The Particular and the General: Essays at the Interface of Ethics and Epistemology

by

A. Selim Berker

A.B. Physics
Harvard University, 1998

A.M. Physics
Harvard University, 2000

Submitted to the Department of Linguistics and Philosophy in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Philosophy
at the
Massachusetts Institute of Technology

September, 2007

Signature of Author: .....
Department of Linguistics and Philosophy
June 11, 2007

Certified by: .....
Judith Jarvis Thomson
Professor of Philosophy
Thesis Supervisor

Accepted by: .....
Alex Byrne
Professor of Philosophy
Chair, Committee on Graduate Studies

The Particular and the General: Essays at the Interface of Ethics and Epistemology

by

A. Selim Berker

Submitted to the Department of Linguistics and Philosophy
on June 11, 2007 in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in Philosophy.

Abstract

This dissertation consists of three chapters exploring the nature of normativity in ethics and epistemology, with an emphasis on insights that can be gleaned by comparing and contrasting debates within those two fields.

In chapter 1, I consider *particularism*, a relatively recent view which holds that, because reasons for action and belief are irreducibly context-dependent, the traditional quest for a general theory of what one ought to do or believe is doomed for failure. In making these claims, particularists assume a general framework according to which reasons are the ground floor normative units undergirding all other normative relations. However, I argue that the claims particularists make about the behavior of reasons undermines the very framework within which they make those claims, thus leaving them without a coherent notion of a reason for action or belief.

Chapter 2 concerns a problem arising for certain theories that take the opposite extreme of particularism and posit a fully general theory of what one ought to believe or do. In the epistemic realm, one such theory is process reliabilism. A well-known difficulty for process reliabilism is the *generality problem*: the problem of determining how broadly or narrowly to individuate the process by which a given belief is formed. Interestingly, an exactly parallel problem faces one of the most dominant contemporary ethical theories, namely Kantianism. I show how, despite their seeming differences, process reliabilism and Kantianism possess a markedly similar structure, and then use this similarity in structure to assess the prospects that each has of ever solving its version of the generality problem.

Finally, in chapter 3, I consider a recent argument by Timothy Williamson that what it would be rational for one to do or believe is not *luminous*, in the following sense: it can be rational for one to do or believe something, without one's being in a position to know that it is. Careful attention to the details of Williamson's argument reveals that he can only establish this limit to our knowledge by taking for granted certain controversial claims about the limits of belief.

Thesis Supervisor: Judith Jarvis Thomson
Title: Professor of Philosophy

Acknowledgements

Two people deserve thanks above all others for helping me complete this dissertation: my wife, Carol, and my primary thesis adviser, Judith Jarvis Thomson. Due to an unfortunate encounter with a certain tick-borne illness, finishing this dissertation proved much more difficult than expected, and I don't think I could have done it without Carol's love and support. Judy, on the other hand, devoted vast amounts of time in discussing this work with me, and I feel that her boundless demand for rigor has made me much of the philosopher that I am today.

I would also like to thank the other members of my committee, Alex Byrne and Caspar Hare, as well as those faculty in the MIT philosophy department with whom I discussed some or all of this work: Stephen Yablo, Robert Stalnaker, Agustín Rayo, Vann McGee, Rae Langton, Richard Holton, and Ned Hall. I also owe a great debt of gratitude to many of my fellow graduate students while at MIT for their help, including Clare Batty, Jason Decker, Tyler Doggett, Andy Egan, Leah Henderson, James John, Sarah Moss, Bernhard Nickel, Dilip Ninan, Chris Robichaud, Eric Swanson, and Seth Yalcin. Tyler in particular deserves singling out for having commented on multiple versions of every single chapter of this dissertation. Finally, I would like to thank my mother, Terrell Sundermann, and my father, Nihat Berker.

Table of Contents

Abstract 3

Acknowledgements 5

1. Particular Reasons 9
1. Introduction . . . 9
2. The Generalized Weighing Framework . . . 13
3. Particularism about Reasons for Action . . . 18
4. Weak Holism about Reasons for Action . . . 20
5. The Argument from Cases . . . 23
6. The Argument from a General Holism . . . 30
7. Whither Non-Combinatorialism? . . . 37
8. Do Particularists Have a Coherent Notion of a Reason for Action? . . . 40
9. A Newtonian Analogy . . . 51
10. Giving Up Non-Combinatorialism? Or Moving Beyond Weighing? . . . 55

2. General Processes, Specific Maxims 61
1. Introduction . . . 61
2. The Generality Problem for Reliabilism: An Initial Characterization . . . 63
3. The Generality Problem for Kantian Ethics: An Initial Characterization . . . 66
4. The Need for a Deeper Characterization of Each Problem . . . 69
5. Alston's Attempted Solution . . . 71
6. Structural Similarities Between the Two Problems . . . 76
7. Desiderata for a Successful Solution to Either Problem . . . 79
8. First Neo-Kantian Strategy: Privileging a Certain Level of Generality . . . 81
9. Second Neo-Kantian Strategy: Counterfactual Tests . . . 84
10. Third Neo-Kantian Strategy: The First-Person Perspective . . . 86
11. Toward a General Hypothesis on the Origins of the Generality Problem . . . 91

3. Luminous Conditions 99
1. Introduction . . . 99
2. Williamson's Core Argument . . . 100
3. Narrowing the Target . . . 105
4. Coarse-Grained Safety . . . 106
5. Fine-Grained Safety . . . 114
6. Do the Relevant Constitutive Connections Obtain? . . . 124
7. Coda: The Lustrous and the Luminous . . . 128

References 133

Chapter 1: Particular Reasons

1. Introduction

Many, if not most, fields of human inquiry are guided by a search for unity. Whether one is studying economics or biology, philosophy or physics, it seems entirely unsatisfactory to say that instances of a given phenomenon (be it inflation or evolution, personal identity or proton decay) form a motley crew with no discernable underlying unity. What we usually want is a common explanation for how the various instances of the phenomenon hang together—some general principle linking those instances to one another—and to insist that such an explanation or principle is not in the offing would seem to invite skepticism about the reality of the very phenomenon being investigated.

