a fuller exercise of his imagination regarding what it would be like if some state of affairs obtained.¹³ In view of these qualifications, we can therefore understand Williams to claim that the agent has reason to pursue only the objects of his relevantly-informed desires, where relevant information means having all true beliefs which would alter the composition of his motivational set (i.e. by leading to the formation or loss of his desires).

¹² Thus he writes: "A member of *S*, *D*, will not give *A* a reason for Φ -ing if either the existence of *D* is dependent on false belief, or *A*'s belief in the relevance of Φ -ing to the satisfaction of *D* is false." Ibid, 103.

¹³ See *ibid*, 104-5. Williams adds that imagination could also have the opposite effect: the agent could also lose desires he has if he "come[s] to have some more concrete sense of what would be involved, and lose[s] his desire for it, just as, positively, the imagination can create new possibilities and new desires." Ibid, 105.

But why does the agent have reason to pursue such objects? Why are the objects of relevantly-informed desires reason-giving while those of misinformed desires are not? How about the objects of some *other* hypothetical desires (e.g. what I would desire if I were born 100 years earlier) or the desires of *others* (e.g. what my President desires)? Williams's answer cannot be that we cannot be motivated by those desires because (i) he would then be inconsistent since we are often not relevantly-informed and, thus, likewise often cannot be motivated by our relevantly-informed desires;¹⁴ (ii) even if we cannot be motivated by those desires, it does not follow that just because we can be motivated by certain desires, we therefore have a reason to be motivated by them.

There is a further problem. Suppose it is true that the agent has a reason to pursue the objects of his relevantly-informed desires. However, to pursue those objects, it is an empirical fact that he sometimes has to pursue objects which are merely instrumental to them. For instance, the agent may desire to stay dry in the rain, but it may be an empirical fact that he has to get under some shelter to do so. It seems possible that an agent desires to stay dry and believes that he would do so by getting under shelter, yet remains unmotivated to get under shelter (assuming there are no countervailing considerations). Let us term him the Unmotivated Agent. It certainly seems that, all else being equal, the agent fails to respond to what he has reason to do: he has reason to get under shelter. But how do we account for that reason?

¹⁴ It may be said that we can be motivated by our relevantly-informed desires once we have the relevant beliefs. But likewise, we can be motivated by the desires we would have if we were in some other hypothetical situation *if we were in that hypothetical situation*. In both instances, the fact about our motivation is conditional and it seems plausible that in some cases, we cannot actually have the relevant beliefs, just as we cannot actually go back 100 years. As for the desires of others, we could similarly construct a relevant conditional, such as "I can be motivated by the desires of my President if I constructed such that I would be motivated by the desires of my President."

It may seem that if he has reason to pursue the object of his relevantly-informed desires, then he therefore also has reason to pursue whatever is instrumental to those pursuits. But why think so? Even if the agent has a reason to pursue some end, there may be many ways the pursuit can take place, some ways being less efficient than others. It is not clear that all else being equal, the agent would of necessity take the *most* efficient way, or even a way with *any* efficiency. I could rightly believe that I have a reason to get to London and know all the ways to get to London but remain without any normative guidance as to *how* to get there. Suppose I rightly believe that I would get there only if I take a plane: how does the fact that I have a reason to get to London entail that I have a reason to take a plane? It may be replied that my reason to take a plane is ‘contained within’ my reason to get to London, given that I would get there only by this means. However, there is no content concerning means in my reason to get to London and it is therefore mysterious how the reason to pursue the means come to be contained within the reason to pursue the end. Yet it seems compelling that I have a reason to take the most efficient means to my relevant end (henceforth, the Instrumental Requirement), although that reason does not seem identical or reducible to my reason to pursue the objects of my relevantly-informed desires (henceforth, the End Requirement). Thus,

Instrumental Requirement: A has a reason to take the most efficient means to her relevant end.

End Requirement: A has a reason to pursue the object of his relevantly-informed end.

The Instrumental Requirement seems to depend on the End Requirement because it seems that the agent has no reason to engage in any pursuit, let alone the most efficient, unless he first has reason to pursue some relevant end. However, one may take the Instrumental Requirement as simply a requirement to be consistent in one's attitudes, *whatever* those may be. For instance, even if we disregard the End Requirement for now, it is not obvious that I can therefore have any configuration of attitudes: some configurations remain unreasonable insofar as the attitudes contradict one another.¹⁵ I shall not argue for a position here but content myself with defending the claim that the Instrumental Requirement is in either case distinct from the End Requirement. Thus, the internalist has to account for both.¹⁶

1.2. A dilemma for the internalist

Because Williams defends (RI), he has to abide by its constraints. Let us denote the claim 'The agent has reason to pursue the object of his relevantly-informed desires' as R_1 . On (RI), R_1 is true only because (2) is true—there is a sound deliberative route from A's subjective motivational set to A's pursuing the object of his relevantly-informed desires. Surely this is a bullet which is difficult to bite for the internalist because its consequence is that it is not always true that the Mistaken Agent is acting against his best reasons. After all, the agent may not *actually* desire to pursue the object of his

¹⁵ This point will be revisited on p.27-30 in addressing the claim that certain desires are in themselves reason-giving.

¹⁶ Jay Wallace thinks that the End and Instrumental Requirements can be combined in a holistic model that is used to "determine which course of action would optimally advance the agent's complete set of ends." This approach seeks the way to maximize the agent's total subjective expected utility. But it remains for the internalist to account for the Maximising Requirement. (See his "Practical Reason" in *Stanford Encyclopedia of Philosophy*.)

relevantly-informed desires. We can express this challenge as part of a dilemma facing Williams. He claims

(RI) A has a reason to Φ only if there is a sound deliberative route from A's subjective motivational set to A's Φ -ing.

However, (RI) admits of two readings. We can interpret it as

(RI*) A has a reason to Φ only if [there is a sound deliberative route from A's subjective motivational set to A's Φ -ing].

or

(RI**) A has a reason to [Φ only if there is a sound deliberative route from A's subjective motivational set to A's Φ -ing].

(RI**) is a reason statement of the form 'A has a reason to Ψ ', where Ψ denotes ' Φ only if there is a sound deliberative route from A's subjective motivational set to A's Φ -ing'. But the question then arises: is the truth of (RI**) independent of the agent's motivational set? If it is not, then it is not clear how it helps avoid the counter-intuitive outcomes since it seems possible that the relevant desire is lacking in some agent's motivational set. If it is, then (RI**) is an external reason statement. If it is true, then an external reason statement is true.

However, Williams's exposition of (RI) does not allow for it to be taken as (RI**), understood in this sense. This is because the truth of (RI**) would then require that the agent acquires a new desire, i.e. the desire to [Φ only if there is a sound deliberative route from A's subjective motivational set to A's Φ -ing]. This is because for (RI**) to be a true reason statement, it must be that the agent can act for that reason. But if he is motivated to act for that reason in virtue of an antecedent desire, then he is acting for an internal reason. Thus if the agent can act for (RI**), understood as an external reason, then he can come to have the desire to [Φ only if there is a sound deliberative route from A's subjective motivational set to A's Φ -ing] as a result of believing (RI**). But Williams's complaint against externalism is precisely that this is impossible. In his view, internalism escapes this problem as it needs only motivation which derives from elements within the agent's actual motivational set, or elements connected to that set by a sound deliberative route. Thus, it would be glaringly inconsistent for Williams to take internal reasons as reasons which obtain independently of the agent's motivational set, even if they concern promoting the satisfaction of elements within it.¹⁷

It seems therefore that Williams takes (RI) as (RI*). However, (RI*) delivers implausible outcomes since it claims that *all* reasons are contingent on the agent's motivational set. If the agent's motivational set does not include the desire to pursue the object of his relevantly-informed desires (or some desire which is linked to that desire by a sound deliberative route), he would have no reason to do so. If the Mistaken Agent has

¹⁷ Another reason for taking Williams as defending (RI*) is that (RI**) is not a claim about what tracks reasons—it is a claim about the sufficient condition for the truth of a reason statement. Only (RI*) is a reasons-tracking thesis. Yet Williams seems to explicitly state that he is only offering a reasons-tracking thesis (although he thinks internalism is also a reasons-making thesis). Thus, he writes of internalism: “Whether this is also a sufficient condition of A's having a reason to Φ is a question which I have left aside; the essence of the internalist position is that it is a necessary condition.” (“Postscript,” 91)

such a motivational set, we cannot claim he is acting against his best reasons. As such, Williams faces a problem because for (RI) to avoid implausibility, he has to understand it as (RI**) rather than (RI*). But (RI*) and (RI**) are inconsistent. If (RI**) is true, then (RI*) is false: the latter is the claim that the truth of all reasons statements are contingent on the agent's motivational set, while the former is taken as a reason statement whose truth is not subject to such contingency. Thus, (RI) is incoherent if it requires accepting both (RI*) and (RI**). Yet without (RI**), (RI) is implausible as it entails highly-counter-intuitive outcomes as explained earlier. This is the dilemma facing Williams and the internalists. I shall now consider three ways that the internalist may deny that (RI*) has counterintuitive consequences.

1.3. Response 1

1.3.1. *Is* and *ought*

The internalist may argue that what *makes* reason statements true need not be itself *a reason statement*. In this way, the internalist may claim that the agent has a reason to drink gin so long as his drinking gin relates to some relevantly-informed desire which he has, without being committed to the claim that the agent has an external reason to pursue his relevantly-informed desires. It is the empirical fact that Φ -ing is related to his relevantly-informed desires that gives him a reason to Φ , not the normative fact that 'he has a reason to pursue his relevantly-informed desires'. If this is right, then (RI) can be construed as (RI*) without yielding counterintuitive results. That said, what does not seem contingent on A's motivational set is the relation between that reason statement and

A's means-end complex. However that relation, as I shall now argue, is tenuous: it is either unjustifiable or implicitly reliant on some external reason statement.

Let us assume

- (1) I desire to drink gin.
- (2) I would drink gin if I drink from the bottle.

The internalist may claim that given (1) and (2), it follows that

- (3) I have a reason to drink from the bottle.

But it is far from clear why that it so. Both (1) and (2) state empirical facts: the former describes a psychological fact while the latter describes a causal relation. Yet, (3) is a normative fact. Just how are normative facts derived from non-normative facts? It may seem obvious that (3) follows from (1) and (2) but that is because (1) and (2) are of normative significance. However, the fact that they are normatively significant is not dependent on the shape of my motivational set. This may not be clear because it is easy to confuse desire-involving reasons with desire-dependent reasons. The latter obtain only when certain desires obtain. On the other hand, the former may have desires as part of their content; for instance, the claim that we have reason to pursue the objects of our relevantly-informed desires. Yet, even when desires are implicated in its content, the truth of that reason statement can be independent of desires: if that statement is true, it is true whatever my desires may be. Such a reason statement, if true, would account for why

(1) and (2) are normatively significant—if that reason statement is true, then the truth of (1) and (2) entails that some normative statement is true of me. Externalists can deny that we have reason to pursue such desires, or any desires in themselves. Such externalists usually hold that reasons are given by the objects desired or the state resulting from the satisfaction of the desire (such as pleasure). But even those externalists would hold that (3) follows not from (1) and (2) alone, but only from their normative significance, given some other normative truth.

The internalist has no such recourse, since his appeal to further normative truths merely defers the issue to why that further truth obtains. Here, Hume seems right when he claims:

“In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary ways of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when all of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. This change is imperceptible; but is however, of the last consequence. For as this ought, or ought not, that expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given; for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it.”¹⁸

¹⁸ *A Treatise of Human Nature*, Bk III, Pt I, Sect I.

In (1), the relation claimed to obtain between the agent and that of his desire is one of ownership, while in (2), it is one of cause and effect. In (3), the claimed relation between the agent and the action is prescriptive, such that if the agent (acknowledges and) failed to act in the relevant way, he would be aptly cast into some negative light. This relation is strikingly different from the earlier two because neither (1) nor (2) is prescriptive nor places him under some form of evaluation. (They are at most of relevance in a classificatory sense, placing the agent into wholly descriptive categories such as ‘someone who desires gin’. Nothing follows from such classification that indicates evaluation.) Presumably the internalist claims that (3) follows from (1) and (2). But if nothing like the relation claimed in (3) is contained in the premises from which it is derived, how can the reasoning be valid? Just as we have reason to suspect necessary truths which are wholly derived from contingent premises, and temporal truths wholly derived from non-temporal premises, I think we have *prima facie* reason to be skeptical of normative truths which are wholly derived from non-normative premises. In all cases, the nature of the claim in the conclusion is wholly different from the nature of the claims in the premises.

One may plug the gap by explaining that A would Ψ by Φ -ing. That is true, but how do we infer from that that A therefore should Ψ ? Φ -ing would serve to satisfy my desire to Ψ , but why should I satisfy my desire to Ψ ? Even if it is true that whenever I desire to Ψ , I would desire to Φ if I believe that I would Ψ by Φ -ing, it is no clearer how it is therefore true that I should Φ . Perhaps given (1) and (2), we can derive (3) through

- (A) A would desire to Ψ if A is relevantly-informed and rational relative to his beliefs.

However, if the claim is that ‘if (A), then (3)’, then we have merely relocated the problem to the derivation of (3) from (A): how do we derive what A has reason to do from the empirical fact of what his relevantly-informed self desires? As it stands, I think it is invalid to infer (3) from (1) and (2). Whatever intuitive appeal that such reasoning has is due to some implicit premise which we may consider obviously true. For instance,

- (B) I have reason to pursue the object of my relevantly-informed desire and to take the most efficient means necessary for that pursuit.

If we take (1) and (2) to describe the object of my relevantly-informed desire and the most efficient means necessary respectively, then we can validly deduce (3) from these three premises. Some externalists reject (B) as they hold that the nature of the object desired, rather than the nature of the desire, provides reasons for action. They can claim

- (C1) I have a reason to pursue pleasure.
(C2) Drinking gin gives me pleasure.

When (C1) and (C2) are taken as premises, along with (2) and the Instrumental Requirement, then (3) can be validly deduced.

In both cases where the reasoning is valid, there is at least one normative claim among the premises, along with empirical claims which are thereby of normative significance. On the other hand, the internalist claims that some empirical facts are of normative significance without there being some normative claim which accounts for that. It is therefore hard to see how a normative conclusion validly follows.

1.3.2. A reductive account

Mark Schroeder notes how it seems natural to account for facts about the particular actions we have reason to perform by appealing to ultimate, desire-independent normative facts. He considers such versions to follow what he terms the Standard Model of normative explanations.¹⁹ Schroeder writes: “The explanation that there is a reason for *X* to do *A* because of *P* follows the Standard Model just in case it works because there is (1) some further action *b* such that there is a reason for *X* to do *b* and (2) not just because of *P* and (3) *P* explains why doing *A* is a way for *X* to do *b*.”²⁰ The Standard Model Theory is the view that all explanations of reasons for action must follow the Standard Model.

However, Schroeder rejects the Standard Model Theory and defends an alternative view of accounting for the agent’s reasons. In doing so, he takes (RI) as (RI*),

¹⁹ In writing of a student, Rachel, who is given an assignment to write about whatever she happens to think about each morning, Schroeder notes: “It is more natural to think it is no accident that on Monday when Rachel finds herself thinking about feather pillows, she acquires a reason to write about feather pillows, given that she already has a reason to write about what she is thinking about. After all, the fact that she is thinking about feather pillows explains why writing about feather pillows is a way to write about what she is thinking about, and that is something she has reason to do... For the facts about what she is thinking about alter her situation by altering which actions are *ways* or *means* for her to do something she already has a reason to do...” (*Slaves of the Passions*, 42-3) Schroeder notes that a similar line of reasoning is taken in explaining a case where a person, Ronnie, desires to dance and knows that there will be dancing at a certain party. He comes to have a reason to go to the party because he has a reason to pursue his desires. This manner of explaining a person’s reasons follows the Standard Model Theory which Schroeder thinks “is an appealing and natural view to take”. (Ibid, 50)

²⁰ Ibid, 43.

rather than as (RI**). He proposes a different response which, if successful, would help the internalist evade the charge of invalid reasoning without following the Standard Model. Recall

- (1) I desire to drink gin.
- (2) I would drink gin if I drink from the bottle.

In Schroeder's view, it is not invalid to conclude that

- (3) I have a reason to drink from the bottle.

This is because (3) can be analysed as

- (3*) There is gin in the bottle.

In his view, normative facts can be reduced to non-normative facts. He thinks such a reduction is successful when (and only when) the reductive account “gets the right results about all the things that are true about reasons”.²¹ Schroeder offers the following understanding of reason:

²¹ Ibid, 80.

Reason: “For R to be a reason for X to do A is for there to be some p such that X has a desire whose object is p , and the truth of R is part of what explains why X ’s doing A promotes p .”²²

Applying Reason, we can see how (3*) follows from (1) and (2): the fact that there is gin in the bottle is clearly part of what explains why my drinking from the bottle promotes my gin-drinking. But can (3) be metaphysically reduced to (3*)?

I cannot respond to all of Schroeder’s points in this essay but I am skeptical that there can be a successful reduction. In support of my skepticism, I shall assess what Schroeder considers a challenge to his view and argue that it is indeed a compelling objection, to which his response is unsatisfactory. He takes one important challenge to his internalist account to be that it allows for too many reasons, even those which we intuitively think we do not have. Schroeder terms this the Too Many Reasons objection. As he explains:

“The Too Many Reasons objection is simply that no matter what desires are and what it takes for an action to promote one, this thesis [i.e. Reason] will have intuitive counterexamples.”²³

Consider two cases:

²² Ibid, 59.

²³ Ibid, 88.

Page-counting: I desire to count, for its own sake, the total number of pages in a given book ninety-nine times. Given Reason, there is a reason for me to locate that book.

Future-disregarding: I desire to disregard, for its own sake, my future wellbeing. Given Reason, there is a reason for me to conduct my life so as to disregard my future wellbeing.

Before we reach conclusions about both cases, let us be clear about the relevant considerations. It is important to see that what gives me reason in both cases, according to Reason, is not what I would *derive* out of satisfying those desires. Even supposing that I would obtain pleasure from *page-counting* and *future-disregarding*, it is not the pleasure I would obtain which gives me reason to perform the relevant actions. To make the issue clearer, we can assume that I would derive no pleasure from my satisfaction of either desire, but I shall consider it sufficient to stipulate that I desire to *page-count* and *disregard my future wellbeing* for their own sakes. It is also crucial to note that when it is claimed that there is a reason for me to perform some act, that is meant as a *pro tanto* reason—one that can be of very little weight and easily overridden.