In ethics, this pressure toward unification takes its starkest form in the long-standing tradition that sees the field as a search for a unified theory of the moral realm, as a quest for a "supreme principle of morality" (as Kant called it) that would give us a single, universal criterion for rightness and wrongness in action. This principle could then be used to *discover* which of the actions available to us in a given circumstance are right and which are wrong, and also to *justify* (both to ourselves and to others) why, exactly, the right actions are right and the wrong ones wrong. The two best-known attempts at such a single principle of morality are Kant's categorical imperative and Bentham's principle of utility, but this tradition lives on to this day, with many contemporary Kantians and consequentialists, among others, actively searching for what they take to be the most plausible version of a fully unified theory of morality.

In the past century a growing number of moral philosophers have expressed dissatisfaction with this *monist* tradition according to which there must be a *single* fundamental principle of morality. Often influenced by Aristotle, these critics find the monist principles thus far proposed unbearably crude and contend that *any* attempt to reduce morality to a single principle will inevitably leave something out. Monism seems to assume that there exists a moral algorithm, and if only we knew it, we could turn the crank and deduce for any given situation what we should and should not do. But, these critics insist, morality isn't the sort of thing that can be reduced to an algorithm; it isn't the sort of thing that could in principle be programmed into a computer. Rather, what is supposedly needed to arrive at the correct

moral verdict in a given situation is *moral wisdom*—a kind of *sensitivity* to the morally relevant considerations present in the case at hand, and an ability to properly *judge* the right thing to do in light of those considerations.

Among the philosophers sympathetic to this sort of criticism of the monist tradition, some have chosen to embrace a *pluralist* approach to ethics such as that found in W. D. Ross' theory of *prima facie* duties or in some (but not all) forms of virtue ethics, where instead of one, solitary moral principle there is posited to be a plurality of basic principles—an "unconnected heap of duties," to use David McNaughton's apt expression, that are all equally fundamental.[1] These differing duties can in principle conflict, but in such cases there are claimed to be no finitely codifiable rules dictating which duty trumps or outweighs the others, for otherwise the basic principles together with the weighing rules could be conjoined into a single master principle. Instead, what is allegedly needed in cases of conflict is the ability to accurately judge whether, say, one's duty of beneficence to perform some action outweighs one's duty of justice to avoid it, or whether it would be better in this situation to be honest than to be loyal. The hope is that by allowing for a plurality of equally fundamental moral principles we can better cover the entire moral landscape, while at the same time providing an ineliminable role for moral judgment in deciding what one ought to do in a given circumstance.

In recent years, though, an even more radical break with the monist tradition has begun to develop.[2] According to what has come to be known as *moral particularism*, even the pluralist's search for a multitude of basic moral principles is in vain. It's not that we need seven, or 27, or even 207 fundamental moral principles to fully capture the moral realm; rather, the particularist insists that *no* finite number of finite, exceptionless principles could account for all the truths there are about right and wrong, good and bad, virtue and vice. In the particularist's eyes, it's judgment all the way down: judgment as to which features of a given situation are morally relevant, and judgment as to how the features that are morally relevant play off each other to determine what one should and should not do in that situation. If the

---

1 McNaughton (1996).

2 See Dancy (1981, 1983, 1993, 2000a, 2001, 2004); Kihlbom (2002); Lance & Little (2004, 2006a, 2006b); Little (2000, 2001a, 2001b); McNaughton (1988); and McNaughton & Rawling (2000). The possibility of such a position was first pointed out and given the name "particularism" in Hare (1963), p. 18.

particularist is right, any attempt to reduce morality to principles, even the sort of *pro tanto* principles appealed to by Rossian pluralists, will inevitably lead to error: the features that make something good are just so complicated, the conditions under which an action is right just so variegated, and the properties that make a person virtuous just so nuanced, that the moral realm resists capture by *any* finite number of finite principles. Therefore the pluralist's quest for a disconnected heap of duties is just as misguided as the monist's dream of a grand unified moral theory.

Such is the particularist's challenge to both monist and pluralist approaches to ethics. But it is one thing to have a sneaking suspicion that there are no substantive moral principles—one thing to place one's bets, as it were, on none ever being found—and quite another thing to adequately establish that this must be so. The main particularist strategy is to argue for their claims about *principles* by first attempting to secure certain claims about *reasons*. More precisely, full-blooded moral particularism is not one thesis, but rather a cluster of three related theses. Roughly stated, these theses are:

> *particularism about reasons for action*: Whether a given feature of an action provides a reason for or against acting in that way—as well as how all the relevant reasons combine to yield an action's overall moral status—is an irreducibly context-dependent matter.
>
> *particularism about moral principles*: There are no substantial, finite, exceptionless moral principles (or at least we should not expect there to be any).
>
> *the particularist moral epistemology*: Knowledge of what one ought to do at a given time is attained by directly attending to the morally relevant features of the particular situation that one is in, not by subsuming that situation under general principles.[3]

Particularists usually proceed by *first* arguing for particularism about reasons for action, and *then* insisting that particularism about moral principles and the particularist moral epistemology follow from that view;

---

[3] See Little (2001a), p. 32, for a similar tripartite division among the particularists' claims. Exact formulations of particularism about moral principles vary significantly from particularist to particularist (and even within the work of the same particularist over time); see McKeever & Ridge (2005a) for a helpful discussion of the various possibilities.

A number of philosophers who do not consider themselves full-fledged particularists have recently endorsed one or another of the three central particularist doctrines while remaining agnostic about, or even explicitly opposed to, the other two. For instance, Cullity (2002) presents an argument for a weak version of particularism about reasons for action while remaining silent on the subject of particularism about moral principles and the particularist moral epistemology, whereas Jackson, Pettit & Smith (2000) appear willing to endorse some form of particularism about reasons for action despite explicitly arguing against particularism about moral principles. On the other side, Holton (2002) explores the plausibility of a version of particularism about moral principles while remaining silent on the subject of particularism about reasons for action and the particularist moral epistemology, whereas Crisp (2000) and Raz (2000) appear willing to endorse some form of particularism about moral principles despite explicitly arguing against particularism about reasons for action. Finally, although several of the articles reprinted in McDowell (1998) express sympathy for positions very similar to particularism about moral principles and the particularist moral epistemology, in conversation McDowell openly disavows particularism about reasons for action.

indeed, that is the general argumentative strategy pursued by the most prominent particularist, Jonathan Dancy, and it is echoed by many of his most ardent allies.[4] The purpose of this chapter will be to assess that strategy by considering the case that particularists make for their views about reasons for action.

The structure of this chapter will be as follows. First (§§2-4), some time will be spent carefully formulating particularism about reasons for action. I think that in the current debates about particularism, insufficient attention has been paid to the fact that particularists frame their entire position within an unargued-for framework according to which reasons are the fundamental normative units whose interactions determine, through a metaphorical balancing of the weight of reason, all other normative properties and relations. Making explicit this framework both clarifies the assumptions of the debate, and allows one to formulate the central tenets of particularism in a perspicuous manner. Second (§§5-7), I will consider the particularists' two main methods of arguing for particularism about reasons for action. Both of these argument will be found wanting, but the main reason for considering them is that one way in which they fail points the way to a deeper problem for particularism. The deeper problem is this (§§8-9): the position that particularists seek to establish about the nature of reasons threatens to undermine the very framework within which they attempt to establish that position, thus leaving them with no coherent notion of a reason for action; or, at least, so I shall argue. Finally (§10), I will consider an important reply to this charge of incoherence, and close with some speculation about the tenability of what I take to be the real position that particularists are after.

By focusing in this way on the particularists' views about reasons, I will for the most part be ignoring the significant further question as to whether particularism about reasons for action genuinely entails particularism about moral principles and the particularist moral epistemology.[5] However, I believe that there is enough cause for concern over the plausibility and even coherence of the particularists' views about practical reasons to render this further question essentially moot. The case for particularism about

---

[4] Dancy's defense of all three particularist doctrines via a defense of particularism about reasons for action can be found in his (1983, 1993, 2000a, 2004), among other places. Other particularists who follow Dancy's lead in pursuing this strategy include Kihlbom (2002), Lance & Little (2004, 2006a, 2006b), Little (2000, 2001a, 2001b), and McNaughton (1988).

[5] For arguments that such an entailment does not hold, see Jackson, Pettit & Smith (2000), pp. 96-99; Holton (2002), p. 197, n. 12; Väyrynen (2004); and McKeever & Ridge (2005b).

reasons for action is sufficiently fraught with difficulties that unless that position is drastically amended or the arguments for it radically improved, the version of the particularist project that starts by first attempting to secure particularism with respect to reasons is in danger of never getting off the ground.

2. The Generalized Weighing Framework

Particularists typically formulate their views within the confines of a certain widely accepted framework for how reasons for action function. Thus it will help to outline this framework before laying out the particularists' own position on the nature of reasons.[6]

Suppose Andy is trying to decide which of two apartments to rent. How might he go about making that decision? One natural suggestion is that Andy should do something like the following. On a piece of paper he should make four columns and, in the first column, write down the positive elements—or "pros"—associated with taking the first apartment; in the second column, write down the negative elements—or "cons"—associated with taking the first apartment; in the third column, write down the pros of taking the second apartment; and, in the fourth column, write down the cons of taking the second apartment. Next, he should decide the weight of each item in the four columns—that is, decide how heavily each pro or con will impact his ultimate decision. Finally, he should survey the relative weights of the various pros and cons, and come to a final decision about which option has the most favorable balance of considerations in its favor. Note that in this process Andy need not be able to represent the weight of each pro or con with anything as precise as a numerical value. Note, also, that the final determination of which option is the weightiest need not involve anything as mechanical as adding up the weights of a given option's pros and subtracting the weights of its cons in order to determine a total weight that can be compared with the total weight of the other option. But despite these deviations from a strict analogy with a weighing of masses on a scale, there still seems to be a useful sense in which we can say, metaphorically at least, that Andy is weighing his options—or, as it is commonly put, that he is weighing the *reasons* for and against each alternative to see where the balance lies.