Let us first consider *Page-counting*. I think the first question we should ask is “Do I have any reason to satisfy my desire to count, for its own sake, the total number of pages in a given book ninety-nine times?” Some tempting answers may be:

- (a) Yes, if you enjoy it.

(b) Yes, if you find it useful.

But it should be clear that (a) and (b) do not relate to my desire to count, *for its own sake*, the total number of pages in a given book ninety-nine times. Thus, if we answer in the positive, we must say

(c) Yes, simply because you desire to do so.

I think (c) is a highly-counterintuitive reply. It is important to note that (c) differs from (a), because (a) justifies an activity by one's enjoyment of it, which seems to be *pro tanto* desirable (i.e. a *pro tanto* reason for desiring the activity). On the other hand, (c) refers only to the fact that one desires it and no appeal is made to the desirability of the object of one's desire. It is the claim that an activity is *pro tanto* desirable simply because one desires it. Typically, we cite desires as justifications for action because of considerations of pleasure, or the nature of the intentional object of the desire, or even how it conduces to my creation of individuality. It does not seem that the desire *itself* gives me a *pro tanto* reason to seek its satisfaction. But if I have no reason to satisfy my desire for *page-counting*, then it may be unclear how I can *still* have a reason to pursue what would promote *page-counting*—not even a reason of slight weight.

One may object here by noting how we would subject to rational criticism someone who desires to page-count and yet seeks to do so by *hiding* the book. We may say of him that he failed to behave reasonably; or, loosely speaking, he should²⁴ have

²⁴ In the sense that he has some reason to.

looked for the book instead of hiding it.²⁵ An examiner of this essay makes this point with an illuminating example:

“Consider the military strategy of Nazi Germany in World War II. We might consider an ineffective bit of German military strategy and say that the Germans acted irrationally, or failed to do what they had reason to do. Of course, the failure of German military strategy was a great thing for the world, and once reasons coming from outside their subjective motivational set enter the picture, it will seem that they had more reason to court defeat than victory, as their victory would have destroyed so much that is objectively valuable in the world. But we still do not say, for example, that Hitler has good reason not to crush the British army at Dunderk because by letting them escape he contributed to his defeat and the happiness of humanity. We say that he failed to do what he had reason to. This seems to be a case where we relativise reasons to the agent’s subjective motivational set.”

Although I consider the conclusion above plausible, it does not seem enough in this context that reasons display a relativity to the agent’s subjective motivational set: the *right* sort of relativity is needed.

In case of the Hitler, I agree that he has a reason to crush British army at Dunderk, given his aim of winning the war. However, I contend that this is due entirely to his being subject to a normative requirement, specifically, the Instrumental Requirement. As I have explained earlier, this requires the agent to take the most efficient means to his ends,

²⁵ I owe this important suggestion to an anonymous examiner of my essay.

whatever those may be.²⁶ It is simply the requirement to be consistent in one's attitudes. It seems to me that it is in virtue of the Instrumental Requirement that Hitler has a reason to crush the British at Dunderk, not *simply* because he has a desire to win the war. However, Schroeder's aim is just to avoid appeal to such ultimate, desire-independent normative facts. Otherwise, he would simply be falling back into what he terms as the Standard Model of normative explanations, and that is just what his reductive account is meant to prevent.

Could it be that Hitler has such a reason even apart from the Instrumental Requirement, and simply in virtue of his desire? Let us examine systematically if this is so. Hitler desires to win the war and he would do so only if he crushes the British army at Dunderk. To find out if he has a reason to crush the British *simply* in virtue of his desire, let us first peel away the reasons given by extraneous considerations. Take away the reasons, if there are, given by the pleasure or good (instrumental or otherwise) he would obtain by winning the war. Take away the reason, if there is, given by the Instrumental Requirement which requires him to take the most efficient means to achieve his end of winning the war. Is there any reason left over?

If desire *itself* gives one a *pro tanto* reason to seek its satisfaction, we should expect that there remains a reason for Hitler to crush the British. But once we have removed the reason which desire *derivatively* provides via the Instrumental Requirement, it seems there is no reason *additionally* given by desire in itself. Even the aforementioned examiner speaks only in terms of the Instrumental Requirement when he concludes:

²⁶ See p. 14-5.

“I fear that the ends of many artists and lovers might not come to any great objective value, but there is still a broadly applicable sort of positive normative evaluation that attaches to resourceful pursuit of their goals.”

However, what Schroeder needs is for desire itself to provide the agent with a reason. Thus, his reductive account cannot rely on the Instrumental Requirement. But it appears that there is no reason given by desire in itself.

I shall adopt a similar approach to considering *Future-disregarding*; that is, to first ask whether I have a reason to satisfy my desire to disregard my future wellbeing. The immediate intuitive response seems to be that we have compelling reason not to do so. However, this response needs to be analysed because it is not immediately clear that it is not relating to what we have *most* reason to do. It may seem that we have weighty prudential reasons not to disregard my future wellbeing but do we nonetheless have other reasons—albeit overridden ones—to disregard my future wellbeing? As before, we must rule out irrelevant considerations. In this case, it seems that I not only have *no* reason to satisfy my desire to disregard my future wellbeing but ample reason *not* to satisfy it—what is being desired is intuitively undesirable in itself. Once we put aside potentially misleading considerations, it appears false that we always have *pro tanto* reason to satisfy our desires.

Schroeder anticipates this challenge and seeks to defuse it by showing that our intuitions concerning when we have no reason to act are systematically inaccurate. He terms such intuitions our ‘negative existential intuitions’. To this end, he puts forth a

number of cases for our consideration. I shall consider only two of them, but my assessment of the other cases is identical in principle to my assessment of these cases.

Case 1: “[Y]ou see Tom Grabit come out of the library, pull a book from beneath his shirt, cackle gleefully, and scurry off. Intuitively, you have a reason to believe that Tom just stole a book from the library. But in a revised version of the case, Tom has an identical twin, Tim, from whom you cannot visually distinguish him. If you’re aware of this, then it turns out that you don’t have any reason to believe that Tom stole a book after all. Right?”²⁷

Case 2: “Suppose that there are three routes that you can take home from work—Timely, Compromise, and Scenic. If you take Timely, you will get home by six. If you take Compromise, you have even odds of getting home by six, but the scenery will be much more pleasant. And if you take Scenic, you will not get home by six, and the scenery will be exactly as pleasant as if you take Compromise. Suppose further that you want to get home by six, but are indifferent to scenery. Is there any reason for you to take Compromise? Intuitively, no—it doesn’t get you anything that you want that you can’t get better by taking Timely.”²⁸

²⁷ Ibid, 93. While this case concerns theoretical rather than practical reasons, it is relevant to my discussion because it exemplifies same the line of reasoning taken by Schroeder in the other cases. He thinks negative existential reason claims in both theoretical and practical contexts are both afflicted with the same problem. My focus is whether he is right that there is such a problem.

²⁸ Ibid, 93.

In both cases, Schroeder claims that there is in fact *a reason* for each of the relevant actions. It is just that both times, the weight of each reason is clearly overridden by another reason. Yet it seems we intuitively judge that there is *no reason* for each of the two actions. Schroeder thinks this striking pattern has a compelling natural explanation which involves two steps.

Firstly, he thinks we can observe that the two reasons in question have relatively little weight. But we consider what reasons for action we have for practical purposes—we want to know what we have *most* reason to do. If so, then reasons of relatively little weight are irrelevant and thus we are likely to overlook them, especially when they abound. Building on that, Schroeder claims that in conversation we seek to be as informative as is feasible in what we say to others. But since most actions have at least some poor reason in favour of them, we would be hardly informative in making true but merely existential claims about reasons. As such, there is a standing presumption that the speaker has some weighty reason to convey when he simply claims that I have a reason to act. Otherwise, he would make clear what the poor reason that he is claiming is—or else he would be hardly informative. Thus Schroeder concludes:

“And so we have our two-step pragmatic explanation of why we often find it unintuitive or inappropriate to say that there is a reason for someone to do something even when, in fact, there is reason for her to do it. It yields two predictions. If I tell you that there is a reason for you to do something that there are only poor reasons to do, what I say will sound wrong. But—first prediction—it will sound less wrong if I tell you what the reason is, because doing so will

remove the pragmatic reinforcement of the standing presumption that I have only relatively good reasons in mind. And second, if I then tell you that I don't think it is a particularly weighty reason, I should be able to cancel the presumption, and so the unintuitiveness of what I say should go down a second time."²⁹

He thinks these predictions are true of all similar cases, including Cases 1 and 2. Take Case 2 for example. Initially we think there is no reason for you to take Compromise and it seems wrong if one claims that there is a reason for you to do so. However, the first prediction of the pragmatic explanation seems right: it sounds less wrong to think you have a reason to take Compromise once it is explained to us that it does get you what you want, albeit in a less efficient manner. The second prediction also seems right: once it is clarified that the reason to take Compromise is not a weighty reason given its inferiority to Timely, we begin to accept how there is indeed such a reason. At each of the two steps, we become clearer about the context of the conversation and note how it is not concerned solely with good reasons. Accordingly, we revise our claims about what there is reason to do. It therefore seems our initial judgments were systematically flawed.³⁰

²⁹ Ibid, 95.

³⁰ Similarly in Case 1, I have a reason to believe that Tom Grabit stole a book based on the context of his behaviour. However, I have as much reason to believe that his identical twin stole the book. Nonetheless, even if I do not have *most* reason to believe that Tom stole the book, I still have *some* reason.

Thus Schroeder claims that the Too Many Reasons objection is without force because it relies on negative existential intuitions about reasons, which in his view “really are misleading, and basic pragmatic principles suffice to predict and explain this.”³¹ Schroeder seems right insofar as we do appear to revise our judgments about what there is reason to do, from making *negative* existential claims about reason to *merely* existential ones.

However, he seems too swift to dismiss the force of arguments based on negative existential claims about reasons. For all he has shown (if he is successful) is that in *certain* situations, we are *prone* to make false claims about what we have no reason to do. This seems only to justify our skepticism about the claims we make in such situations *until* we subject them to careful analysis. Even if he is right in the above cases, he has shown us how we are able to systematically correct our judgments in recognition of how we are likely to be misled—we need to remove the constraints of conversational pragmatics by making clear that we are not only concerned with particularly weighty reasons. Moreover, these constraints seem only to arise given that we are considering our reasons for *practical* purposes. Couldn’t we be free of those constraints when our purpose is *theoretical*, aimed at being sensitive to the nuances of what we have reason to do?

Aside from the aforementioned concern, it seems Schroeder has not succeeded in showing us that our negative existential judgments about reasons are truly misled, such that in each case we have to revise them. For us to *revise* our judgment about what we

³¹ He adds: “I conclude from this that intuitively overgenerating existential predictions about reasons is not in and of itself a bad problem for the Humean Theory. So long as the Humean can get the right results about what those reasons would be, if there were such reasons, and allow that they are not particularly weighty, she can explain our negative existential intuitions about reasons just as well as views that deny that there are any such reasons. Provided the Humean Theory satisfies these two constraints, therefore, we can’t rely on negative existential intuitions about reasons to decide between our theories.” Ibid, 96.

have reason to do, the meaning of the phrase “what we have reason to do” must stay constant, or else we are simply making a judgment about a *different* issue. But in the cases which Schroeder mentions, it seems what we are revising is not our judgment about what we have reason to do, but what we understand by the phrase “what we have reason to do”. In making our initial judgments, we overlook how the notion of “what we have reason to do” in fact means what we have *a/some* reason to do. It may be that conversational pragmatics misleads us to understand it as what we have *most* reason to do, and therefore make a negative judgment in each case. However, once we clarify the issues at hand, it seems that we have no reason to doubt such judgments of ours. This goes to show that we are not revising our initial judgments about *what* we have reason to do, but our judgments about what the phrase ‘what we have reason to do’ *means*. This does not suffice for Schroeder’s claim that our negative existential claims about reasons are systematically flawed, since it is not truly those claims which are flawed, but our recognition that we were asked to make those claims. This error in recognition seems easily remediable by clarifying the meaning of the relevant terms.

The issues of conversational pragmatics and the limits of practical deliberation do not seem relevant to philosophical reflection on carefully specified cases—or at least these two factors do not render our judgments irreparably flawed. After all, we are still confident about various actions that we have no reason to perform in the aforementioned cases: in Case 1, I have no reason therefore to believe that Tom has a sister; in Case 2, I have no reason therefore to paint my nails. Moreover, Schroeder himself presumably thinks he can make such claims, or else it is hard to see how his account of practical

reason is to work—surely he would accept that I have no reason to perform the action which does *not* promote the satisfaction of *any* of my desires.

To sum up, Schroeder's response to the Too Many Reasons objection is problematic because he fails to show that our negative existential claims about reasons are flawed in principle. At most, he has shown that we are prone, in certain situations, to mistake the meaning of the term 'reason' for 'weighty or most reason'. But this seems remediable and in any case, does not amount to our being mistaken in judgments about whether we have reasons, where the term 'reason' is properly understood. In considering *Page-counting* and *Future-disregarding*, I have explicitly considered whether there are reasons of *any* weight to act in the relevant ways and argued that none are to be found. The considerations Schroeder adduces do not undercut my judgments in those cases, *if* my arguments are sound. In that regard, I have argued that we do not always have *pro tanto* reason to satisfy our desires because desires in themselves do not provide us with reason to satisfy them. We often have *pro tanto* reason to satisfy our desires because of considerations like the pleasure we would derive from doing so. But such considerations are distinct from the fact that we *have* those desires. In some cases such as *Page-counting* and *Future-disregarding*, those considerations are lacking and it seems implausible to think we nonetheless have *pro tanto* reason to satisfy those desires simply because we have them.

I therefore conclude that Schroeder's response to the Too Many Reasons objection is inadequate. Although I have not considered how he responds to other difficulties involved in his attempted reduction of normativity, this objection gives me some reason to doubt that Schroeder can succeed. Unless there is a compelling reductive

account, I see no reason to think the internalist can avoid accepting some external reason statement while claiming that the agent has some reason for action.

1.4. Response 2

The second response accepts that (RI*) entails the claims described above but denies that those claims are truly counterintuitive, for it remains that there are rational requirements. It is only when we confuse *rationality* with *reasons* that we consider (RI*) problematic. Consider the Instrumental Requirement. It may be claimed that we need not consider the Unmotivated Agent to be acting against his best reasons. Instead we should simply consider him irrational. After all, the key feature of the case is that the agent fails to respond appropriately to what he believes he ought to do. The Instrumental Requirement is a requirement of rationality and further argument is needed to claim it is what the agent has reason to abide by.

A clear application of this distinction would show the problems with Christine Korsgaard's attack on internalism. Korsgaard focuses on what she considers motivational skepticism—the view that reason cannot motivate except with the help of the agent's already existing desires, arguing that even the internalist has to accept that reason can motivate.³² As such, there *can* be external reasons since the agent can be motivated by reason without the help of his already existing desires. In her argument, Korsgaard appeals to the possibility of (a version of) the Unmotivated Agent. She claims, plausibly, that we would consider that agent irrational. However, in Korsgaard's view, the

³² She distinguishes between two forms of skepticism about practical reason, motivational and content skepticism, arguing that the former has no force apart from the latter. While motivational skeptics hold that reason cannot motivate, content skeptics think it can but does not. After all, it is conceptually possible that reason statements have content with motivational relevance while substantively, no reason statement has such content.

admission of the possibility of such irrationality is fatal to motivational skepticism. His irrationality lies in his failing to be motivated by the rational consideration that drinking what is in the bottle is the means to his end. If we admit that the person is thereby irrational, then we are admitting that he is governed by certain norms of rationality. Korsgaard thinks that since Williams holds that rational norms apply to someone only if he is capable of being motivated by them, it follows that Williams must accept that he could have been motivated by that consideration. Thus she writes that

“it looks as if a theory of means/end rationality ought to allow for at least one form of true irrationality, namely, failure to be motivated by the consideration that the action is the means to your end. Even the skeptic about practical reason admits that human beings can be motivated by the consideration that a given action is a means to a desired end.”³³ (underlining mine)

David Sobel thinks Korsgaard’s argument is irrelevant to her stated aim. This is because he observes that the concern of Hume and Humeans like Williams is with the agent’s reasons for action, not his rationality or decision procedure. Thus Sobel notes:

“Korsgaard claims that Hume and the Humeans lack the resources to label any action irrational... [Her] central mistake in these two papers is that she fails to appreciate that Hume's account is an account of an agent's true reasons for action and not an account of rationality. Thus Korsgaard's criticisms of Hume and his followers are misdirected and leave unaddressed the Humean theory because she

³³ “Skepticism about Practical Reason,” in *Moral Discourse and Practice*, 378.

misunderstands what such theories are accounts of. She is correct when she writes of the sorts of cases in which Hume thinks we act contrary to reason that ‘these are cases of [factual] mistakes; the actions that result are not, strictly speaking, irrational’. Yet such actions certainly do count as contrary to reason for the Humean.”

If Korsgaard means to attack their view of reasons based on her case of irrationality, she must make clear the link between reasons and rationality such that it is clear how the Humean theory of reasons is threatened by its inability to account for cases of irrationality. Sobel thinks—and I agree—that she fails to do so.

However, it is not obvious that Korsgaard’s argument is *irreparably* irrelevant. This is because although theories of reasons and rationality can be clearly distinguished, it does not seem that they are without any important relation. Derek Parfit puts forth two questions:

- (A) What do we have most reason to want, and do?
- (B) What is it most rational for us to want, and do?

Parfit claims that if we assume that we have all the relevant beliefs, these two questions have the same answer with respect to an agent. This seems true. In the case of the Mistaken Agent, there is a divergence between the agent’s rationality and reasons because he does not have the relevant beliefs. It seems plausible to hold that rationality consists in responding appropriately to reasons, or at least acknowledged reasons: it

seems difficult to make any sense of the notion of rationality without appealing to the concept of reason. This does not mean that the rational agent cannot be mistaken about what he has reason to do, but that the agent cannot be understood as acting rationally unless we invoke reasons—what he has reason to do, or, what he thinks he has reason to do. When we come to the Unmotivated Agent, it is hard to explain why he is irrational (as he clearly is) without understanding him to be inappropriately responding to either his actual or acknowledged reason (such as the Instrumental Requirement).