---

[6] See Broome (2004), §3, and Horty (2004), §3.3, for similar accounts of the general framework assumed by particularists in their discussions of reasons for action.

According to what I will call the *generalized weighing model of practical deliberation*, all practical deliberation does—or at least should—take the above form. Whenever an agent is in a given situation, there are a variety of actions that the agent might choose to perform in that situation.[7] Each of these actions in turn possesses a variety of non-normative properties, such as the property of causing pleasure in someone or the property of being a telling of a lie. According to the generalized weighing model, some of these non-normative features give rise to a *reason in favor of* performing the action possessing them, and some of the features give rise to a *reason against* performing the action possessing them.[8, 9] Moreover, each of these reasons has a metaphorical *weight* or *strength* corresponding to how heavily it counts in favor of or tells against the action it counts in favor of or tells against. Finally, the generalized weighing model holds that each available action's *overall moral status* (such as its being right or wrong, obligatory or forbidden, permissible or supererogatory) can be determined by balancing the weights of the relevant reasons against each other in order to see where the overall weight of reason lies.[10] When so construed, the generalized weighing model is a widely held conception of practical deliberation with an undeniable attraction to it; indeed, talk of reasons and their weights almost makes this model seem inevitable.[11]

As presented thus far, the generalized weighing model is an epistemological position concerning how one does (or should) determine the overall moral status of the actions available to an agent. However, it is also tempting to assume that the deliberative order of discovery mirrors the metaphysical order of explanation—or in other words, it is tempting to make a slide from the *epistemological claim* that the overall

---

7 Following the particularists, I will be using the terms "situation," "circumstance," and "context" interchangeably.

8 The type of reasons under consideration are sometimes called "contributory reasons" or "*pro tanto* reasons" to represent the fact that they need not be decisive reasons for a given course of action. (A slightly older term for the same notion is "*prima facie* reason"—a term that was misleading, since the entities in question are not merely "at first glance" reasons, but retain their normative force even if outweighed.) For brevity I will usually omit the "contributory" or "*pro tanto*" qualifier.

9 A note about the ontology of reasons: throughout I talk about the *features* of an action as *providing* (or *giving rise to*) reasons for or against performing that action, and about the *fact* that an action would have a given feature as *being* a reason for or against performing that action. By taking reasons to be facts, I am regimenting our moral terminology in a way that not all theorists might endorse. However, nothing of consequence hangs on the particular regimentation I have settled on, and one can easily translate my talk of reasons as facts into a terminology adverting to one's ontological category of choice.

10 My discussion here concentrates on the overall *moral* status of actions, but the generalized weighing model is often applied more generally to the overall *normative* status of an action, such as its being what one ought (all things considered) do, or its being what it is most in one's interests to do. Particularists usually intend their claims to extend to these normative categories as well.

11 The two contemporary works most responsible, I believe, for making the generalized weighing model of deliberation as widely accepted as it currently is are Baier (1958) and Nagel (1970), both of which quite explicitly endorse the model: see ch. 3 of the former and ch. 7 of the latter. A more recent endorsement of essentially the same model can be found in Scanlon (1998), pp. 65-66, and Parfit (forthcoming).

balance of reason *reveals* which actions are right and which actions are wrong, to the *metaphysical claim* that the overall balance of reason *is what makes* the right actions right and the wrong actions wrong (and similarly for other overall moral verdicts). According to what I will call the *generalized weighing model of morality*, this metaphysical claim is indeed the case: the non-normative features of the actions available to an agent in a given circumstance give rise to genuine metaphysical normative entities, called "reasons for action," and the interplay of these reasons is what makes the available actions have the overall moral status that they do.

To simplify our discussion, let us focus on two overall moral statuses that an action might have, namely that of being *right* and that of being *wrong*. Also, drawing on a partial analogy with chemistry that has become standard in the particularist literature, let us say that a reason for action has a *positive* or *negative valence* depending on whether it is a reason *for* or *against* action, respectively. Then on the generalized weighing model of morality, we can think of the metaphysical moral picture as having three layers to it:

1. *The underlying level*: the facts about the non-normative properties of the actions available to a given agent in a given circumstance.

2. *The contributory level*: the facts about the valences and weights of the reasons for and against performing each available action, which obtain *in virtue of* the facts at the underlying level.

3. *The overall level*: the facts about the rightness and wrongness of the available actions, which obtain *in virtue of* the facts at the contributory level.[12]

So once this model is in place, we can say that a non-normative feature of an action that provides a reason for acting in that way is a *right-making* feature, since it contributes toward the overall balance of reason in favor of that action being such as to make the action right. Similarly, a feature that provides a reason against an action can be said to be a *wrong-making* feature. What determines the rightness and wrongness of actions on this conception is the overall balance of reason in favor of each, and what determines the reasons for and against each action are its non-normative properties. To have a name for it, let us call the

[12] Note that this talk of normative facts obtaining is all unabashedly realist-sounding; indeed, I take it to be a basic assumption of the debate about particularism that moral realism holds. (Sometimes particularists insist that their talk of normative facts and properties can be construed minimalistically if one has anti-realist persuasions, but whether this is really so is open to debate.)

conjunction of the deliberative and metaphysical weighing models the *generalized weighing framework.*

Two slightly more concrete examples will help illustrate how the generalized weighing framework works. Although classical utilitarians rarely talk of reasons for or against action, we can easily imagine one way of fitting hedonistic act-utilitarianism (henceforth: utilitarianism) into the confines of the generalized weighing framework. According to such a version of utilitarianism, there are two non-normative properties at the underlying level that give rise to reasons for and against action at the contributory level: the property of bringing about pleasure in someone, which always gives rise to a reason for action, and the property of bringing about pain in someone, which always gives rise to a reason against action. The "always" here is important: on the utilitarian story, not only is the property of bringing about pleasure in someone right-making, but it is *necessarily* right-making (in every possible situation it provides a reason for action), and not only is the property of bringing about pain in someone wrong-making, but it is *necessarily* wrong-making (in every possible situation it provides a reason against action). Moreover, the utilitarian has a correspondingly simple story about what fixes the weight of a given reason: the weight of every reason *for* action is directly proportional to how much *pleasure* the action would bring about in the relevant person, and the weight of every reason *against* action is directly proportional to how much *pain* the action would bring about in the relevant person; indeed, we can even represent these weights with numbers if we wish. Finally, our utilitarian holds that the overall level is determined by the contributory level in the following manner: for each action, add up the numbers representing the weights of the reasons in favor of acting in that way and subtract the numbers representing the weights of the reasons against; the actions with the highest such total are right, and all other available actions are wrong.

A second ethical theory easily shoehorned into the generalized weighing framework is W. D. Ross' theory of *prima facie* duties.[13] Adapted so as to involve explicit talk of reasons for action, Rossianism maintains that there are seven sets of non-normative features that in every possible circumstance give rise to reasons for or against the actions possessing them. For instance, on this account it follows from Ross'

[13] At least since Urmson (1975), if not earlier, it has been common to reinterpret Ross' theory of *prima facie* duties as a theory of (contributory or *pro tanto*) reasons for action.

prima facie duty of fidelity that the property of being a breaking of a promise always provides a reason *against* action; from his *prima facie* duty of self-improvement that the property of contributing to the improvement of one's intelligence always provides a reason *for* action; and so on, for all seven of Ross' *prima facie* duties.[14] Thus in regard to what determines the valence of a given reason, the Rossian story is similar to the utilitarian story, in that each posits a certain range of non-normative features that are necessarily right- or wrong-making, although the Rossian list of such features is lengthier and of a more varied nature. Moreover, Ross insists that "the ground of the actual rightness of [an] act is that, of all acts possible for the agent in the circumstances, it is that whose *prima facie* rightness in the respects in which it is *prima facie* right most outweighs its *prima facie* wrongness in any respects in which it is *prima facie* wrong," which strongly suggests that he too would determine how the facts at the overall level depend on the facts at the contributory level through the simple additive procedure endorsed by the utilitarian.[15]

Where the Rossian account diverges in spirit from the utilitarian one is in how the *weight* of a given reason is determined. According to Ross, which of the various, possibly competing *prima facie* duties is most binding in a particular situation depends upon "the circumstances of the case," since "for the estimation of the comparative stringency of these *prima facie* obligations no general rules can . . . be laid down."[16] Thus on one plausible interpretation of Ross, he would hold that the relative weights of the reasons stemming from his *prima facie* duties vary from context to context: a non-normative feature of an action that gives rise to a reason of one weight in one situation might give rise to a reason of a different weight in a different situation, depending on the details of the two cases. This is the sense in which Ross' theory provides an ineliminable role for *judgment*—as there are no general, context-independent rules for how the non-normative features of an action determine the weights of the reasons for and against it, discerning the weight of an individual reason in a given situation requires *judging* how demanding that reason is in the current circumstance.[17]

---

[14] See Ross (1930), p. 21, for the original list of Ross' seven *prima facie* duties.

[15] Ross (1930), p. 46. He says something similar on p. 41.

[16] Ross (1930), pp. 19, 41.

[17] I have mentioned two theories that can be easily fit into the generalized weighing framework, but it is worth noting that not all moral theories can be so easily fit into the framework. In particular, it is far from clear how to formulate Kantian ethical

3. Particularism about Reasons for Action

Particularists applaud Ross' inclusion of an ineliminable role for judgment in his theory but feel he did not go far enough. As mentioned before, for particularists it's judgment all the way down: not only judgment as to the weight of a given reason, but *also* judgment as to the valence of a reason *and* judgment as to how the valences and weights of the relevant reasons play off each other to determine the overall moral status of the available actions. Dancy summarizes the particularists' views about reasons as follows:

> I see ethical particularism as merely one expression of an overall holism in the theory of normative reason. . . . Such an overall holism can be expressed as follows:
>
> 1. What is a reason in one situation may alter or lose its polarity [i.e. valence] in another.
> 2. The way in which the reasons here present combine with each other is not necessarily determinable in any simply additive way.[18]

Thus particularism about reasons for action is a two-fold thesis: it concerns both the way in which reasons for action arise out of a situation's non-normative features, and the way in which the reasons for action that are present in a given situation combine to yield the overall moral status of the available actions. Let us consider each of these claims in turn.