The attempt to understand the Instrumental Requirement as a rational requirement but not *also* as a requirement of reason is suspect because it takes rationality to be independent of reason. I doubt such independence makes sense: even if rationality concerns our decision-making procedure, it seems the only way that what we are procedurally-rational to do differs from what we have reason to do is when we fail to have the relevant information.³⁴ Otherwise, it seems that the answers to (A) and (B) are the same. Therefore, I conclude that the rationality/reason distinction is of no help to the internalist in defusing the counterintuitive outcomes of (RI*).

1.4. Response 3

1.4.1. Motivation by desire

The internalist can deny the Instrumental Requirement, thereby denying that agents have reasons to take the means to their relevant ends. To justify that denial, one can adopt Stephen Finlay's argument that there is no place for reason to figure in how

³⁴ Thus I think we can usefully distinguish between objective and subjective rationality: the latter concerns what it is rational to do given the subject's beliefs, while the former assumes the subject has all the relevant beliefs. On the other hand, it does not make sense to distinguish between what we are objectively rational to do and what we have reason to do: what could that difference be?

agents are motivated towards their (believed) means, because motivation from one's ends to the means is fully explained by the nature of desire. As he writes:

“On this account of instrumental motivation, the motivation is provided by the desire. In following instrumental reasons and seeking means, desire is not obeying any authoritative normative principles but simply acting according to its essential nature. There is no place in this account for rational norms or a motivational susceptibility to any.”³⁵

Finlay aims to defend what he terms the ‘Humean Motivation-by-Desire Principle’ (hereafter, the HMD principle), which he explains as the thesis that “all motivation of action—causal influence towards activity exerted by minds—requires and is produced by conative psychological states like desire and aversion.”³⁶ He wants to defend that against rationalist objections centered on the claim that motivation from ends to means sometimes involves us acting as “rational agents motivated by recognition of reasons.”³⁷ This claim is not just about the agent's rationality as it also involves the agent recognising and acting for genuine reasons. If Finlay shows that the nature of desire fully accounts for means-end motivation, then it is unclear how reason has any role in such motivation.

It is noteworthy that the HMD principle is compatible with instrumental reason, so long as reason has an indirect role in motivating action. It can be the case that my recognition of a reason to Φ (all else being equal) leads to my desiring to Φ , which then

³⁵ “Motivation to the Means,” in *Moral Psychology Today*, 27.

³⁶ *Ibid*, 1.

³⁷ *Ibid*, 3.

accounts for my being motivated to Φ . It seems to me the claim that the nature of desire fully accounts for means-end motivation is distinct from the claim that the HMD principle is true. The former denies reason any role in means-end motivation while the latter does not. I am not sure which claim Finlay means to defend and if he recognizes that they are distinct.³⁸ As I see it, his argument (if sound) shows the former, but also shows the latter, given that the first entails the second. Here I shall only be concerned with the claim that reason has no role in means-end motivation. I shall not be contesting the HMD principle in its entirety.

1.4.2. Finlay's account of means-ends motivation

Accounts of means-end motivation, as Finlay notes, explain how “a desire for one thing motivate[s] action of doing something else.”³⁹ For instance, suppose I desire to play golf and believe that I would do so only if I broke a vase. How does my desire to play golf motivate my breaking a vase? Finlay approaches that question by first examining what desires are. He notes that all desires are desires for ends. However, there are different types of ends. As such, Finlay distinguishes between two kinds of desire: desire-to and desire-that. The former involves ends which are forms of activities or behaviour, while the latter involves states of affairs. Consider a father desiring that his children are happy. Compare that to the same person desiring to play golf. There is certainly a difference in the two ends which he desires: while *playing golf* is something he can *do*, it

³⁸ There are other indications that he does not recognize their difference. For instance, he writes on p. 6 that “if normative reasons are true propositions, they could be the contents of the relevant beliefs, while the source of their normativity (status as reasons) is their connection with desires. The motivation may then derive entirely from those desires. Motivation by reasons would then be identical with a form of desire-belief causation”. Even if (A) the source of normativity lies in relation to desires, it does not follow that (B) motivation is fully explained by those desires. Finlay hereby seems to be either making the unjustified claim that if (A) then (B), or wrongly equating (A) with (B).

³⁹ Ibid, 1.

is clear that *his children's being happy* is not something he can do. For a concept of desire to be plausible, it must capture these two kinds of desires.

Finlay considers but rejects the construal of desires as motivation towards *doing* the end. This is because if a desire is the motivation towards an end, which is in turn understood as motivation towards doing the end, then we cannot desire various states of affairs since these ends are not a form of activity. But surely that is absurd. To accommodate such desires, Finlay thinks we should take desires to be motivation towards making the end obtain. Even if we cannot be motivated towards *doing* states of affairs, we can be motivated towards *making them obtain*. Thus he writes:

“The only viable solution for a motivation view of desiring is that motivation towards an end *E* is motivation towards making *E* so... Desiring an end does not necessarily involve motivation towards doing the end, but rather towards facilitating the end.”⁴⁰

Finlay thinks this does not remove the distinction between desiring that *E* and desiring to make it so that *E*. He notes: “These desires may involve the same activity, but are responsive to representation of different goals.”⁴¹ However, I think it is not obvious that ‘desires that’ cannot be understood as ‘desires to’. He considers Davidson’s claim that ‘desires that’ are really misdescribed ‘desires to’, and applies it to the case of my desire that my children are happy. However, Finlay rejects Davidson’s claim, writing: “It is implausible that (following Davidson) this is really a misdescribed desire to make my

⁴⁰ Ibid, 13.

⁴¹ Ibid, 13, footnote 28.

children happy, because it may just as well be satisfied if my children make themselves happy.”⁴²

Finlay seems to have a point here. However, that can be accommodated by understanding my desire that my children are happy as my desire to know that my children are happy. In this case, it does not matter if I am not the one who makes it the case that they are happy. We could then understand this desire as the motivation towards knowing that my children are happy. This construal would make it clearer how the motivation towards this end is distinct from the motivation towards the means, e.g. the motivation towards painting my nails. The former is a motivation to know, which is a mental activity; the latter is a motivation to paint, which is a physical activity. *Prima facie*, the two activities are distinct. One may respond that my desire that my children are happy can be satisfied even if I do not know that they are happy—they can be happy without my knowledge. This concern can be addressed by qualifying my proposed analysis: we can understand my desire that my children are happy as my desire to know (if I am fully-informed) that my children are happy. Alternatively, we can deny that my desire to make it so that E is satisfied only if I made it the case that E. That would be true only if I desire to make it so that *I make it so that E*. The former desire does not require my role in bringing E about, unlike the latter. However, I shall not argue for this claim in this essay and shall assume, for the sake of argument, that Finlay is right on this point. As such, when I *desire to* make it the case that E, I shall take it that this desire would be satisfied only if *I* made it the case that E. However, if I *desire that* E, I would be satisfied if E obtains even if I did not make it so.

⁴² Ibid, 13.

Finlay proceeds to consider what a means-end belief is and finds it crucially linked to desiring *E*. He writes:

“To believe some action *M* to be a ‘means’ to an end *E* is simply to believe that *M* is an action (or part thereof) of making *E* so, or a facilitating condition for *E*... The motivation involved in desiring *E*, we have seen, must be motivation towards making *E* so. In other words, motivation towards the end *just is* motivation towards the means to that end.”⁴³

Finlay therefore has his account of means-end motivation: since motivation towards the means is constitutive of motivation towards the end, the agent who is motivated towards *E* is motivated towards *M* as a matter of conceptual truth. The nature of desire therefore fully explains means-end motivation. Finlay’s account can be understood in this way:

- (I) Necessarily, *A* desires *E* only if *A* is motivated towards⁴⁴ making *E* so.
- (II) Necessarily, *M* is a means to *E* if and only if *M* is an action of making *E* so.
- (III) Necessarily, if *A* is motivated towards making *E* so and believes that *M* is an action of making *E* so, then *A* is motivated towards making *M* so.
- (IV) Necessarily, if *A* is motivated towards making *E* so and believes that *M* is a means to *E*, then *A* is motivated towards making *M* so. [II, III]

⁴³ Ibid, 14.

⁴⁴ Finlay uses the term “motivated towards” rather than “motivated to” because he thinks the former involves taking ‘mental steps’ (like searching for means) towards bringing about the end but not actually bringing the end about, unlike in the case of the latter. Nothing that I shall argue for hangs on this distinction. See Ibid, 20.

- (V) Therefore, necessarily, if *A* desires *E* and believes that *M* is a means to *E*, then *A* is motivated towards making *M* so. [I, IV]

1.4.3. Motivation towards ends and means

(III) seems implausible. *A*'s being motivated towards making *E* so is a distinct mental state from *A*'s believing that *M* is an action of making *E* so, and it therefore seems conceptually possible that they exist independently of each other within an agent. It is possible that *A*'s motivational and belief systems are divorced from each other such that whatever goes on in one system has no effect on the other. If that is possible, then it is possible that *A* is motivated towards making *E* so and believes that *M* constitutes making *E* so (i.e. *M* is a means to *E*), *and yet* *A* is not motivated towards making *M* so. There seems no conceptual necessity that the antecedent in (III) entails the consequent. Therefore, (III) seems false. But if so, (IV) clearly does not follow. It seems the nature of desire therefore does not fully account for means-end motivation and Finlay has not shut out the rationalist's challenge.

He may respond that (III) must be true if (I) and (II) are true: if *A* believes that *M* is a means to *E* and yet is not motivated towards making *M* so, then it simply shows that *A* does not desire *E*. As Finlay writes:

“Granting that desiring can motivate action only in conjunction with belief, it may still remain a problem how desiring an end (e.g. drinking) can motivate an action that is not identical with that end (e.g. inserting a coin in a slot). But I have already argued that these scruples are misconceived. Desiring involves motivation

towards making its end the case, and that just is to be motivated towards believed means.”⁴⁵

However, it is false that (I) and (II) jointly entail (III), and therefore wrong to think that if (III) is false, then either (I) or (II) is false. While (II) states that means are actions of bringing about the ends, it does not claim that the concept of bringing about some end is identical to the concept of bringing about the means to that end. What is claimed in (II) is just that the bringing about of the means to some end is metaphysically constitutive of the bringing about of that end. In like manner, the concept of ‘the only balloon in Jane’s house’ is distinct from the concept of ‘the whole of Peter’s property’. However, suppose that the only balloon in Jane’s house is in fact owned by Peter. Thus, it is partly constitutive of the whole of his property. It seems clear that I can be motivated to paint all of Peter’s property red, without being motivated to paint the only balloon in Jane’s house red—I may lack the belief that that balloon is Peter’s property.

Similarly, the agent’s motivation towards making E so is a distinct motivational state from his motivation towards the believed means. These two states are distinct since it is possible that the former obtains while the latter does not. The agent must first *believe* that M is the means to E before he is motivated towards M, but he need not have that same belief in order to be motivated towards E. Thus, if they are distinct states, it is not clear why it is conceptually necessary that if the agent is motivated towards making some end, then he is motivated towards the believed means. These two motivational states are mediated by his belief about the means and it seems possible that the connections between his means-belief and motivational states can break down or not exist at all.

⁴⁵ Ibid, 16.

To secure (III), Finlay needs a stronger connection between the agent's motivation towards making *E* so and his motivation towards the believed means—(II) must be replaced by

(II*) Necessarily, *M* is a means to *E* if and only if making *M* so is conceptually identical to making *E* so.

If making *M* so is conceptually identical to making *E* so, it is conceptually impossible that the agent is motivated to make *E* so and yet not motivated to make *E* so: he cannot *both* be motivated and not motivated to make *E* so. But (II*) is implausible. Even if the agent is motivated towards making *E* so, surely he must first *believe* that *M* is the means to *E* before he comes to be motivated towards making *M* so. Otherwise, it is difficult to see why he would be motivated towards making *M* so in virtue of his desire for *E*—absent his belief about their relevant connection, *M* would seem irrelevant to *E* in the agent's view.

The rationalist can accept that there is no conceptual connection between the agent's motivation towards his end and his motivation towards his means, while maintaining that there is a rational connection: the agent who desires *E* and believes that *M* is the means to *E* would be motivated towards *M*, insofar as he is rational. Since the nature of desire does not secure, by itself, motivation from ends to means, it remains open for the rationalist to mount a challenge. If I am right, we should reject (III) even if we accept (I) and (II). The rejection of (III) renders Finlay's argument unsound since (IV) and (V) require (III).

I have argued that even *if* it is true that making M so is constitutive of making E so, we should not think that motivation towards the believed means is constitutive of motivation towards the end. But is it true that making M so is constitutive of making E so? In other words, is (II) true? Here is a reason for doubting (II): M is not constitutive of E, since M is stated as only a means to E. If so, it appears implausible that making M so is constitutive of making E so—if M is distinct from E, why is making M so constitutive of making E so? Allow me to elaborate by means of some examples.

1.4.4. Ends and means

Suppose I desire that my children are happy. I also desire to play golf. *Playing golf* is something I can do, while *my children's being happy* is not something I can do. However, what I *can* do in that regard is to *make it the case that my children are happy*. While we are fairly clear about what falls under the description 'playing golf', it is less clear what falls under 'making it the case that my children are happy'. Even if there may be controversial cases concerning what counts as 'playing golf', we are fairly sure that nail-painting does not. How about what falls under the description 'making it the case that my children are happy'?

One may think that this surely depends on several factors, such as the personality/interests of my children and the circumstances they are in. But suppose we know what those factors are. It may therefore seem that we can then be confident that some actions do not count as 'making it the case that my children are happy' (and vice versa). Suppose my children would not be happy if I paint my nails, unlike how they would be happy if I buy them chocolates. But suppose I can nonetheless make them

happy *indirectly* by painting my nails: if I paint my nails, I would make my wife happy and she would therefore buy my children chocolates, which would make them happy. In this case, one may claim that my painting my nails should not count as an action which is properly described as ‘making it the case that my children are happy’, just as my breaking a vase does not count as ‘playing golf’ even if it would lead to my playing golf.

However, I think this view is wrong. This is because in the case above, *both* my buying them chocolates and my painting my nails should not count as actions which are described as ‘making it the case that my children are happy’. Both actions should be understood as *what* makes it the case that my children are happy. With respect to some state of affairs S, it is crucial to note that *making it the case that S* is distinct from *making it the case what makes it the case that S*. I can make it the case that my finger moves at time T, by moving my finger at T. On the other hand, I can attach my finger to a machine at T-1 which would move my finger at T. I make it the case that S in the first case, and make it the case what makes it the case that S in the second. When I desire that my children are happy, I do not desire for its own sake what would make them happy. I desire to give them chocolates only because I believe that would make them happy, and not for the sake of giving them chocolates in itself.

One would then ask what counts as ‘making it the case that my children are happy’. Unlike the case of playing golf, I do not think we are capable of any action which falls under that description.⁴⁶ Thus, we have to speak of purely logically possible actions. This class of actions would include my direct actualization of their happiness in the manner of my moving my finger. Such actions are not what I can actually perform but they seem conceptually possible. Consider my desire that the sun is square. I am certainly incapable of any action which counts as making it the case that the sun is square, but it is possible that God is capable. Making the sun square may be for him what it is for me to move my finger.

There is a similarity between actions which count as ‘making it the case that my children are happy’ and actions with abstract content. Consider my desire to greet the tenth person I see tomorrow. On Finlay’s account, I am motivated towards greeting the tenth person I see tomorrow. We can make sense of such a motivation even if we do not know exactly what action that entails. Suppose the tenth person I see tomorrow is John. But it would be wrong to think that my motivation towards greeting the tenth person I see tomorrow is identical to my motivation towards greeting John tomorrow. There are significant differences between the two motivations: for instance, (i) my motivation towards greeting John tomorrow cannot imply my motivation towards greeting Mary

⁴⁶ It may be said that one can be capable of physically manipulating the ‘happiness centers’ of his children’s brains, thereby directly making them happy. I have two reservations about this suggestion. Firstly, it seems we are directly manipulating their brains and thereby only indirectly making them happy, i.e. by actualizing the sufficient conditions for their happiness. Secondly, this suggestion depends on the understanding of happiness as some mental state. It is controversial whether we are not instead mistaking happiness for pleasure. Some construe happiness as more than the agent’s mental state(s) and include also his partaking of goods which are independent of his mind. Happiness, according to such views, cannot be wrought by brain-manipulation. However, even if we can directly make our children happy, my argument against Finlay’s view is not weakened in any way. It merely follows that we need not take indirect actions to bring about our children’s happiness. The distinction between allegedly direct actions (such as brain manipulation) and indirect actions (such as nail-painting) remains. This is all my argument based on *Happiness* requires.

tomorrow, but that can be implied by my motivation towards greeting the tenth person I see tomorrow, and (ii) I am *now* motivated towards greeting the tenth person I see tomorrow but I am not now motivated towards greeting John tomorrow.

The point I want to draw from this comparison is that we can be motivated and yet be unable to act out of those motivations unless some additional conditions obtain. For instance, I need to identify the tenth person I see tomorrow before I can greet him/her. Likewise, I must find out how I would make my children happy before I can (intentionally) go about making them happy. But even so, in such cases, my motivational goal (i.e. what I am motivated to do) *before* the additional conditions obtain is distinct from my motivational goal *after* those conditions obtain. Thus in such cases, the desired end may involve motivation which I cannot act out of consideration for. However, I can realize my desired end when certain conditions obtain, such as my being relevantly informed about how I would realize that end. But we must be careful not to conflate the motivational goal I thereby *derive* with my *initial* motivational goal: the former is only a derivative motivational goal, i.e. the means to some end. I think we are in a position to detect a problem in Finlay's account once we distinguish between (i) what counts as the desired end and motivation towards that end, and (ii) the means to the desired end and motivation towards that means.

1.4.5. Evaluating Finlay's account

Let us consider how his account explains two cases of means-ends motivation. I shall first focus on a case in which the agent 'desires to' rather than 'desires that'. I think that Finlay's account fails to explain cases of 'desiring to' and it is instructive why it

fails. This allows us to see more clearly how it *also* fails with respect to cases of ‘desiring that’. I shall therefore start with a case involving an agent desiring to Φ . Consider the following case:

Golf: I desire to play golf. It happens that I believe I would play golf only if I break a vase. Is it possible that I am not motivated to break the vase?