According to what has come to be known as *holism about reasons for action*,[19] the very valence of the reason (if any) provided by a given non-normative feature of an action can alter as we change contexts. Here is one sort of example offered by particularists in support of this thesis: in most situations the fact that an action would bring the agent pleasure is a reason *in favor of* performing that action, but when Jim tortures a cat for fun, the fact that doing so brings him pleasure is (allegedly) a reason *against* his acting in that way. Another example: in most situations the fact that I borrowed a book from you is a reason *in favor of* my returning it to you, but if you stole the book, then the fact that I borrowed the book from you is (allegedly) *no reason at all* for me to return it. The particularists' bold claim is that *all* reasons function in this way: for every fact that is a reason for action in one possible context, there is another possible context in which that same fact is either a reason against action or no reason at all, and an analogous claim is taken

---

theories within the generalized weighing framework without doing major damage to both the content and intent of those theories. (For more on the structure of Kantian ethical theories, see chapter 2 of this dissertation.)

[18] Dancy (2000a), p. 132.

[19] Some might find the use of the label "holism" here a bit odd, since the position in question is quite different from the other sorts of holisms that one finds in philosophy. However, by now the name "holism about reasons for action" (originally coined by Dancy) is too firmly entrenched in the philosophical lexicon to resist.

to hold for reasons against action. So whereas our utilitarian and Rossian theories identify certain *necessarily* right- and wrong-making features, particularists insist that all right-making features are only *contingently* right-making, and all wrong-making features only *contingently* wrong-making. In sum, particularists hold:

> *holism about reasons for action*: For every non-normative feature of an action that gives rise to a reason for/against action in one possible context, there is another possible context in which that feature either gives rise to a reason of opposite valence or else provides no reason one way or the other.[20]

Or as Margaret Little, another prominent particularist, puts it: "A consideration that in one context counts for an action, can in another count against it or be irrelevant."[21]

Holism about reasons for action concerns how the facts at the contributory level depend on the facts at the underlying level. The second half of particularism about reasons for action concerns how the facts at the overall level depend on the facts at the contributory level. Let us call the function that takes as input the valence and weight of all the reasons present in a given possible situation and gives as output the rightness or wrongness of each action available in that situation the *combinatorial function*; given the generalized weighing model's assumption that the facts at the contributory level determine the facts at the overall level, such a function must exist.[22] Let us say that the combinatorial function is *additive* if it can be calculated by adding up the weights of the reasons in favor of each action and subtracting the weights of

---

[20] In order to avoid certain cheap counterexamples in which one builds a complete description of a given context into the property being considered (for instance: the property of being a telling of a lie in such-and-such a situation, where the "such-and-such" provides an exhaustive description of some particular situation), particularists may need to restrict this thesis so that it pertains only to non-normative features that can be specified "in finite or helpful propositional form." (The latter phrase is Little's; see her (2000), p. 280). Because it is ancillary to my main concerns, I ignore this complication in what follows.

[21] Little (2001a), p. 34.

[22] Kagan (1988) uses the expression "governing function" for much the same notion, and then goes on to argue that this function is not additive. Nagel (1970) prefers the term "combinatorial principle," but there is a worry that this choice of terminology begs the question against the particularist by assuming that the combinatorial function can be represented as a finitely expressible principle.

Note my talk of *the* combinatorial function, as opposed to *a* combinatorial function: since I am using the term "function" in the mathematical sense, there is only one combinatorial function, which holds for all contexts. Suppose I ask you to think of two natural numbers and either add the numbers together if they are both even, or multiply them if at least one is odd. Have I just defined one function, or two functions that change depending on which numbers you choose? In the mathematical sense, there is only one function here—only one mapping from the set of pairs of numbers you might choose to the set of numbers you might end up with after you have done what I tell you to do. One way of representing this function is as follows:

$$f(x,y)=\begin{cases} x+y & \text{if } x \text{ and } y \text{ are even} \\ x \bullet y & \text{otherwise} \end{cases}$$

Similarly, I intend there to be only one combinatorial function. Even if, intuitively, reasons combine one way in some contexts and combine a different way in other contexts, this can always be represented by a single function from the morally relevant factors in any given possible context to the rightness and wrongness of the actions available in that context.

the reasons against, and then assigning an overall moral status (right or wrong or neither) to a given action on the basis of comparing the total weight of reason in favor of that action to the total weight in favor of the other available actions. Then particularists are quite explicit in denying that the combinatorial function is additive.[23] Moreover, their general rhetoric makes it clear that they intend to deny that the combinatorial function can be written down in *any* finite formula, additive or otherwise.[24] In sum, particularists hold:

> *non-combinatorialism about reasons for action*: The combinatorial function for rightness and wrongness is not finitely expressible (and so in particular, not additive).

Thus for particularists, reasons for action are inextricably context-dependent twice-over: which non-normative features give rise to reasons for or against action varies from context to context, and how the various reasons that are present combine to yield the overall rightness and wrongness of actions also varies from context to context.[25] Together, the twin theses of holism and non-combinatorialism about reasons for action make up the view that I am calling *particularism about reasons for action*.

4. Weak Holism about Reasons for Action

Before turning to the main arguments for particularism about reasons for action, let us take a short digression in order to look more closely at the holist portion of that position. Not only does holism about reasons for action need to be distinguished from a weaker—and more readily acceptable—thesis, but pursuing this matter will allow us to introduce a distinction that will prove useful in what follows.