Let us apply to this case Finlay’s general argument against motivational failure from ends to means. If we insert relevant details into the general argument, then it goes like this:

- (A1) Necessarily, I desire to play golf only if I am motivated towards playing golf.
- (A2) Necessarily, my breaking a vase is a means to my playing golf if and only if my breaking a vase is an action of my playing golf.⁴⁷
- (A3) Necessarily, if I am motivated towards playing golf and believe that my breaking a vase is an action of playing golf, then I am motivated towards breaking a vase.
- (A4) Necessarily, if I am motivated to play golf and believe that my breaking a vase is a means for me to play golf, then I am motivated towards breaking a vase. [A2, A3]

⁴⁷ I take the expression “if breaking a vase is an action of my playing golf” to be logically equivalent to “if breaking a vase is an action of my making it the case that I play golf”. I used the former expression instead of the latter because the latter is cumbersome. However, the point I wish to make applies equally if the latter expression is adopted.

(A5) Therefore, necessarily, if I desire to play golf and believe that my breaking a vase is a means for me to play golf, then I am motivated towards breaking a vase. [A1, A4]

(A2) is clearly false. Even if my breaking a vase is a means for me to play golf, it is not *constitutive of* my playing golf such that it is an action of my playing golf. Rather, it is *contributive to* my playing golf—it is an action which is necessary for my playing golf but it is not part of playing golf. Suppose we amend (A2) to

(A2*) Necessarily, my breaking a vase is a means to my playing golf if and only if my breaking a vase is necessary to bring about my playing golf.

However, even if we assume that (A3) is true, we cannot infer (A4) from (A2*) and (A3). In the general argument, (II) and (III) jointly entail (IV) because M is an action which is constitutive of ‘making E so’. If (A2) is modified to (A2*), that constitutive relation between M and ‘making E so’ is lost. Supposing that I am motivated towards playing golf, it seems clearly possible that I can be so motivated without being motivated towards doing what brings about my playing golf—after all, *what* is necessary to bring about my playing golf is not *itself* my playing golf nor constitutive of it. As I have explained earlier, breaking a vase clearly does not fall under the description ‘playing golf’ even if it is causally related to my playing golf.

In *Golf*, it seems that motivational failure from end to the believed means is possible. Even granting (A3), Finlay’s argument is unsuccessful because it conflates my

given end with the means to it. Let us now turn to a case involving the agent desiring that S and compare it with the earlier case. Consider

Happiness: I desire that my children are happy and believe that they would be happy only if I painted my nails. Is it possible that I am not motivated to paint my nails?

By applying Finlay's general argument, we get:

- (B1) Necessarily, I desire that my children are happy only if I am motivated towards making it the case that my children are happy.
- (B2) Necessarily, my painting my nails is the means to its being the case that my children are happy if and only if my painting my nails is an action of making it the case that my children are happy.
- (B3) Necessarily, if I am motivated towards making it the case that my children are happy and believe that my painting my nails is an action of making it the case that my children are happy, then I am motivated towards painting my nails.
- (B4) Necessarily, if I am motivated towards making it the case that my children are happy and believe that M is the means to its being the case that my children are happy, then I am motivated towards painting my nails. [B2, B3]

(B5) Therefore, necessarily, if I desire that my children are happy and believe that my painting my nails is the means to its being the case that my children are happy, then I am motivated towards painting my nails. [B1, B4]

As I have explained earlier, there is a class of actions which fall under the description ‘making it the case that my children are happy’. By virtue of my desire that my children are happy, I am motivated towards performing actions of this class. The nature of desire accounts for such motivation. There is another class of actions which does not fall under the aforementioned description, even though these actions help make it the case that my children are happy. These actions are means to bringing about my end, but they do not count as ‘making it the case that my children are happy’ just as my breaking a vase does not count in my earlier example as ‘playing golf’. However in *Happiness*, I am incapable of any of the actions which count as ‘making it the case that my children are happy’—I cannot simply make it the case that my children are happy unlike how I can play golf. I must therefore avail myself of the actions of the second class: to do what would make it the case that my children are happy.

With this distinction in mind, I think (B2) should be rejected. The motivational goal that I have in virtue of my desire that my children are happy is distinct from the motivation to paint my nails. With reference to *Golf*, my breaking a vase is not an action of my playing golf; likewise, my painting my nails is not an action of making it the case that my children are happy. Painting my nails is contributive to—not constitutive of—making it the case that my children are happy. I do not have the former simply because I

have the latter and believe that the two motivational goals are causally connected. Something else is needed to bridge the gap and it is not clear how the nature of desire provides it. Desire accounts for motivation to the ends but not necessarily also to the means: since the means is not constitutive of the ends, it is not clear why motivation to the ends *must* include motivation to the means.

It is crucial to note that on a motivation account of desire, my desire for E is just my motivation towards making E so, not both making E so and making M so. Finlay thinks making M so is constitutive of making E so, but these are distinct and there is no reason to think the former constitutes the latter. However, in premise (II) of his general argument, actions which only instrumental are conflated with actions which are ends in themselves. Finlay writes: “To believe some action to be a ‘means’ to an end E is simply to believe that M is an action (or part thereof) of making E so, or a facilitating condition for E.”⁴⁸ Suppose we take Finlay here to mean that M is constitutive of making E so because by ‘making E so’, he means to include *all the causal processes involved in bringing E about*. In this case, he would include my vase-breaking as part of the process of how I come to play golf, and my nail-painting as part of how my children come to be happy. This would be the loose sense of ‘making E so’ and the strict sense involves only the immediate bringing about of E.

But if this is the case, it is not clear why we should accept (I) and correspondingly, (A1) and (B1). If my desire is just for E, why must my motivation be towards what is *not* E, even if I believe they are the means to making E so? It cannot be replied that because in desiring E, I am motivated towards making E so and *thus* I am motivated towards bringing about what is *not* E, but merely a means to making E so. This

⁴⁸ Ibid, 14.

is because it is clearly conceptually possible to understand my motivational goal to be identical to my desired end. Within the motivation account of desire, when I desire to play golf, it is possible to take me as being motivated just to play golf. Whether I come to be motivated to do what is necessary to allow me to play golf is a separate matter and *need not* be built into my desire to play golf. If it is conceptually necessary to take the loose sense of ‘making E so’, some further argument is needed. The motivation account of desire does not necessitate it.

One may ask why there is a significant difference between the strict and loose senses of ‘making E so’ and the corresponding kinds of motivation. If the agent who desires E can be motivated to make E so in an immediate way, why can’t he be motivated to make E so in less immediate ways? I shall explain what I see as the crucial difference by an example. Suppose I believe that desires are best explained as mental states that are motivating. Take my desire to play golf. If one finds a motivational account of desire plausible, it seems plausible to thereby take my desire to play golf as my motivation to play golf. If desires are explained in terms of motivations, the desired end is plausibly understood as being identical to the motivational goal. It seems less plausible to take my motivational goal as involving more than what my desired end involves: if desires are motivations, why are their objects different? There should be an explanation why this is so, let alone why it must be so. I have argued that Finlay has not shown why this must be so and this leaves open the conceptual possibility that the goal of the motivation is just what the desired end is. Thus, the significance between the two senses of ‘making E so’ is not fundamentally about more and less immediate ways of bringing E about. It is about how the desire for one thing can motivate action of doing something *else*.

It may be objected that my making it the case that my children are happy is merely a means to my desired end, which is its being the case that my children are happy. As such, the nature of desire does itself explain motivation to the means.⁴⁹ My response is to note that on Finlay's account, my desire that my children are happy just is my motivation towards making it the case that my children are happy. Here, is *that* motivation the motivation towards making the end so, or is it the motivation towards making the means so? For there to be means-ends motivation, the ends must be conceptually distinct from the means, or else there is no motivational transfer. Suppose my motivation towards making it the case that my children are happy is my motivation towards making the means so. But what then is my motivation towards making the ends so? Where is the motivation towards the means transferred from?

Since Finlay defends the claim that my desire for E is my motivation towards making E so, it seems to me that he would *not* consider my motivation towards making it the case that my children are happy an instance of *my making M so*. On his view, I would be making E so. But this means that I am not merely making the means obtain for some distinct end; I am making the end obtain. As such, I conclude that this objection misunderstands the view which Finlay defends, leading to confusion between the end and the means.⁵⁰

⁴⁹ It is noteworthy that this objection does not seem to arise in the case of 'desires to'. It seems evident that when I desire to play golf and therefore play golf, I am not playing golf as a means to the further end of 'making it the case that I play golf'. My playing golf is identical to my making it the case that I play golf.

⁵⁰ It could also be objected that my desire to play golf is more accurately described as my desire to play golf *by any means*. Thus, on a motivation account of desire, I am thereby motivated to play golf by any means. It seems then that my desired end has the means built into it. My response is that in such a case, there is no genuine motivational transfer from ends to means. This is because the stated means are already included *as part of the end*. The motivation involved is only motivation towards that given end. (The spirit of this objection is taken from Alexander Pruss, "Means-End Reasoning" in *Alexander Pruss's Blog*.)

I have not argued that Finlay's account of means-end motivation is wrong, only that he has not shown it to be right. As such, it remains that his account is a possibility, and our desires are possibly as he explains. Nonetheless, he has not shown the impossibility of our having desires for ends which do not necessarily involve motivation towards the means. Insofar as such desires are possible, we can exploit them in thought experiments to argue that the Instrumental Requirement is possible and needs to be accounted for by a theory of reasons.

1.5. Conclusion

In sections 1.1 and 1.2, I have argued that the internalist is faced with a dilemma: he either has to accept that there are two requirements of reason—the End and Instrumental Requirements—or accept claims which are highly-counterintuitive. I then considered three possible ways for the internalist to dissolve that dilemma. The first way involves avoiding those implausible claims without invoking any problematic requirement of reason. This way fails because it relies on reasoning that is invalid. The second way involves exploiting a distinction between rationality and reason to argue that only requirements of rationality are needed in certain cases, which pose no problems for the internalist theory of reasons. This way fails because it is hard to make sense of rational requirements without requirements of reason. The third way is to argue that the Instrumental Requirement can be replaced by the nature of desire. This way fails because it requires a concept of desire which we are given no compelling reason to accept, either as the concept of our desires or as the only possible concept of desire. As such, the Instrumental Requirement has not been undermined.

If my arguments are successful, the internalism is implausible on either side of the dilemma. If it entails the rejection of the requirements of reason, then it seems an implausible view about what tracks reasons since it entails outcomes which are difficult to accept. But if it allows for those requirements, then it is false since some external reason statements are true. Therefore the distinction between internalism and externalism as defined by Williams is of limited use because only externalism is not clearly false. Insofar as Williams accepts those requirements, he is an externalist in the sense given by *his* distinction between internalism and externalism. It seems nonetheless that there are important differences between William's view of reasons and those whom he considers externalists. In Part 3, I shall attempt to reformulate the internalism/externalism distinction to capture the most significant differences. But for now, I shall consider Williams's argument against externalism, as he defines it. If his argument is sound, then we have no reasons for action.

Part 2: Can there be external reasons?

2.1. Williams's argument against external reasons

Williams's distinction between internal and external reasons is followed by his argument that there are no external reasons, and only internal reasons. He begins his argument with the claim that the agent has a reason to Φ only when the agent can Φ for that reason; the reason can figure in the explanation of the agent's action.⁵¹ In other words, a reason R is a normative reason for the agent only if R can be his motivating reason. This is an adaptation of the principle 'Ought implies can'. There is dispute over what sense of *can* is involved which I wish to avoid as far as possible in this essay. I shall therefore understand the sense of 'A can Φ ' as either (i) it is naturally possible for A to Φ or (ii) it is naturally possible for A to Φ if A so chooses to. I think nothing significant hangs on accepting either (i) or (ii) for the purposes of my essay. I shall agree with Williams that

- (A1) If there is a reason for A to Φ , it must be possible for A to Φ for that reason.

Externalists hold that

- (A2) Whether A has a reason to Φ is not contingent on whether there is a sound deliberative route from A's subjective motivational set to A's Φ -ing. [RE]

⁵¹ Williams terms this "the explanatory dimension" of reasons. In this regard, he writes: "If there are reasons for action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action." *IER*, 102.

It is possible that the agent is ascribed an external reason to Φ and in fact Φ s, but yet not having acted for the external reason. This can happen when A Φ s simply out of brute instinct. It can also happen because A Φ s for an internal reason and the action merely coincided with what was prescribed by the external reason statement.

Thus, the agent Φ s for an external reason only when two conditions are satisfied: (a) he Φ s in rational response to his belief that he has reason to Φ , and (b) it is not the case that he is motivated to Φ through deliberation from his subjective motivational set (or else he is Φ -ing for an internal reason). This means that the externalist is committed to claiming that it is possible for the agent to Φ in rational response to his reason-belief even when there is no sound deliberative route from A's subjective motivational set to A's Φ -ing. More than that, the externalist cannot avail himself of any such deliberative route, even the unsound ones. This is because the agent who Φ s due to unsound deliberation from her motivations (perhaps she has some relevantly false beliefs) is merely Φ -ing for what she (wrongly) believes to be an internal reason, not for an external reason. Thus, according to Williams, the externalist must accept

- (A3) If it is the case that whether A has an external reason to Φ is not contingent on whether there is a sound deliberative route from A's subjective motivational set to A's Φ -ing, then it is possible for A to Φ for that reason without being motivated through a deliberative route from A's subjective motivational set to A's Φ -ing.

Williams argues that the externalist is thereby committed to an impossibility, for there is no way for the agent to act for a reason apart from deliberation from his motivations.

Thus, Williams claims

(A4) It is impossible for A to Φ for that reason when there is no sound deliberative route from A's subjective motivational set to A's Φ -ing.

Given (A1) and (A4), it follows that

(A5) There are no external reasons for action.

Why think (A4) is true? Williams thinks the externalist has to show how the agent comes to be motivated to Φ through sound deliberation, whatever his original motivations may be.⁵² He argues that we have good reason to think this cannot be done:

“But if this is correct... it is very plausible to suppose that all external reason statements are false. For *ex hypothesi*, there is no motivation for the agent to deliberate *from*, to reach this new motivation. Given the agent's earlier existing motivations, and this new motivation, what has to hold for external reason statements to be true, on this line of interpretation, is that the new motivation could be in some way rationally arrived at, granted the earlier motivations. Yet at the same time it must not bear to the earlier motivations the kind of rational relation which we considered in the earlier discussion of deliberation—for in that

⁵² Ibid, 109.

case an internal reason statement would have been true in the first place. I see no reason to suppose that these conditions could possibly be met.”⁵³

At this point, it may be thought that Williams is begging the question against the externalist by simply assuming that the only form of deliberation is procedural, where the agent is deliberating *from* his existing motivations. Thus Brad Hooker writes:

“The external reasons theorist will not only reject Williams’s claim that external reasons are impossible but will also be dissatisfied with Williams’s conception of rational practical deliberation—in contrast to Williams, the external reasons theorist is likely to think that (at least some) rational deliberation about reasons for action starts not from the agent’s present motivations, but from some objective (‘external’) values or requirements, fixed independently of the agent’s present motivations.”⁵⁴

Unless Williams shows why practical deliberation must be procedural, Hooker thinks the externalist would simply take (at least part of) it to be substantive, where deliberation aims to track values or “features that would make the thing worth desiring as an end in

⁵³ Ibid, 109. John McDowell agrees on this point: “As Williams describes his position, the external reasons theorist must envisage a procedure of correct deliberation or reasoning that gives rise to a motivation, but that is not ‘controlled’ by existing motivations... [he] has to envisage the generation of a new motivation by reason, in an exercise in which the directions it can take are not determined by the shape of the agent’s prior motivations—an exercise that would be rationally compelling whatever motivations one started from. As Williams says, it is very hard to believe there could be a kind of reasoning that was pure in this sense—owing none of its cogency to the specific shape of pre-existing motivations—but nevertheless motivationally efficacious.” (“Might There Be External Reasons?,” *Mind, Value, and Reality*, 99.)

⁵⁴ “Williams’ Argument Against External Reasons,” in *Varieties of Practical Reasoning*, 100. See also Hooker, B. and Streumer, B. “Procedural and Substantive Practical Rationality,” *The Oxford Handbook of Rationality*, 70-1.

itself”.⁵⁵ The externalist would then be able to accommodate Williams’s claim that if the agent rationally deliberated—whatever his present motivations—he would come to be motivated to Φ .

However, this is too swift. May there not be a better way to understand Williams’s argument? I shall now explore this question by formulating in detail what seems to be his reasoning. Williams thinks that a belief by itself cannot motivate action: the agent can be motivated to act only if he desires to perform that action, or desires an end which would be brought about by that action.⁵⁶ I shall follow Williams in accepting

(B1) It is possible for A to Φ only if A desires to Φ . (where Φ refers to an action)⁵⁷

As explained earlier, the externalist must meet two conditions, (a) and (b). Let us express this requirement as

⁵⁵ “Procedural and Substantive Practical Rationality,” 68.

⁵⁶ (B1) articulates the Humean view of motivation and is not uncontroversial. For instance, David McNaughton argues that if desire is taken as a non-cognitive state, (in some cases) the agent’s belief is sufficient in explaining his action because some beliefs possess directions of fit that run both ways: the belief aims to fit the world and is about how the world should be. McNaughton’s view is that motivation should be understood as a cognitive state and there is no need for the agent to first desire to Φ in order for him to Φ . See his *Moral Vision*, especially chapter 7. Russ Shafer-Landau thinks this is the wrong way to object to Williams’ argument, for once we allow that evaluative beliefs are sufficient to motivate, these beliefs count as part of the agent’s motivational set. Externalists then have to show that the agent could be motivated independently of his desires and evaluative beliefs, and in his view that is far harder than to defend externalism while accepting (B1). (See his *Moral Realism*, 175, footnote 11.) Shafer-Landau seems wrong here for a core issue dividing internalists and externalists is whether one can be motivated independently of one’s *antecedent* motivations. If the agent forms an evaluative belief (the truth of which is independent of his motivations) and that is sufficient to motivate him, then he can be motivated independently of his antecedent motivations—he can thereby act for external reasons.

⁵⁷ Or if A desires E and believes that Φ -ing brings about E.

(B2) For A to Φ for some reason, there must be a motivational route from A's belief that he has a reason to Φ to A's Φ -ing..⁵⁸

It is at this point that Williams makes the crucial claim that the externalist cannot make sense of such a motivational route because “there is no motivation for the agent to deliberate from, to reach this new motivation”.⁵⁹ If there is such a motivation, then the agent would instead come to Φ for an internal reason.

There is an ambiguity here: what sense of *reach* is Williams employing in the quote above? He may mean that the agent cannot *rationally* reach the new motivation from his original motivations, since his motivational set does not give him any *reason* to have that new motivation. This would indeed be begging the question against the externalist because Williams's argument is aimed to show that there are no reasons that obtain independently of the agent's motivational set, and Williams has assumed in a premise that there are no such reasons. Hooker's charge would therefore be accurate. However, that is not the only sense of *reach* which is available to Williams. He may instead mean that the agent cannot *psychologically* reach the new motivation from his original motivations. The difficulty in taking this interpretation is that Williams does not provide any reason to think so. But is there support for his claim even if he does not offer it? I shall consider two such arguments.

2.2. The *Direction-of-fit* argument

⁵⁸ As stated by (a) and (b), there are at least two ways that the motivational route could be of the wrong kind.

⁵⁹ *IER*, 109

The first argument is based on the nature of desire—I shall refer to it as the *Direction-of-fit argument*. To be in a state of desiring is to desire the world to be in a certain way: when I desire to drink water, I desire that the world is such that I drink water. I would not stop desiring to drink water just because it is not true that the world *is* such that I drink water. On the contrary, it seems that I would stop desiring to drink water if it is already true that I drink water.⁶⁰ A prominent way of expressing the nature of desire is contrasting its direction of fit with that of belief.⁶¹ Beliefs aim to capture how the world is like. Put another way, they seek to fit the world. Thus, when I am given compelling reason to think that the world is not what I thought it is like, I would revise my beliefs accordingly. But that is not so with desire, as explained earlier: desires aim to make the world fit their contents. Therefore, beliefs and desires have opposite directions of fit.

Reasoning is concerned only with truth. However, only beliefs are truth-apt while desires are not. It seems to follow that desires are by their nature not responsive to reasoning. We therefore cannot reach a desire simply by reasoning and having some belief: desires cannot be affected in that way. But do we not gain/lose desires in response to what we believe about the world, such as how it is like for us to be in some state? We can accommodate that while denying that this constitutes reasoning about desires: we need only claim that such gains/losses of desires are not in *rational* response to what we believe about the world. Thus I desire E and also desire M because I believe that M makes it the case that E. However, I come to lose that belief. It may be claimed it is not

⁶⁰ I may instead desire that I continue to drink water. I leave it open that I could continue to have a second-order desire to drink water.

⁶¹ This idea is commonly said to originate in Elizabeth Anscombe's *Intention* (see especially p. 56), although she does not herself employ the phrase 'direction of fit'.

through reasoning that I stop desiring M, but simply what happens given the nature of desire. To desire E just is to desire to make it the case that E. If I desire E and believe that M makes it the case that E, then I automatically come to desire M and not through reasoning that I should desire M. If I hold such a belief but do not come to desire M, then it cannot be true that I desire E. Stephen Finlay defends such an account of desire and I have examined it in an earlier section. It is sufficient for my present purposes just to introduce this account to indicate the resources available to the defender of the *Direction-of-fit* argument.

We can now see why one may think that deliberation must proceed from one's present motivations: because that is the only way the agent can be motivated to move down the deliberative route which leads to a new motivation. It may be thought that beliefs are in themselves motivationally inert. They lead to new motivations only by channeling motivational force from already existing motivations. If there is no relevant motivation, beliefs can have no effect on the agent's motivations.⁶² The agent desires E and since he believes that he would bring about E if he Φ s, he is moved *by virtue of the nature of desire* to take the path shown by reason, connecting his desire for E with the motivation to Φ . Reason tells him the way but does not motivate him.

Let us suppose the agent has no deliberative route to the desire for E, but still the externalist ascribes to him a reason to Φ . There is no desire in the agent to motivate him

⁶² This may seem a problem for Williams since he accepts a broad understanding of deliberation, including instances where the agent gains/loses a completely new motivation through imagination and reflection—instances where beliefs are formed. I think these are compatible with the account of desire under consideration. Motivations which occur with such beliefs are only derivatively affected by them. What happens is similar to how means-ends beliefs affect motivation, according to the *Constitutive* account of desire. In the case of imagination and reflection, new motivations are formed in non-rational ways, the causes of which are the agent's (brute) disposition to respond from such motivations in response to such beliefs. Williams could claim that the agent is disposed to Φ iff he would Φ if he believes P, where Φ denotes the loss or gain of a motivation.

to Φ . How then is that reason ever to figure in a correct explanation of the agent's action? There may be non-rational ways of motivation, like the use of moving rhetoric, but these are of no help to the externalist as they are of the wrong kind. After all, the agent must be able to act *for* the external reason and non-rational responses are not acting for reasons.

As such, Williams need not be merely begging the question against the externalist if he claims

- (B3) There is a motivational route from A's belief that he has a reason to Φ to A's Φ -ing only when there is a sound deliberative route from A's subjective motivational set to A's Φ -ing.

Recall (A3). The externalist is committed to

- (B4) If it is possible for A to Φ for an external reason, then it is possible for A to Φ for that reason without being motivated through a deliberative route from A's subjective motivational set to A's Φ -ing.

But given (B3), it follows from (B4) that

- (B5) If it is possible for A to Φ for an external reason without being motivated through a deliberative route from A's subjective motivational set to A's Φ -ing, then it is possible for A to Φ for that reason when there is no

motivational route from A's belief that he has an external reason to Φ to A's Φ -ing. [B3, B4]

The externalist is then in trouble since (B2) and (B5) entail

(B6) It is impossible for A to Φ for an external reason when there is no motivational route from A's belief that he has an external reason to Φ to Φ -ing.

Given (B4)-(B6), we must accept

(B7) It is impossible for A to Φ for an external reason.

And by virtue of (A1), Williams concludes that

(B8) External reasons are impossible.⁶³

2.3. Is the *Direction-of-fit* argument sound?

The *Direction-of-fit* argument can be stated as follows:

(C1) Belief is concerned only with truth.

⁶³ Williams does not deny we *seem* to ascribe external reason statements to others but suspects that they are in fact "optimistic internal reason statements" which we attribute to others in "hope that somewhere in the agent is some motivation that by some deliberative route might issue in the action we seek." ("Internal Reasons and the Obscurity of Blame," in *Making Sense of Humanity*, 40) In this regard, see also Stephen Finlay's "The Reasons That Matter," in *Australasian Journal of Philosophy*, especially p. 18.

(C2) Desires are not truth-apt.

(C3) Desires are responsive to beliefs only if they are truth-apt.

Therefore,

(C4) Desires are not responsive to beliefs.

The argument is valid and I shall accept (C1) and (C2). However, I reject (C3). The justification for (C3) is that since desires do not aim to fit the world, they are immune to beliefs, which aim to fit the world. The claim seems to be that unless the aims of beliefs and desires are identical or sufficiently alike, the latter is not responsive to the former. There is some apparent force in this claim since beliefs appears irrelevant to desires unless there is some nexus for interaction. A common aim provides this and without it, it seems hard to see how interaction can occur between such states. As argued earlier, to make sense of acting for a reason, this nexus must be of the right sort. A purely causal connection fails to capture what it means to act for a reason.

While the demand for a nexus of interaction seems justified, (C3) is unconvincing. Why must desires be truth-apt in order to be responsive to beliefs? The need for a common aim which connects beliefs to desires casts an unwarranted restriction upon the nature of the connection. This is because even if we accept that beliefs and desires have opposing directions of fit in virtue of their respective natures, it can be the case—as I shall explain—that they are connected, nonetheless, precisely because they are of their respective natures.

(C3) derives its force from ignoring how else beliefs and desires can be connected: it can be *the agent* who is believing and desiring. Insofar as the agent is rational, he responds appropriately to what he believes he has reason to do. Therefore, it is possible that he guides his various activities *with the same aim*: to be responsive to reason. As stated in (C1), beliefs aim only at discovering truth but there can be truths about what we ought to desire (or do). There is then a relevant nexus of interaction between beliefs and desires: the rational agent himself. If I believe that I ought to stop disregarding my future wellbeing, I would cease to desire to do so insofar as I am rational. Given the agent's rational capacity, desires need not be truth-apt in order to be responsive to beliefs. They need instead to be apt for rational assessment and that seems at least possible. Therefore, I conclude that the *Direction-of-fit* argument is unsound.

2.4. The *Desire-out, Desire-in* argument

Michael Smith offers another argument for the claim that the agent cannot psychologically reach a new motivation apart from his original motivations. This argument, which I shall refer to as the *Desire-out, Desire-in* argument,⁶⁴ supports (B3) and we can then arrive at (B8) in the manner explained earlier. Smith's first distinguishes between motivating and normative reasons. He takes motivating reasons to be teleological explanations of the agent's actions: "explanations that explain by making what they explain intelligible in terms of the pursuit of a goal."⁶⁵ In his view, motivating reasons need not be causal explanations (or what I shall refer to as explanatory reasons) although he thinks that is possible. On the other hand, normative reasons are independent

⁶⁴ Following Jay Wallace in his "How to Argue about Practical Reason," in *Normativity and the Will*, 30.

⁶⁵ "The Humean Theory of Motivation," in *Mind*, 44.

of the agent's perspective. To say that the agent has a normative reason to Φ "is to say that there is some normative requirement that he Φ s. It is therefore to justify his Φ -ing from the perspective of the normative system that generates that requirement."⁶⁶

Smith argues that for the agent to be motivated for a reason is for him to have a goal:

"It is thus natural to suppose that an agent's motivating reasons are, as we might put it, psychologically real, for it would seem to be part of our concept of what it is for an agent's reasons to have the potential to explain his behaviour that his having those reasons is a fact about him; that is, that the goals that such reasons embody are his goals."⁶⁷

He then proceeds to analyse what it is for the agent to have a goal, and eventually argues for these three claims:⁶⁸

- (1) "Having a motivating reason is, *inter alia*, having a goal."
- (2) "Having a goal is being in a state with which the world must fit."
- (3) "Being in a state with which the world must fit *is* desiring."

Thus, Smith claims that motivating reasons consist of belief/desire pairs. He expresses this as

⁶⁶ Ibid, 39.

⁶⁷ Ibid, 38.

⁶⁸ Ibid, 55.

(P1) “R at t constitutes a motivating reason of agent A to Φ iff there is some Ψ such that R at t consists of a desire of A to Ψ and a belief that were he to Φ he would Ψ .”⁶⁹

P1 entails what Smith considers a weaker claim:

(P2) “Agent A at t has a motivating reason to Φ only if there is some Ψ such that, at t, A desires to Ψ and believes that were he to Φ he would Ψ .”⁷⁰

The externalist wants to deny (A4) and so he has to deny (B3), yet he also accepts (B1). Thus he has to accept (P2) while denying (P1). This cannot be done, Smith argues, because (P2) is true only if (P1) is true.

To see why he thinks so, consider the following example of Thomas Nagel’s:

Headache: I want my headache to go away and I believe that taking aspirin makes it go away. Therefore, I take aspirin.

Nagel observes: “I take aspirin because I recognize that my desire to be rid of the headache gives me a reason to take it, justifies my wanting to take it. That’s why I take it, and from inside I can’t think that it would be no less rational to bang my head against a fire hydrant instead.”⁷¹ Nagel distinguishes between motivated and unmotivated desires.

⁶⁹ Ibid, 36. Here, R refers to a reason statement, t refers to a particular point in time, while Φ and Ψ refer to actions.

⁷⁰ Ibid, 36.

⁷¹ *The View from Nowhere*, 152.

Unmotivated desires have only causal explanations while motivated desires admit of rational explanations, i.e. justifications. He thinks my desire to take aspirin is a motivated desire, which is explained not by another desire but by my recognition that I have a *reason* to take aspirin in virtue of my desire to stop the headache. It is this recognition which explains why I would not consider taking aspirin to be no less rational than banging my head against the fire hydrant. If a desire lies behind every action and there are no motivated desires, then it would be true that in *Headache*, it is no less rational for me to bang my head against the fire hydrant than for me to take aspirin. This is because there would be no rational response *at all* to my having a headache: any subsequent action of mine would be due to some unmotivated desire and has only a causal explanation.⁷² Nagel thinks it can be true that a desire lies behind every action but some desires are motivated. That would allow the externalist to accept (P2) while rejecting (P1). Thus he explains:

“if the desire is a motivated one, the explanation of it will be the same as the explanation of his pursuit, and it is by no means obvious that a desire must enter into this further explanation. Although it will no doubt be generally admitted that

⁷² An anonymous examiner of this essay thinks that an internalist “should here advance a view of rationality that does not require the possibility of motivated desires.” He adds: “If rationality is just a matter of acting in ways that maximize expected desire-satisfaction, it can be rational to take aspirin. This does not require the possibility of motivated desires.” It seems to me that such a view is unpromising, although I can only briefly explain my worries about it. The defender of this view faces a dilemma. If the agent is motivated by the consideration of maximizing expected return, then it seems he is acting out of a motivated desire. Since motivated desires are not available within this view, the remaining option seems to be that the agent’s rationality consists wholly in acting in a specified manner, i.e. the manner which maximizes his expected return. However, it does not seem sufficient that the agent does so, since his actions could match the specified description purely by chance. Such an agent may seem to be acting rationally, but surely it matters — if we are assessing if he is in fact acting rationally — why he acts the way he does. Afterall, given the appropriate conditions, he could maximize his expected return through wholly capricious actions, or even actions aimed to minimize his expected return (we need only tweak how his intentions are causally related to his actions).

some desires are motivated, the issue is whether another desire always lies behind the motivated one, or whether sometimes the motivation of the initial desire involves no reference to another, unmotivated desire.”⁷³

Smith argues that Nagel’s reply is inconsistent with either (1) or (3) and should therefore be rejected:

“A motivated desire is a desire had for a reason; that is, a desire the having of which furthers some goal that the agent has. The agent’s having this goal is, in turn, *inter alia*, the state that constitutes the motivating reason that he has for having the desire (from (1)). But if the state that motivates the desire is itself a reason, and the having of this reason is itself constituted by his having a goal, then, given that the having of a goal is a state with which the world must fit rather than vice versa (from (2)), so it follows (from (1)) that the state that motivates the desire must itself be a desire. Thus, the Humean will say, the idea that there may be a state that motivates a desire, but which is not itself a desire, is simply implausible.”⁷⁴

He thinks Nagel cannot drive a wedge between (P1) and (P2) by distinguishing between motivated and unmotivated desires because the former must ultimately be explained by the latter—or else we face an infinite regress of desires, which is problematic at least because we think the sum of our desires are finite. Likewise, if we are to avoid

⁷³ *The Possibility of Altruism*, 46.

⁷⁴ “The Humean Theory of Motivation,” 59.

postulating an infinite number of motivating reasons, we must hold that no motivating reason can be a state which is not itself a desire. This means that if we accept (P2), we must also accept (P1). This argument, if sound, shows that whenever the output of deliberation is a desire, it must have been that there was an input of some other desire.

2.5. Is the *Desire-out, Desire-in* argument sound?

Smith's argument is problematic because it either equivocates between two senses of reasons or it simply begs the question against the externalist. There are three senses of reasons to be distinguished here: *explanatory* reasons are causal explanations of the agent's action; *normative* reasons are justifications of the agent's actions; *motivating* reasons are what the agent takes to be normative reasons which he thereby acted for. Consider a case of my firing a gun. Here, the content of the explanatory reason would include my moving my finger and pressing it against the trigger (along with the relevant causal mechanisms which are thereby activated). My motivating reason is to execute a person whom I believe ought to be executed. However, my normative reason in that case is to prevent the execution of that person because he was maligned. In this case, my false belief led to a divergence between my motivating and normative reasons: I acted for what I wrongly believed to be my normative reason.

Smith is aware of these three kinds of reasons but nonetheless seems to confuse motivating reasons with explanatory reasons (although he clearly differentiated them). Consider his claim that motivating reasons are *constituted* by belief/desire pairs.⁷⁵ Motivating reasons are the normative reasons for which the agent acted. But the

⁷⁵ As Smith claims in (P1), "R at t constitutes a motivating reason of agent A to Φ iff there is some Ψ such that R at t consists of a desire of A to Ψ and a belief that were he to Φ he would Ψ ." (Ibid, 36) See also p.54, 55, 60, 61.

normative reasons cannot be some psychological state since they are propositional rather than psychological. The content of the agent's justification of his own action cannot be a psychological state since that cannot enter into justificatory relations with propositional claims—and that is the role of normative reasons.⁷⁶

Thus Smith may instead be claiming that explanatory reasons are constituted by belief/desire pairs. We need then to substitute all his claims about motivating reasons into claims about explanatory reasons. But if that is the case, Smith's response to Nagel equivocates between explanatory and motivating reasons. In speaking of motivated desires as desires had for reasons, Nagel is concerned only with motivating reasons. (1) and (2) are then not inconsistent with his claim, but irrelevant to it. Perhaps Smith is in fact concerned with motivating reasons but failed to properly express his claims. He may be claiming that motivating reasons are *given by* belief/desire pairs rather than being constituted by them. On that interpretation, we would take Smith to claim that the agent's motivating reasons are therefore reasons to satisfy his desires. We should then amend (1) and (P1) to reflect this. Thus,

(1*) Having a motivating reason is, *inter alia*, dependent on having a goal.

(P1*) R at t is a motivating reason of agent A to Φ iff there is some X such that
R at t is given by a desire of A to X and a belief that were he to Φ he
would X.

⁷⁶ Thus Finlay notes: "Smith's error is to identify the motivating state as the motivating reason. Insofar as something motivates us qua reason, it does so by appearing at least minimally normative to us." ("Motivation to the Means," 5.)

This interpretation faces its own problem, namely, that it is not clear why (1*) is true and likewise, (P1*). The externalist would deny both claims and Smith provides no argument for them. *Prima facie*, it seems plausible that I can be motivated by reasons which are not goal-dependent, such as my refusal to engage in the murder of an innocent person for the reason that it is evil to do so.⁷⁷ It is still the case that (P2) is entailed by (P1) but we have no reason to think (P2) is true only if (P1) is. The externalist therefore need not accept (B3) and likewise (A4). Therefore, Williams's argument has not been shown to be sound.

However, even if motivating reasons need not involve desires, is that true of explanatory reasons? Consider

(P3) R at t is an explanatory reason of agent A to Φ iff there is some X such that R at t consists of a desire of A to X and a belief that were he to Φ he would X.