As stated so far, holism about reasons for action is a rather strong thesis: it maintains that *all*

---

[23] In addition to the passage already quoted from Dancy (2000a), p. 132 ("The way in which the reasons here present combine with each other is not necessarily determinable in any simply additive way"), Dancy (2004), p. 190, denies that "[o]nce one has assessed the separate weight of each element, evaluative judgment consists of adding up the pros and cons to see which side is weightier," and Little (2000), p. 280, rejects a view according to which each moral reason "goes in the hopper to be weighed against whatever other independent factors happen to be present." See also Dancy (2004), pp. 15, 105-106, 127, 143.

[24] Little (2000), p. 279, n. 3: "I'm reserving the 'particularist' label for those who deny codification at both levels [i.e. at both the contributory and the overall level]." Little (2001b), pp. 166-167: "The particularist begins by rejecting attempts to codify relations between nonmoral and moral properties. The resultant picture also leads to a rejection of efforts to systematize relations among moral properties." Dancy's discussion in his (2004) of Kagan (1988) makes it clear that he too intends to deny that there is a finitely expressible combinatorial function: "For the particularist, it is going to be variability all the way down" (p. 10). See also McNaughton & Rawling (2000), p. 260, n. 12, where they claim that "[t]he relation between an agent's reasons and her obligations is, we think, complex" and insist that there is "no weighing algorithm" for reasons.

[25] It is important to notice that the type of context-dependence at issue here is very different from the sort of context-dependence at issue in debates about epistemic contextualism. Unlike contextualists in epistemology, particularists are not making a semantic claim about how a given word such as "knowledge" or "reason" picks out a different relation in different contexts; rather, the context in question is the *context of the subject*, not the *context of utterance*.

reason-giving features of actions are only contingently right- or wrong-making. However, might not a particularist instead hold a weaker version of this thesis, according to which only *some* reason-giving features are contingently right- or wrong-making? That is, might not a particularist instead adhere to the following:

> *weak holism about reasons for action*: There are *some* non-normative features of actions that in one possible context give rise to a reason for/against action, and in another possible context either gives rise to a reason of opposite valence or provide no reason one way or the other.

Particularists and those with particularist leanings are somewhat divided over which version of the holist thesis to hold. Dancy usually endorses the original, stronger version of holism according to which all right- or wrong-making features are only contingently so.[26] Others only support the weaker version of the view.[27] However, this split in the particularist camp is puzzling, for two reasons. The first is that it is the stronger version of the thesis that is needed to derive particularism about moral principles; weak holism on its own is not enough.[28] The second is that weak holism about reasons for action is such a weak thesis that almost *any* moral theorist should accept it.

Suppose Tyler tells me to do something. It should be clear that, in some situations, the fact that Tyler told me to do something is a reason in favor of acting in that way.[29] For example, Tyler might be the captain of an airplane performing an emergency landing, and I might be his co-pilot. However, it should be equally clear that, in a multitude of other possible situations, the fact that Tyler told me to do an

---

[26] Dancy (2000a), p. 130. Later in the same article, though, Dancy concedes that he might be forced to admit the existence of a "privileged few" reasons whose valence is not sensitive to context, "including probably the intentional inflicting of undeserved pain" (p. 131), and he says similar things in his (2004) at pp. 77-78. I believe Dancy should be more wary than he is of admitting, however begrudgingly, that holism might only be true in its weaker form, for, as I will argue, not only is weak holism a trivial thesis held by nearly all moral theorists, but moreover, only the stronger version of holism can do the work Dancy needs it to do in securing particularism about moral principles.

[27] Cullity (2002), for example, argues in favor of weak holism and against the stronger version of the thesis.

[28] Proof: Suppose it turned out that the only reason-giving features of actions were the following: (a) the property of bringing about pleasure, which always gives rise to a reason in favor of action; (b) the property of bringing about pain, which always gives rise to a reason against action; and (c) the property of being a breaking of a promise, which gives rise to a reason against action in some contexts and gives rise to no reason one way or the other in some other contexts. Then weak holism about reasons for action would be true, since the property of being a breaking of a promise would be only contingently wrong-making. However, we would also have the following two true, exceptionless principles: "For all agents A and circumstances C, if an action X available to A in C would bring about pleasure, then that fact is a reason in favor of A's performing X in C," and "For all agents A and circumstances C, if an action X available to A in C would bring about pain, then that fact is a reason against A's performing X in C." So weak holism about reasons for action is compatible with the falsity of particularism about moral principles. (After writing this chapter I discovered that McKeever & Ridge make a similar point in their (2005b), p. 96.)

[29] Or more precisely, and more in keeping with our previously noted regimentation of terminology (cf. n. 9): in some situations, *the fact that an action would be one that Tyler told me to do* is a reason in favor of acting in that way.

action is neither a reason for nor against carrying out the action in question (say if Tyler on a whim told me to do some utterly inane activity), or might even be a reason *against* performing the action (say if Tyler is a sociopathic mass-murderer telling me to empty the contents of some canister into a town's water supply). Or to adapt an example of Judith Jarvis Thomson's,[30] the fact that an action would be a dancing of a jig is a reason against acting in that way in some possible circumstances (such as when one is attending a somber funeral), but is neither a reason for nor against acting in that way in other possible circumstances (such as when one is in the solitude of one's own home). Examples such as these are plentiful and should make it clear that *everyone*, particularist and non-particularist alike, accepts—or at least should accept—weak holism about reasons for action. But then it becomes rather puzzling why some philosophers have taken weak holism to be a substantive thesis worth dignifying with its own -ism and then arguing for.