The externalist can however accept (P3) because it can be the case that the agent Φ s due to his belief/desire pair even while his relevant desire is in turn explained by a state which is not another desire. The externalist encounters a problem only when Φ -ing is taken even to include desiring. Thus, the explanatory reasons for all desires in turn have to invoke some antecedent desire. But surely this poses a problem for the internalist too. For taken that way, (P3) requires that we accept either an infinite regress of desires or desires that exist uncaused. Both options are unpalatable for externalists *and* internalists. We do not have to face this dilemma because there can be causes of desires that are not themselves desires.

⁷⁷ I consider counterexamples to (1*) and (P1*) further on p. 89-90 of this essay.

Williams considers one such explanation of desires when he writes:

“It might be said that the force of an external reason statement can be explained in the following way. Such a statement implies that a rational agent would be motivated to act appropriately, and carry this implication, because a rational agent is precisely one who has a general disposition in his S to do what (he believes) there is reason for him to do. So when he comes to believe that there is reason for him to Φ , he is motivated to Φ , even though, before, he neither had a motive to Φ , nor any motive related to Φ -ing in one of the ways considered in the account of deliberation.”⁷⁸

I have argued in the preceding part of this essay that there is no reason to think an antecedent desire is required—either in motivating or explanatory reasons—to account for the agent’s desire to Φ when he believes he has reason to Φ . As Williams is considering, it may be that the rational agent’s disposition is such that his desires are directly responsive to his beliefs. But wouldn’t this disposition count as part of the agent’s antecedent motivational set?⁷⁹ Since the allegedly external reasons motivate the agent in this manner, he then seems to be deliberating from his antecedent motivation—he would not be acting for genuinely external reasons.

⁷⁸ *IER*, 109. The account considered by Williams is explained by Derek Parfit: “We can also respond directly to our object-given reasons to have certain desires, by coming to have them, and then continuing to have them. But, at least in most cases, these responses are not voluntary. These responses are, in this respect, like our responses to epistemic reasons. Though our responses to these two kinds of reason are seldom under our direct voluntary control, they are not things that merely happen to us, like an automatic knee-jerk, or our slipping on a banana skin. Our being rational consists in part to our responding to such reasons in these non-voluntary ways.” (*Climbing the Mountain*, 32-3.)

⁷⁹ Here, one may point to how Williams explicitly includes “dispositions of evaluation” as part of the agent’s subjective motivational set. See *IER*, 105.

Admittedly, it is conceptually true that the agent cannot be motivated *except* by the involvement of his existing motivational abilities. However, the key issue is whether that truth entails that all practical deliberation proceeds from his existing motivations.⁸⁰ I do not think there is such an entailment. While the agent comes to be motivated through his antecedent motivational abilities, the *considerations* which he is deliberating on are not themselves those abilities. The internalist claims that all practical deliberation concerns how best to satisfy the relevantly-informed desires of the agent, while the externalist thinks that it can go beyond that—the agent can deliberate from considerations which do not ultimately concern the satisfaction of such desires. Considerations like suffering and the worth of various objects or state of affairs can ground practical deliberation—so the externalist claims. Thus, when the agent is motivated to Φ upon believing that he has an external reason to Φ , he is not deliberating from his antecedent motivations since the truth-condition for that reason is independent of his motivational set. He does not reason: since I am disposed to be motivated in accordance with my external reason beliefs, I have a reason to act upon my disposition, and derivatively, my external reason beliefs.

That said, it is true that the externalist appeals to the agent's antecedent motivational disposition to explain *how* he is motivated to act, upon believing that he has a reason to act. Nevertheless, it does not follow the agent is therefore deliberating from his antecedent motivations. The crux of the debate between internalists and externalists is

⁸⁰ Thus Jay Wallace writes: "...the sense in which motivation presupposes a prior disposition to action is the truistic sense in which it is conceptually impossible to acquire through deliberation a motivation that is independent of the dispositions to which one was antecedently subject. So interpreted, however, the claim fails to rule out any of the interesting options in the theory of rational motivation that one might wish to endorse. In particular, it does not rule out the idea that correct reflection on one's moral reasons might suffice to generate a corresponding motivation, even if, prior to the episode of deliberation, one lacked any identifiable desire from which the new motivation might have been derived." ("Three Conceptions of Rational Agency," in *Normativity and the Will*, 46)

about whether the agent's practical deliberation must proceed from his antecedent desires, such that all reasons for actions are grounded on such desires. The internalist holds that the agent is motivated by *reasons*, even if those reasons are all based on his desires. Thus, even the internalist has to explain *how* (internal) reasons motivate the agent. If he accepts (B1), then his task is to show how the agent's beliefs about his reasons lead to the formation of his corresponding desires. It must be clear how reason has the appropriate explanatory role in accounting for motivation. Whatever account is provided by the internalist can be adopted by the externalist, since the task here is a common one: to explain how the agent genuinely acts for a reason. If we accept the plausible claim that agents are passive with respect to the (direct) formation of their desires, it seems most plausible to hold that rational agents are disposed such that their desires are responsive to their beliefs. But the externalist can avail himself of this explanation of rational agency.

Thus, if the agent can come to be motivated to act in response to what he believes he has external reason to do, it seems sufficient to show that he can act for an external reason. Although his motivation is (partly) accounted for by his disposition, his action can nonetheless be correctly *justified* in terms which do not involve his motivational set.⁸¹ As an example, consider our reasons for belief. It is plausible to think that we are disposed to form beliefs in accordance to what we think we ought to believe. When I survey the evidence for some claim and find it to be overwhelming, I simply come to believe that the claim is true. My justification for my belief involves the evidence I have surveyed; no mention of my disposition to form beliefs is needed. Although the formation of my belief is (partly) accounted for by my disposition, I can nonetheless accurately say

⁸¹ The full account of his motivation should include also his external reason belief.

that I believe this claim because of the evidence that I see for it. In a similar way, we can be said to act for external reasons even when our motivation to act is due to our disposition to form desires in response to our beliefs about our reasons for action.

However, Williams thinks this explanation is unsatisfactory:

“But this reply merely puts off the problem. It reapplies the desire and belief model (roughly speaking) of explanation to the actions in question, but using a desire and belief the content of which are in question. What is it that one comes to believe when he comes to believe that there is reason for him to X, if not the proposition, or something that entails the proposition, that if he deliberated rationally, he would be motivated to act appropriately? We were asking how any true proposition could have that content; it cannot help, in answering that, to appeal to a supposed desire which is activated by a belief which has that very content.”⁸²

Williams seems to be charging that such a response is vacuously circular since it makes use of the disputed belief (about some external reason statement) to defend the possibility of external reasons, without first explicating the content of that belief. He thinks the only intelligible content of reason claims is, or entails, that if the agent deliberated rationally, he would be motivated to act appropriately. But this seems to be the internalist’s claim. However, the externalist can understand rational deliberation in a *non-procedural* sense, such that the content of external reason claims is intelligible and

⁸² *IER*, 109-10.

distinct from internalism.⁸³ He can explain that the agent can be *substantively* rational, which involves his coming to be aware of aims that are worth having and responding by being relevantly motivated. That theorist does not claim that the agent would be so motivated upon believing the external reason statement, but only if the agent is substantively rational. Here is where the agent's rational disposition figures: an agent is substantively rational if and only if he is disposed such that he is motivated to Φ when he knows that he has reason to Φ . If the agent can be so disposed—and it is not clear why he cannot—then we can understand how he can arrive, through rational thought, at a motivation which cannot be reached through Williams's procedural means. This enables us to make sense of the content of external reason statements.

One may charge the externalist with postulating an *ad hoc* disposition just so the externalist's notion of motivation may work. However, it should be noted that the internalist is likewise guilty of this charge. The internalist likewise accepts that if there are internal reasons, it must be the case that agents can act for these reasons—their actions cannot merely match what is prescribed by those reasons. As such, even the internalist has to claim that the agent's rational disposition is such that his motivations are responsive to his beliefs about his reasons for action. It may be objected that I have misunderstood the spirit of internalism, for internalists would think of desires as responsive to other desires, not beliefs in themselves. They therefore need not follow the externalists in involving additional dispositions. However, I think they have thereby merely deferred the *ad hoc* charge to the disposition of the agent to form desires in

⁸³ Hooker and Streumer defends such a response in their "Procedural and Substantive Practical Rationality," 70-2. For a different response, see Parfit's "Reason and Motivation," in *The Aristotelian Society*, 120-30.

response to desires (or belief/desire pairs)—a disposition that an externalist account need not involve.

Aside from noting that both parties are allegedly partners in guilt, we should better understand how the *ad hoc* charge is justified. One makes an *ad hoc* claim only when that claim has no independent justification and is motivated only by the need for a view to accommodate data which does not fit with what would be the case if that view is right. I doubt the charge works in this case because there seems to be ample independent evidence for our having such rational dispositions: our first-person experience. Just as we come to believe P upon recognition that we have good reasons to believe P, it seems just as evident that we come to be motivated to Φ upon recognition that we have good reasons to Φ —what constitutes genuine reasons for actions seems to be a separate issue. It seems to me that we already believe that we sometimes act for reasons, before we proceed to examine what the truth-conditions of such reasons are.

I have argued that the crucial step in Williams's argument against external reasons is his defense of (B3). However, his defense seems to involve a question-begging move, where he appears to *assume* that practical deliberation can only be procedural in nature. I then considered and rejected two arguments for (B3) which do not employ that move: the *Direction-of-fit* and the *Desire-out, desire-in* arguments. So far I have focused on whether external reasons can be *motivating* reasons. If I am right, we have no reason to think they cannot be. I shall now consider an argument that external reasons cannot be *normative* reasons.

2.6. The *Unimportance* argument

Stephen Finlay thinks there is no trouble in understanding how external reasons motivate agents. The problem however lies in their status as normative reasons, given that their normative authority is “troublingly *sui generis* and inexplicable”.⁸⁴ Thus in his view, the force of William’s argument lies in exposing this aspect of external reasons: the trouble with externalism is not how external reasons *can* motivate us, but why it *should*. Finlay explains:

“Williams’s demand for explanation acquires traction when it is addressed rather to a normative problem. We may grant that in principle any psychological state might be able to cause any other psychological state. But this applies equally to various kinds of neurological disorders... Reasons aren’t merely things that happen to motivate us: they are also things that ought to do so.”⁸⁵

To properly understand Finlay’s objection, we first have to grasp his account of the *meaning* of reasons, which departs significantly from Williams’s. Finlay thinks the contemporary concept of normativity misleadingly combines two elements: prescriptive content and importance. It is possible for statements to have prescriptive content but lack importance. Finlay thinks institutions like chess and etiquette issue standards in the form of prescriptive statements—statements which “direct us as to what we ‘ought’ to do”.⁸⁶ He argues that institutional ‘oughts’ or reasons should be understood as normative reasons because (i) they are clearly not just motivating reasons since we sometimes are unmotivated by them, and (ii) they involve norms, prescribing various actions. Finlay

⁸⁴ “The Reasons That Matter,” 2.

⁸⁵ Ibid, 3.

⁸⁶ Ibid, 5.

notes that it is common to treat institutional reasons as motivating reasons while denying that they are also normative reasons. However he thinks this is due to the conflation of motivating reasons with motivating states or (causal) explanatory reasons. As he explains, “our psychological states (motivating beliefs and desires) do not motivate us *qua* reasons—they generally are not (and do not figure in) *considerations* that lead us to act. What is normally meant by ‘motivating reasons’ are... normative reasons (answers to questions of the form ‘Why ought I to Φ ?’) that motivate us.”⁸⁷ Motivating reasons are reasons pertaining to normative questions but because we act for those reasons, they can be invoked in causal explanations of why we acted as we did. But to deny that motivating reasons are normative reasons (for which we act) is, in Finlay’s view, to confuse them with their related motivating states or explanatory reasons. Since institutional reasons answer to normative questions, we should regard them as normative reasons. Finlay proposes the following definition of normative reasons (NR):

(NR) R is a normative reason for agent A to Φ iff R would answer a question as to why A ought to Φ or that it would be good that A Φ s.⁸⁸

Finlay thereby extracts the element of importance from the concept of normativity since normativity does not entail importance: “there are reasons that favour actions in ways that do not matter”.⁸⁹

⁸⁷ Ibid, 8.

⁸⁸ Note that on (NR), for a reason statement to qualify as a normative reason, it need only *answer* the normative question of why A ought to Φ or that it would be good that A Φ . There is no requirement that it answers that question *correctly*—it needs only provide an answer of the relevant kind.

⁸⁹ Ibid, 5.

Having argued that the meaning of the term ‘normative reason’ is captured by (NR), Finlay proceeds to advance a claim about the nature of normative reasons which he terms the *Unifying Thesis* (UT):

(UT) A fact is a reason for Φ -ing, relative to a system of ends E, iff it explains why Φ -ing is conducive to E.”

(UT) states that all normative reasons are end-relational. This applies both to the reasons that matter and the ones that do not matter. In support of (UT), Finlay writes:

“Observe that providing a reason for Φ -ing is often explained as demonstrating the ‘point’ of Φ -ing. To furnish an action with a point just is to demonstrate that it serves some purpose or end. Hence for a fact to count as a normative reason for Φ -ing, it must constitute an explanation relative to some end.”⁹⁰

He thinks that if (UT) is true, it would account for normative reasons in purely non-normative terms, thereby providing “a naturalistic reduction of the relation ‘counting in favour of’”.⁹¹ If all reasons are end-relational, then it is possible that some reasons do not relate to *our* ends—the *ends we care about*. This is where the distinction between internal and external reasons comes to the fore, for they differ precisely over how the truth-conditions of reasons relate to the agent’s ends.

⁹⁰ Ibid, 8-9.

⁹¹ Ibid, 8.

But instead of focusing on how it is impossible that external reasons can motivate us, Finlay argues that it is impossible that they should motivate us. Externalists would agree with Finlay that reasons ought to motivate us. This is why we do not consider the agent to lack the reason to Φ just because he *is* unmotivated to Φ : it remains that he *ought* to be motivated. He thinks externalists need to account for why external reasons matter since it is far from obvious that they do. However, it is not clear how that is to be done. Finlay considers what he deems a prominent approach: to account for the importance of external reasons in terms of rationality. Externalists may claim that the reasons that matter are those which would motivate us insofar as we are aware of them and are rational. But such an approach is, in Finlay's view, uninformatively circular since we need then to account for rationality and cannot do so without appeal to reasons that matter. Thus he writes:

“Rational requirements are such that, once we understand them as such, we cannot coherently or intelligibly challenge their authority over us. This seems to be all that is uncontroversial about the concept: but in that case all that it uncontroversially means to say that real reasons are rationally demanding on us is that they are reasons that matter. The real difficulty—concerning what it is to matter—has merely been transposed into the question of what rationality is. Invoking rationality here is thus opaque and unhelpful.”⁹²

This is where Finlay thinks internalism and externalism differ crucially. For while externalists cannot account for the importance of external reasons, sometimes even

⁹² Ibid, 16-7.

attempting to make such reasons important by *stipulation*,⁹³ internalists have little difficulty in this regard:

“The principal virtue of internalist theories of reasons is that they offer an explanation for why we ought to act for such reasons: if Annie’s reasons are considerations that answer to her existing concerns, then (unlike her thoughts about horses) their significance to her practical deliberations is available to her and nonmysterious. In contrast, the normativity of external reasons seems quite inexplicable. Why ought we guide our deliberations and actions by certain facts just because they possess some unexplainable property that does not connect with anything we already care about?”⁹⁴

Finlay therefore thinks *importance internalism* is true; that is to say, reasons are important only if they are motivationally internal, relating to our cares and concerns.⁹⁵

Externalism holds that some reasons are important *simpliciter* and is therefore false since such a claim is inexplicable.

⁹³ “It is part and parcel of Parfit’s nonnaturalism that there cannot be a noncircular explanation of the normative significance of reasons: it is something we cannot explain, yet adequately comprehend. The normative challenge is sometimes countered by definitional fiat—to ask why we ought to do anything just is to ask for reasons for doing it, so the objection is absurd. But to those like me for whom the question seems reasonable when addressed to the nonnaturalist’s account of reasons, this suggests instead that the account must be wrong.” (Ibid, 4.)

⁹⁴ Ibid, 3-4.

⁹⁵ He elaborates: “This is to relativise importance: the claim is not that facts about motivational connections are themselves important simpliciter (a claim no less mysterious than those made by externalists like Parfit, Scanlon, Dancy, and Raz) but rather that there is no such thing as importance simpliciter, there is only importance for particular subjects. There is nothing more to something’s being (intrinsically) important for A than its ‘seeming’ (intrinsically) important to A, and nothing more to this ‘seeming’ than A’s caring intrinsically about it. Our intrinsic concerns are the source of importance for us not by constituting intrinsically important facts, but by making it the case that their objects matter intrinsically to us.” (Ibid, 17.)

It is useful to observe that Finlay's usage of the term 'normative' appears to switch between what he considers the contemporary usage (where the elements of prescriptivity and importance are combined) and his recommended usage (NR). Thus when he writes

“Williams's demand for explanation acquires traction when it is addressed rather to a normative problem”⁹⁶ [emphasis mine]

it should be clear that he is not claiming that Williams's argument shows external reasons cannot be normative reasons in the sense given by (NR). Rather, it seems to me that we should take Finlay to be employing the contemporary use of normativity here and to be pointing out that externalist cannot account for the element of importance within it. Likewise, when he writes

“Reasons aren't merely things that happen to motivate us: they are also things that ought to do so.”⁹⁷ [emphasis mine]

we should take him to be using the term 'ought' in the contemporary sense. After all, Finlay would not be satisfied if externalist merely points to some institutional standard stating that we ought to be motivated by external reasons. Lastly, we should understand Finlay's claim that externalists cannot account for the “normative authority”⁹⁸ as the

⁹⁶ Ibid, 3.

⁹⁷ Ibid, 3.

⁹⁸ Ibid, 2.

claim that they cannot account for the importance of normative reasons, where normativity is employed in his recommended sense.

2.7. Is the *Unimportance* argument sound?

We can formulate Finlay's *Unimportance* argument as follows:

- (D1) All reasons are end-relational. [UT]
- (D2) The only ends that matter to the agent are ends which relate to his intrinsic concerns.
- (D3) All reasons matter to the agent only if they relate to his intrinsic concerns.
[D1, D2]
- (D4) External reasons are independent of the agent's intrinsic concerns.

Therefore,

- (D5) External reasons do not matter to the agent. [D3, D4]

Finlay's defense of (UT) seems rather thin. Externalists deny that all reasons are end-relational and would thus reject (UT). They would claim instead that we have reason to pursue some ends, not as means to some other end, but for their own sake. Those who accept (UT) cannot make such a claim. An additional issue is whether facts, understood as non-normative facts, can be identical to (or constitutive of) reasons for action even if they relate means to ends. It seems to me that (UT) combines two distinct issues since

one can hold that all reasons are end-relational while rejecting that reasons can be understood in non-normative terms. I shall steer clear of the second issue in this part of my essay and focus only on the first. This is because my present concern is whether Finlay is right that external reasons do not matter to the agent and this task does not require me to argue that external reasons cannot be explained in non-normative terms. Put another way, even if Finlay is right that normative reasons can be explained non-normatively, this applies to both internal and external reasons, and does not serve to distinguish between them.

Finlay's argument for (UT) is that we often take the giving of some reason for action as explaining the 'point' of that action. An action with a 'point' just is one which serves an end. He then concludes that normative reasons for Φ -ing "must constitute an explanation relative to some end".⁹⁹ This conclusion is too swift. Firstly, even if we often take providing the reasons for Φ -ing as demonstrating the 'point' of Φ -ing, it may simply show that we *often* act for instrumental reasons. But it does not follow that *all* reasons are instrumental. The claim that all normative reasons for Φ -ing are end-relational is a *conceptual* claim about such reasons and must therefore show that reasons which are not end-relational are conceptually impossible. But such reasons seem at least conceptually possible, as I shall explain later.

Secondly, it is not clear that we must understand the 'point' of some action as the end it serves. We can instead explain it as the justification of that action, serving to differentiate it from capricious actions which are 'pointless'. We have not yet been given reason to think that justifications for actions must be in the form of showing that those actions serve certain ends. Consider this example: Theresa is offered a million dollars to

⁹⁹ Ibid, 8-9.

help in the trafficking of child sex-workers. However, she rejects the offer and explains: “I will not help in such an activity because it is evil!” Alternatively, suppose she says: “I will not help in such an activity because it denigrates children and causes the innocent great suffering!” It seems that both explanations are reasons for her refusal of the offer and do not seem relative to some end. The reasons involve only references to the facts about (or intrinsic disvalue of) the actions and the consequences, and yet they justify Theresa’s refusal. Perhaps one may attempt to interpret them as relative to some implicit end but surely it is far from clear that we must understand them in that way. Given these two considerations, I think Finlay’s observation does not furnish us with compelling reason to think that all normative reasons are end-relational.

On the contrary, there is a consideration which suggests that not all normative reasons are end-relational: it seems that sometimes we seek for and recognize our reasons *all-things-considered*. Finlay thinks such reasons can be understood as our reasons “[a]ll my ends considered”.¹⁰⁰ That may be the case but how can the view that all reasons are end-relational account for reasons all-my-ends-considered? On that view, even this class of reasons must be end-relational.¹⁰¹ But how should we understand *an* end to capture the notion of *all-my-ends-considered*? It seems that either a reason is end-relational or it is not. If it is not, then the end-relational view of reasons is false. If it is, then it is one end among my many ends. As such, it is weighing the reasons given by various ends with reference to some end, namely, its own. The criterion for assessing the comparative weight of reasons given by different ends is itself end-relative. Yet in considering my all-

¹⁰⁰ Ibid, 11.

¹⁰¹ As he writes: “That there can be normative deliberation and decision altogether independent of the point of view of particular ends I simply deny, and I don’t believe I am overlooking any significant evidence.” (Ibid, 11.)

ends-considered reasons, I should be evaluating *all* my relevant ends in a manner which is not loaded towards some end—otherwise, the evaluation is not neutral with respect to all ends. It thus seems impossible for all-ends-considered reasons to be themselves end-relational. For if they were end-relational, how can they involve evaluating all ends, including their own end? End-relational reasons cannot transcend the end that they are relative to, in order to evaluate the weightiness of the reasons given by their own end. To assess the reason given by some end, a criterion which is independent of that end is needed, or else the criterion is assessed using only itself—this seems question-begging.

Consider the example of three people—A, B and C—assessing how a pot of gold should be shared among them. Let us assume that each can only make the assessment relative to his own interest. C claims to provide an assessment which properly adjudicates between all their interests. But recall that C's assessment is likewise one which is relative to his own interest. If so, how can C provide an assessment which properly adjudicates between all their interests? In what sense is his assessment more valid than those given by A and B in accordance to their own interests? It should not matter what C's interest is, for the point is that he is not in a position to assess the weight of his own interest insofar as he is making that assessment relative to his own interest.

In this way, the end-relational view of reasons makes reasons pertaining to each of my ends incommensurable with those of my other ends. If my all-ends-considered reason is itself end-relative, then it can consider the relative weight of all ends *except itself*. Yet it needs to consider *all* ends. The end-relational view of reasons therefore cannot account for all-things-considered reasons. However, in our deliberation, we weigh end-relational reasons against one another. Thus, for instance, I may conclude that

prudential considerations are outweighed by moral ones. This indicates the presence of considerations of reason which are independent of ends.

(D1) therefore seems poorly supported. However, the force of Finlay's argument can be retained even without it. For even if it has not been shown that all reasons relate to ends, one may still ask of the externalist why non-end-related reasons have any importance in the same way it is asked why there is importance to reasons unrelated to the ends I care about. At this point, note that Parfit does not attempt to account for the importance of external reasons in terms of rationality, contrary to Finlay's charge.¹⁰² On the contrary, Parfit thinks externalists would account for rationality in terms of reasons, unlike internalists.¹⁰³ How then would externalists like Parfit explain the difference between important and unimportant reasons? I think they would take a path broadly similar to Finlay's.

Finlay claims that importance arises from our intrinsic cares and concerns. For him, the difference between important and unimportant reasons is simply that only the former are based on (or relate to) what is importance-giving (i.e. our intrinsic cares and concerns). Parfit claims that importance arises from facts and properties. Thus for him, that difference is simply that only important reasons are based on facts and properties that are importance-giving. Finlay may say that merely defers the problem, for Parfit then has to explain why some facts and properties are importance-giving while others are not. I am not sure if this is a fair demand, for it seems to imply that for an explanation to be adequate, we must first be able to explain the *explanans* (what explains some phenomenon) and perhaps also higher and higher orders of *explanantia*. In any case,

¹⁰² See p. 76-7 of this essay.

¹⁰³ See Parfit's "Reason and Motivation," 101-2.

Finlay does not escape the same demand, if it is a legitimate one. For we may ask: why are our intrinsic cares and concerns importance-giving while other possible sources (like our kitchenware or the concerns of our enemies or our hypothetical selves) are not?

Perhaps Finlay's point is not (just) that externalists fail to provide a criterion of importance, but that the criteria they propose are evidently implausible—unlike the internalist's proposal. Thus Finlay asks why we should “guide our deliberations and actions by certain facts just because they possess some unexplainable property that does not connect with anything we already care about”. He thinks it is non-mysterious that we should be guided in our deliberation by considerations that answer to our existing concerns, but mysterious otherwise.

There are two possible senses of mystery here and one has already been shown to be without force—at least without force applied only to externalists. For it seems Finlay thinks the agent's existing concerns are fittingly deliberation-guiding because those concerns are importance-giving, and to engage in practical deliberation is to weigh the importance of possible courses of action. But it is available to Parfit to say the same, even if he identifies a different source of importance: *if* the facts and properties identified by Parfit are importance-giving, then it is not mysterious why we ought to guide our deliberation by them, even if they do not connect with anything we already care about. In that case, even if we do not care about what is importance-giving, we *should* care.

The remaining sense of mystery is this: it is obvious that Finlay's criterion of importance is right while it is mysterious how we should think the same of the externalists's (and therefore it is mysterious why their criteria should figure in our deliberation). The dispute is then over the substantive issue of which criterion of

importance is right, rather than conceptual issue of what importance is. If so, Finlay has to do more than to claim that the internalism is obviously true: it is not *obviously* true.

I conclude that we should not accept importance-internalism. Moreover, I shall not reformulate the internalism/externalism debate in terms of importance as Finlay has done. This is because I think we should avoid employing his conception of normative reasons as it unnecessarily complicates the discussion by introducing an additional distinction, i.e. between important and unimportant reasons. It is unnecessary because even when we avoid conflating motivating reasons with motivating states (or explanatory reasons), there is still no reason to think that if a reason statement *R* is a motivating reason, then *R* is also a normative reason in virtue of its prescriptive content. After all, we are sometimes motivated to act in part because of our beliefs that there are certain *facts* (such as “there is a fruit in front of me”, or some means-ends relation). But surely we should not take motivating facts—‘facts by which we are motivated’—to *be* facts just because they have *descriptive* content, and thereafter distinguish between facts that matter and do not matter. We can simply treat motivating facts as ‘*what* we believe to be facts and act for’ and still avoid conflating such propositional content with motivating states; we can take motivating states in such cases to be beliefs with descriptive content, e.g. the belief that *P* is a fact.

What may mislead us into taking motivating reasons to be also normative reasons is the claim that motivating reasons are the normative reasons we act for. It seems to me that this claim is either loosely-worded or simply false. It is more accurate to explain motivating reasons as ‘*what* we believe to be normative reasons and for which we act’. If the reason for which we act *is* a normative reason, then by the same token, the fact for

which we act is a fact. Surely that is false. A statement does not become a fact just because it answers to a factual question. To avoid unnecessarily expanding our conceptual vocabulary, I shall abide by what Finlay terms as the contemporary concept of normativity. This has the advantage of being congruent with the usage of most involved in the internalism/externalism debate. My assessment of the *Unimportance* argument shall therefore be couched in contemporary parlance: it fails to show that external reasons cannot be normative reasons.

Part 3: Reasons Subjectivism and Objectivism

3.1. Taking stock

In the preceding two parts, I have tried to defend these two claims:

- (1) Reasons internalism is false.

- (2) Reasons externalism is not conceptually unsound.

In Part 1, I argued that Williams's distinction between internalism and externalism is of limited use because internalism is clearly false. I explained that internalists face a dilemma: internalism entails highly-counterintuitive outcomes which threatens its plausibility, such as making it merely a *contingent* truth that an agent who acts out of false beliefs is thereby acting against his best reasons. Yet, the only way for internalism to avoid such outcomes is for some form of externalism to be true. But this renders internalism false, since it denies the truth of externalism—(1) is true.

In Part 2, I considered Williams's argument against the conceptual soundness of external reasons. The force of this argument lies in its charge that externalism faces a motivational problem: it cannot explain how we *can* be motivated by reasons, since motivation requires desires and externalism needs us to acquire desires the objects of which are unrelated to those of our actual desires. I argued that this charge is without adequate justification because (i) we have no reason to think that desires of the rational agent cannot be responsive to his beliefs or that (ii) the agent's motivating reasons must

involve his antecedent desires. I then considered Finlay's interpretation of Williams's argument which centers on showing how externalism faces a normative problem: it cannot account for why we *should* be motivated by reasons, since those reasons do not relate to our intrinsic cares and concerns. However, as I argued, externalism faces a normative problem only if normative reasons¹⁰⁴ must have the relation he claims. Finlay has not provided compelling support for that claim and thus, it remains possible that we should be motivated by reasons which are unrelated to our intrinsic cares and concerns. We have no reason to think externalism faces a normative problem. I then concluded that Williams's argument has not shown reasons externalism to be conceptually unsound—(2) is true.

If I have successfully defended (1) and (2), then we should accept that either reasons externalism is conceptually sound, or that there are no reasons for action. The issue therefore concerns whether there are reasons for action. However, this is far from the only crucial issue with regards to reasons for action, for there is an important division *within* reasons externalism. This division is engendered by the role of desires in explaining reasons. One view accords a central role to desire, such that the agent's reasons for action are all *indirectly* determined by their relevantly-informed desires.¹⁰⁵ The contrary view denies desires such a significant role. Defenders of this view claim that the agent's reasons are determined by facts and properties,¹⁰⁶ or values,¹⁰⁷ which need not be related to the agent's antecedent desires. I shall now formulate a distinction between these two views, before proceeding to explain its importance.

¹⁰⁴ In Finlay's terms, the reasons that matter to the agent.

¹⁰⁵ I shall explain the nature of this determination later. Proponents of this view include Dreier, 1997, and Sobel 2001.

¹⁰⁶ Proponents of this view include Parfit, 1997, 2006.

¹⁰⁷ Proponents of this view include Adams, 1999.

3.2. Two views of reasons

We can better grasp the distinction between the two views of reasons by first noting their commonality. Defenders of both views agree that the agent can have a reason for action which is *not* contingent on his motivational set, thereby making them united in embracing reasons externalism. Their disagreement lies in what they take to be the *content* of reason claims. Let us call Reason Subjectivists those who hold that the agent has reason only to satisfy his relevantly-informed desires in the most efficient manner. Although this reason obtains independently of the agent's motivations, it has only formal content since the specific actions which the agent has reason to perform depends on the objects of his relevantly-informed desires. On the other hand, let us call Reason Objectivists those who claim that the agent has reasons for action which are not limited to (or do not include) the efficient satisfaction of his relevantly-informed desires. Let us take Reason Subjectivism (RS) as

(RS) A has only reasons to satisfy his relevantly-informed desires in the most efficient manner.

Reason Objectivism can be explained as

(RO) A has reasons to act which are not (or not only) reasons to satisfy his relevantly-informed desires in the most efficient manner.

The key difference between (RS) and (RO) concerns the scope of practical reason.¹⁰⁸ If (RS) is true, the reach of reason in the practical domain of human agency is constrained by our desires. This has clear relevance to morality, as noted by several writers.¹⁰⁹ If (RS) is true, then reasons with distinctively moral content are true only of agents whose relevant desires are such that acting morally promotes their satisfaction. This leaves open the possibility of agents who have no reason to act morally. However this seems to conflict with the authority of moral *requirements*—it seems that all agents have *overriding* reason to abide by moral requirements, which apply universally. This should not be taken to mean that all moral *reasons* are overriding, but only that moral *requirements* are—not all moral reasons are moral requirements. Agents can have reasons which override even moral reasons. For instance, I have a moral reason to not to steal. But suppose that I would lose my life unless I stole a trivial item from a rich person (who would suffer no significant harm in that regard). In that case, it seems implausible to claim I have an overriding moral reason not to steal, even at the expense of my life. Concerning moral requirements, David McNaughton writes:

¹⁰⁸ In drawing the distinction between Subjectivism and Objectivism, rather than Subjectivism and non-Subjectivism, I am leaving aside those who are both non-Subjectivists and non-Objectivists about reasons, i.e. those who hold that reasons are conceptually impossible. It seems to me that the former is the most significant distinction since there are few who are non-Subjectivists and yet not Objectivists.

¹⁰⁹ Thus Talbot Brewer writes: “The upshot of the internalist position is that no reason for action applies to any person regardless of that person’s subjective motivational states. If part of what we mean when we call a reason moral is that the reason applies to us regardless of our subjective motivational states, then internalism implies that there are no moral reasons.” (“The Real Problem with Internalism about Reasons,” in *Canadian Journal of Philosophy*, 443) John McDowell makes a similar point: “The question is quite abstract and general, but it is obvious that it bears on a familiar problem that arises about ethical reasons in particular, in view of the evident possibility of being left cold by them. The implication of Williams’s skepticism is that ethical reasons are reasons only for those for whom they are internal reasons: only for those who have motivations to which ethical considerations speak, or can be made to speak.” (“Might There be External Reasons,” 95)

“What is distinctive of moral requirements is that they are thought of as providing a reason to act which outweighs or overrides any reason the agent may have to act in some other way. Moral requirements are thus seen as independent of desire in the further sense that they have a claim on our obedience that is not conditional on there being nothing else which we want more. In other words, once an agent has formed a conception of the situation in which he takes it that morality requires an action, then the thought that there might be some other action which he has more reason to do is no longer available to him.”¹¹⁰

It seems perfectly plausible to claim that the rapist ought not seek another victim, without first consulting the desires that the rapist has—the fact that such an action is morally wrong seems sufficient to make that reason statement true. If he continues to rape, then it is appropriate to hold him morally blameworthy.

¹¹⁰ *Moral Vision*, 114-5.

However the notion of blame seems to carry with it the belief that the blameworthy person did what he knew he ought not do. However, if (RS) is true, our common practices of ascribing moral reasons and blame either appears terribly presumptuous or requires a revisionary understanding.¹¹¹ The force of morality is thereby held hostage to our desires.¹¹² The distinction between Reason Subjectivism and Objectivism is therefore of much import, even if both views are externalist in nature.¹¹³

By focusing on the Subjectivism/Objectivism distinction, where the dispute is clarified as involving the content rather than motivational power of reason, we can be clearer about the striking similarities between the two dominant views of practical reason. This in turn helps us see the sorts of argument which would *not* directly advance one view over the other. For instance, since both views accept some form of normativity and differ only with respect to what they identify as the sources of normativity, arguments from queerness would affect them both. It is hard to see why one source would be queerer than the other, although it would be hasty to rule out arguments to such

¹¹¹ This is not to say that Subjectivists are without responses. For instance, Williams suggests that ascription of external reason statements to others can be thought of as optimistic internal reason claims which are ascribed to them. This is done in hope that some relevant motivation (perhaps out of a desire to be socially accepted) will be engendered in them in virtue of their awareness of our judgment, such that they would perform the relevant action. (“Internal Reasons and the Obscurity of Blame,” 40) Finlay claims that when we ascribe reasons to others, it could be motivationally internal *to us*, rather than those we are making reason judgments about. This would indicate to others that it matters to us that they act on those reasons. (See his “The Reasons that Matter,” 18) See also the discussion on moral criticism (which I take to be similar to moral blame) at http://peasoup.typepad.com/peasoup/2006/11/subjectivism_an.html

¹¹² It does not seem enough to claim that all agents *in fact* have desires upon which moral reasons can be based, for even this does not reflect the authoritative nature of morality. It seems a conceptual truth that moral requirements are overriding because a person is not exempt from them simply by proving that he lacks the relevant desire. Since it appears logically possible that an agent lacks the relevant desire, it seems that the authority of moral requirements do not derive from some desire of the agent’s.

¹¹³ This tension between the nature and scope of moral requirements, and (RS) is termed by Finlay and Schroeder as the “Central Problem” of recent ethical theorizing. They observe that this close link between practical reason and moral requirements makes the internalism/externalism debate of such philosophical significance. This is because the debate concerns the limit of practical reason, expressed in terms of the necessary conditions of reasons for action, and therefore also the limit of moral requirements. As I have explained earlier, this remains a live issue on the Subjectivism/Objectivism distinction. (See Finlay and Schroeder’s “Reasons for action: Internal vs. External”.)

conclusions. Thus, queerness arguments may *indirectly* convince us to prefer one of the two views, but it would have to work in conjunction with some further argument which dislodges the apparent parity between them with regards to normativity.

However without such additional arguments, one apparent attraction of Subjectivism may fade. Here I am referring to the view that Subjectivism avoids the metaphysical and epistemological commitments which Objectivism requires. Because Subjectivism ties reasons for action to the satisfaction of desires, it may seem like reasons are nothing more than empirical facts about the agent's psychology. Robert Adams notes this point with regards to theories of wellbeing, but it applies as well to theories of practical reason:

“Another apparent advantage of a desire-satisfaction theory is the fact that it offers a reduction of the notion of a person's good to empirical facts about preferences. As Sidgwick puts it, it is supposed to be a notion of good that ‘is entirely interpretable in terms of fact, actual or hypothetical, and does not introduce any judgment of value fundamentally distinct from judgments relating to existence; —still less any ‘dictate of Reason’.’”¹¹⁴

What is an empirical matter for Reasons Subjectivism is the content of true reason statements, and even that content is but partly empirical. The agent's psychology does not determine in itself what the agent's reasons for action are; it is of relevance only because certain elements of his psychology figure in the content of some ‘dictate of Reason’. It is far less clear that reasons are *identical to* rather than merely *dependent on* empirical

¹¹⁴ *Finite and Infinite Goods*, 86. Here Adams cites Sidgwick, *Methods of Ethics*, 112.

facts—at least that is not entailed by or made more plausible by Reasons Subjectivism. The attractions of Subjectivism must therefore come from other considerations and not from its involving less queerness than objectivism.

The Subjectivism/Objectivism distinction also indicates that arguments based on motivational considerations do not directly confer an advantage to either view. This may not seem obvious given that Reasons Subjectivism, unlike Objectivism, accords a central role to desires. However, desires figure in Subjectivism not by motivating agents to act, but by making reason claims true. Desires are the source of normativity for Subjectivists, not the ultimate sources of motivation. Both Subjectivists and Objectivists accept that reason can motivate agents without the aid of their antecedent desires. In view of this consideration, comparative evaluations of Reason Subjectivism and Objectivism should steer away from focusing on motivational issues, unless some relevant difference can be found between the two views which are otherwise on a par.

3.3. Skepticism about Motivational arguments

Subjectivists may nonetheless be tempted to take the relevant difference to lie in the manner in which they, unlike Objectivists, relate reasons to motivation. But herein lies the danger. When desires are seen as the exclusive source of motivation for action rather than (or in addition to being) the source of normativity, there is a tendency to run motivational arguments in support of Subjectivism. By a motivational argument, I mean an argument based on how antecedent desires explain the motivational effect of reason. The danger for Subjectivists is that antecedent desires are made to do all the motivational work, with reason assuming a superficial role. Allow me to explain.

Motivational arguments typically claim that all motivations derive from antecedent desires. Since externalism requires motivations which do not derive from antecedent desires, it is therefore false. Here we may ask: how do antecedent desires work in motivating the agent to act for reasons? One reply may be as follows: suppose I desire E. Reason reveals that the means to E is M. In virtue of my desire for E, I am motivated to bring about E. But how is that motivation directed towards M? The subjectivist can follow Finlay in holding that it is a conceptual truth that you would desire M, if you desire E and believe that the means to E is M. But such a transfer of motivation has no role for reason: it is wholly explained by the nature of desire. While it is true that reason reveals what the means is, it remains that when we act we cannot claim that we are acting for *that* reason since that is a reason for *belief*, not *action*.

The Subjectivist can claim that reason does more than reveal what our means are; it also tells us to take the means to our ends. But what happens when the agent recognizes his reasons? The Subjectivist may reply that the agent is disposed such that when he believes he has a reason to bring about M, his motivation to bring about E (in virtue of his desire for E) is channeled into the relevant direction, resulting in his motivation to bring about M. This account of motivation is strikingly similar to some Objectivist accounts. Some Objectivists hold that the agent's rational disposition is such that he comes to desire E when he believes he has reason to E. Such Subjectivist and Objectivist accounts both invoke dispositions and understand the link between the agent's belief about his reason and his corresponding motivation to be *causal*. The key difference is that some Subjectivists may require the presence of some antecedent desire to make possible the motivational transfer. It is difficult to see why this is or must be the case.

Motivational arguments therefore focus on why certain causal relations are more plausible than others when it comes to the motivation of agents. I do not find compelling reason either to prefer some causal relations over others in this regard, or to think that the prospects of justifying such a preference are promising.

In fact, I do not think any causal account properly captures rational agency. In Part 2, I have assumed for the sake of argument that

(B1) It is possible for A to Φ only if A desires to Φ .

I tried to argue that *even if* (B1) is true, Williams's argument is unconvincing because the agent who believes he has an external reason to act can come to have the relevant desire through his rational disposition. While it is conceptually true that the agent cannot be motivated *except* by the involvement of his existing motivational abilities, it does not follow that he is therefore deliberating from his antecedent motivations. This is because the truth condition of the external reason for which he is acting is independent of his motivational set. His motivation is (partly) accounted for by his disposition, but his action can nonetheless be correctly *justified* in terms which do not involve his motivational set. This is sufficient to demonstrate the possibility of reason motivating the agent without the aid of his antecedent desires. Without (B1), it is harder for Williams's argument to work since no desire, let alone an antecedent desire, is needed for an agent to be motivated to act for a reason. It is therefore more difficult to see why the agent's motivation to act must relate to his antecedent desires.

The problem with causal accounts of rational agency is that it merely shows how the agent's action matches what he believes he has reason to do. However, it does not capture how that match is due to the agent's acting *for* that reason. To illustrate this point, consider Christine Korsgaard's use of an example from Kant to explain how a being can engage in reasoning and yet not be motivated by it. She writes:

“He [i.e. Kant] is talking about what the world would have been like if nature had had our happiness as her end. Our actions would have been controlled entirely by instincts designed to secure our happiness, and ‘... if, over and above this, reason should have been granted to the favoured creature, it would have served only to let it contemplate the happy constitution of its nature.’ The favoured creature is portrayed as able to see that his actions are rational in the sense that they promote the means to his end (happiness); but he is not motivated by their reasonableness; he acts from instinct. Reason allows him to admire the rational appearances of what he does, but this is not what gets him to do it—he has the sort of attitude towards all his behaviour that we in fact might have toward the involuntary well-functioning of our bodies.”¹¹⁵

Here, we can imagine that God has arranged for that creature's instincts to be such that they result in his always doing what he has most reason to do. We may say that God is motivated in doing so by the reasonableness of such a pattern of behaviour. In this way, considerations of reasonableness figure *indirectly* in what the creature does even if he

¹¹⁵ “Skepticism about Practical Reason,” 110-1. Here Korsgaard cites Kant's *Groundwork of the Metaphysics of Morals*, 395.

never acts for those considerations. It is clear, however, that the creature is not acting for reasons, although he recognizes those reasons and his deeds exactly match what he has most reasons to do.

In a similar way, it seems merely a happy coincidence that the agent is disposed such that he is motivated in accordance with his beliefs about his reasons. His beliefs figure indirectly in explaining his deeds just as considerations of reasonableness figure indirectly in the deeds of Kant's creature. However, the manner in which they figure does not suffice for us to accept that they acted *for* those reasons, and that their deeds are instances of rational agency. In both cases, the relevant behaviour is caused by some mechanism which is engineered such that the output is a specific function of the input.

Consider the opposite case: Agent A is disposed such that he responds to what he believes he has most reason *not* to have, by having the desire for it. Thus A believes he has most reason not to consume the packet of heroin but because he has that belief, A desires to and consumes the heroin. Is it therefore the case that A's belief (that he has most reason not to consume the heroin) explains his behaviour? It appears implausible. But why? I believe this is because the belief's motivational effect is entirely derivative from A's disposition to respond to it. This remains so even if the belief appears to have a more rational link to the desire, as in the case of an agent desiring to do what he believes he has most reason to do.

In these cases, it is hard to see how the agent is able to act for reasons when his make-up determines his responses to the considerations that he recognizes as reasons.

Given

(A1) If there is a reason for A to Φ , it must be possible for A to Φ for that reason.

it is difficult to even see how we can take such agents to have reasons for action: what he does *cannot* motivated by reason. One may respond that it is sufficient for rational agency that reasons cause the agent's actions. When he is disposed to desire what he believes he has reason to do, it is (his believed) reason that causes what he does and this may be sufficient to consider him a rational agent. However such a response would also consider Kant's creature rational since God could have set him up to do what he believes he has most reason to do. Is it not the case then that reason causes what the creature does? I believe that we should not take either case to demonstrate rational agency. This is because what the agent does is caused by his *state* of believing rather than the *content* of his belief. What may be misleading is how his belief about his reasons plays a role in his behaviour. The issue is however whether it plays the appropriate role and that is doubtful.

Instead of causal accounts of rational agency where the agent's motivation is passively brought about, I think we need an account which allows the agent an active role in bringing out his motivation. To this end, it seems most plausible that we understand the agent's motivation as his intention to act, which is directly brought about by the agent through the exercise of his power to form intentions. When he acts for a reason, the content of that reason is part of his intention: at time T, he directly brings about the intention to Φ for reason R. This is in accord with Jay Wallace's Volitionist account of rational agency, where the agent "is guided by their conception of their reasons when that conception is reflected in the content of the intention on which they act; in that case, one

will be able to understand what the agent is doing only by grasping what speaks in favour of so acting, from the agent's own point of view."¹¹⁶ The strength of the Volitionist account is that it makes sense of how the agent's comprehension of his reasons is guiding his action, since his intention combines both his self-understanding of his action (i.e. why he acts as he does) and his control over what he does.¹¹⁷

It is not my purpose in this essay to defend in detail an account of rational agency. I wish only to explain my doubts about causal accounts and briefly explain an account which I consider more plausible. This is in support of my claim that Motivational arguments for Subjectivism have dim prospects since they tend to imply a causal account of rational agency, which I have argued that we should reject.

3.4. Content arguments

The debate between Subjectivists and Objectivists should therefore center on how well they account for the content of true reason statements. Obviously there is no clear agreement on *that* content: virtually any claim concerning the truth of a reason statement requires defense. However we can at least narrow the scope of disagreement such that Motivational and Queerness arguments are not longer obviously applicable, given that the Subjectivism/Objectivism debate takes for granted that normative reasons can motivate the agent independently of his antecedent desires. The kind of argument that is of clear relevance is what I shall term Content arguments—arguments based on how a view is superior in getting the most plausible results in cases regarding an agent's reasons

¹¹⁶ "Three Conceptions of Rational Agency," 60.

¹¹⁷ This account does not imply that we have direct power over our desires. It is compatible with our understanding that we have no direct control over what we desire; it claims only that motivation *need not* derive from desires.

for action. I shall now briefly explain, without evaluating, some Content arguments to provide a clearer view of how they operate.

Derek Parfit's arguments against Reasons Subjectivism (or what he terms desire-based theories of practical reason) are Content arguments. In his argument based on "future Tuesday indifference," Parfit considers a person who cares about his pains and pleasures except when they occur on any future Tuesday. This person therefore prefers severe agony on a future Tuesday to slight agony on a future non-Tuesday—given the choice, he would settle for the former. Since it is possible that his "future Tuesday indifference" would remain even after informed deliberation, Parfit argues that Subjectivists have to accept that this person has no reason choose slight agony on a future non-Tuesday over severe agony on a future Tuesday. However, it is obvious that he—and anyone else—has a reason to make such a choice. While Subjectivists have to accept what Parfit considers an absurdity, Objectivists have no difficulty in accounting for what we think the person has reason to choose, given that they do not take the truth of reason statements to be contingent on the agent's desires. Thus in Parfit's view, Reasons Subjectivism is less plausible than Objectivism, since the former is unable to account for an obvious truth unlike the latter.

Content arguments can be formulated in a less direct manner. Objectivists may argue that practices which we consider well-justified make sense only when certain reason statements are true. However, Subjectivism cannot account for these truths and therefore cannot make sense of such practices. Subjectivism is thus less plausible than

Objectivism since only the latter has such explanatory power. Arguments from blame are instances of such Content arguments.¹¹⁸ Here is one version:

- (1) An agent is blameworthy only if he does what he has overriding reason not to do.¹¹⁹
- (2) All agents who engage in recreational baby-torture are blameworthy, only if all agents have overriding reason not to engage in recreational baby-torture.
- (3) If Reasons Subjectivism is true, then it is false that all agents have overriding reason not to engage in recreational baby-torture.
- (4) All agents who engage in recreational baby-torture are blameworthy.
- (5) Therefore, Reasons Subjectivism is false.

The overriding reason not to engage in recreational baby-torture is of a moral nature. However, if Reasons Subjectivism is true, then agents who lack any desire which would be promoted by acting morally have no reason not to engage in baby-torture. Given (1) and (2), such agents would therefore not be blameworthy. But as the argument goes, this seems implausible. We think people ought to be blamed if they tortured babies for fun, even if they prove that they lack the relevant desires not to do so.

¹¹⁸ Russ Shafer-Landau presents such an argument in his *Moral Realism*, 187-8.

¹¹⁹ Some may consider a person blameworthy only if he does what *he knows* he has overriding reason not to do. For my present purpose, (2) is sufficient although it can work just as well if we require the agent's knowledge of what he has overriding reason not to do. This is because it seems possible that an agent does what he knows he has overriding reason not to do, and we need only to change the specification of the case considered.

I shall not evaluate this argument but only use it to point out how arguments based on practices which we consider well-grounded can ultimately be a form of Content argument. But if Content arguments ultimately fail to decisively favour a theory over its alternatives, we may then have to look to features of the theories to evaluate their comparative merit. For instance, if two theories have similar explanatory power, we then have reason to consider the simpler theory more epistemically preferable. Here is where considerations of queerness are important: the theory with metaphysical commitments which are more in accord with our background knowledge is epistemically superior, in virtue of its ontological simplicity.

In summary, even aside from Motivational and Queerness arguments, there remain many points of dispute between Subjectivists and Objectivists. The primary aim of my essay is to argue that it is more fruitful to focus on those disputes instead of motivational and queerness issues. To this end, it is helpful to shift our attention away from the internalism/externalism debate because that distinction centers on those very issues where there is, or should be, agreement.

-End-

Bibliography

Anscombe, E. *Intention*, 2nd Ed. Cornell University Press, 1963: Ithaca.

Brewer, T. "The Real Problem with Internalism about Reasons," in *Canadian Journal of Philosophy*, vol. 32 (2002), 443-74.

Finlay, S. "Motivation to the Means," in *Moral Psychology Today: Essays on Values, Rational Choice, and the Will*, ed. Chan, D. Springer, 2008: Netherlands.

Finlay, S. "The Reasons that Matter," in *Australasian Journal of Philosophy*, vol. 84 (1), 1-20.

Finlay, S. and Schroeder, M. "Reasons for Action: Internal vs. External," in *Stanford Encyclopedia of Philosophy*, ed. Zalta, E. 2008. URL:
<http://plato.stanford.edu/entries/reasons-internal-external/>

Hooker, B. and Streumer, B. "Procedural and Substantive Practical Rationality," in ed. Mele, A. and Rawling, P. *The Oxford Handbook of Rationality*. Oxford University Press, 2004: Oxford.

Hooker, B. "Williams' argument against external reasons," in ed. Millgram, E. *Varieties of Practical Reasoning*. MIT, 2001: Cambridge, MA: 99-101.

Jay, Wallace, R. "Practical Reason," in *Stanford Encyclopedia of Philosophy*, ed. Zalta, E. 2003. URL: <http://setis.library.usyd.edu.au/stanford/entries/practical-reason/>

Jay Wallace, R. "How to Argue about Practical Reason," in *Normativity and the Will*. Clarendon Press, 2006: Oxford: 15-42.

Jay Wallace, R. "Three Conceptions of Rational Agency," in *Normativity and the Will*. Clarendon Press, 2006: Oxford: 43-62.

Korsgaard, C. "Skepticism about Practical Reason," in *Moral Discourse and Practice*, ed. Darwall, S., Gibbard, A. and Railton, P. Oxford University Press, 1997: USA, 373-88.

Korsgaard, C. "The Normativity of Instrumental Reason," in *Ethics and Practical Reason*, ed. Cullity, G. and Gaut, B. Clarendon, 1997: Oxford.

McDowell, J. "Might There be External Reasons?" in *Mind, Value, and Reality*. Harvard University Press, 1998: Cambridge, MA: 95-111.

McNaughton, D. *Moral Vision*. Blackwell, 1988: Oxford.

Nagel, T. *The View From Nowhere*. Oxford University Press, 1986, New York.

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

Parfit, D. *Climbing the Mountain*. Unpublished. URL:

<http://individual.utoronto.ca/stafforini/parfit/parfit - climbing the mountain.pdf>

Parfit, D. "Reasons and Motivation," in *The Aristotelian Society* (suppl. vol.), vol. 71 (1997), 99-130.

Pruss, A. "Means-End Reasoning," in *Alexander Pruss's Blog*. 2008. URL:

<http://alexanderpruss.blogspot.com/2008/08/means-end-reasoning.html>

Shafer-Landau, R. *Moral Realism: A Defence*. Oxford University Press, 2003: Oxford.

Smith, M. "The Humean Theory of Motivation," in *Mind*, vol. 96 (381), 36-61.

Sobel, D. "Subjectivism and Moral Criticism," in *PEA Soup*. 2008. URL:

http://peasoup.typepad.com/peasoup/2006/11/subjectivism_an.html

Sobel, D. "Desires vs. the Reduction Basis of Desires," in *PEA Soup*. 2005. URL:

http://peasoup.typepad.com/peasoup/2005/01/desires_vs_the_.html

Sobel, D. "Subjective Accounts of Reasons for Action," in *Ethics*, vol. 111 (2001), 461-92.

Williams, B. "Postscript: Some Further Notes on Internal and External Reasons," in ed.

Millgram, E. *Varieties of Practical Reasoning*. MIT, 2001: Cambridge, MA: 91-7.

Williams, B. "Internal Reasons and the Obscurity of Blame," in *Making Sense of Humanity*. Cambridge University Press, 1995: Cambridge.

Williams, B. "Internal and External Reasons," in *Moral Luck*. Cambridge University Press, 1981: Cambridge: 101-13.