However, is weak holism at least enough to refute the utilitarian and Rossian theories we considered earlier? After all, don't both theories assume that all right- and wrong-making features are necessarily right- and wrong-making? In fact, they need not make any such assumption. According to the utilitarian theory, there is one necessarily right-making property (that of bringing about pleasure in someone) and one necessarily wrong-making property (that of bringing about pain in someone), while according to the Rossian theory, there are seven sets of necessarily right- or wrong-making properties corresponding to Ross' seven *prima facie* duties; however, neither theory is committed to these being the *only* properties that are right- and wrong-making. Both theories can appeal to a distinction between *basic* and *derived* reasons for action, and then claim that it is only the features that provide basic reasons for action that are right- or wrong-making in all possible situations.

Let us say that the fact that action X would have feature F in circumstance C is a *derived reason* for (or against) performing X in C just in case, although that fact is indeed a reason for (or against) performing X in C, it is only a reason because: (i) if X were to have feature F in circumstance C, it would also have some other feature G, and (ii) the fact that X would have G in C is a reason for (or against)

---

[30] Harman & Thomson (1996), p. 193.

performing X in C. Call a reason for (or against) action that is not derived a *basic reason*, and call a feature of an action that gives rise to a basic or derived reason a *basically* or *derivatively reason-giving feature*, respectively. Then according to the utilitarian, there are only two basically reason-giving features: the property of bringing about pleasure in someone and the property of bringing about pain in someone. However, there are also any number of other features that in some contexts provide a derived reason for action, due to the fact that an action's having that feature in the given context would thereby bring about pleasure or pain; for example, consider the property of being a dancing of a jig, or of being something Tyler told me to do. Clearly on the utilitarian's story some of these features can give rise to reasons of opposite valences in different contexts, or even give rise to no reason at all, depending on whether an action's possessing such a feature would bring about any pleasure or pain in the situation in question. So the utilitarian can allow for the possibility that some derivatively reason-giving features are only contingently right- or wrong-making; it is only the *basically* reason-giving features that the utilitarian insists are necessarily right- or wrong-making. And similarly for the Rossian: she can distinguish a collection of basically reason-giving features corresponding to Ross' seven *prima facie* duties, all of which are claimed to be necessarily right- and wrong-making, from a collection of derivatively reason-giving features, some of which may only be contingently right- or wrong-making.[31]

Thus I think we should conclude that all moral theorists who endorse the generalized weighing framework ought to accept weak holism. The interesting question, as well as the one that is directly relevant for establishing the rest of the particularist creed, is whether the stronger version of the holist thesis is true.[32]

5. The Argument from Cases

On, then, to the particularists' main arguments for particularism about reasons for action. Recall that the

---

[31] Note that once we make a basic/derived reason distinction, to avoid overcounting issues only *basic* reasons should be inputted into the combinatorial function that determines the rightness and wrongness of the actions available to the agent.

[32] What about a version of holism, intermediate in strength between the two versions discussed here, which insists that some (though not necessarily all) *basic reasons* switch in valence from context to context? Such a thesis is controversial enough to be interesting, but the same example used in n. 28 to show that weak holism is compatible with the falsity of particularism about moral principles can likewise show that this intermediate version of holism is compatible with the falsity of that thesis.

view consists of the following two theses:

particularism about reasons for action:

1. *holism about reasons for action*: For every non-normative feature of an action that gives rise to a reason for/against action in one possible context, there is another possible context in which that feature either gives rise to a reason of opposite valence or else provides no reason one way or the other.
2. *non-combinatorialism about reasons for action*: The combinatorial function for rightness and wrongness is not finitely expressible (and so in particular, not additive).

Particularists have two main positive arguments for particularism about reasons for action, both of which originate in Dancy's writings: the first proceeds by appealing to our moral intuitions about cases, the second by appealing to a general holism about normative reasons (both for action and for belief). In this section I consider the first of these positive arguments; in the next I consider the second.

By far the most prevalent particularist strategy in arguing for particularism about reasons for action is to cite a host of examples allegedly supporting the view.[33] Or rather, the most prevalent particularist strategy in arguing for the *holist* half of that view is to cite a host of examples, since the particularist who appeals to examples usually insists on the general truth of particularism about reasons for action without actually offering any examples in support of non-combinatorialism. The *locus classicus* for such an argument by cases is chapter 4 of Dancy's book *Moral Reasons*, where Dancy opens his defense of particularism by providing a series of cases designed to illustrate the ways in which reasons can change or lose their valence from context to context, and many of the examples Dancy gives in this chapter have become oft-repeated staples of the particularist repertoire. Each example works by fixing on a feature that in many circumstances appears to provide a reason for (or against) performing the action possessing it, and then pointing to a case in which that feature is claimed either to give rise to a reason against (or for) performing the action possessing it, or else not to provide a reason one way or the other. Obviously none of these examples has any direct bearing on the truth of non-combinatorialism, but do they at least establish holism?

All of the examples that Dancy proposes in chapter 4 of *Moral Reasons*, as well as every other

[33] See, among other places, Dancy (1993), pp. 56, 60-62; Kihlbom (2002), pp. 40-46; Lance & Little (2006b), p. 579-581; Little (2000), pp. 280-281; Little (2001a), p. 34-35; Little (2001b), p. 165; and McNaughton (1988), pp. 192-194.

particularist example I have encountered in the literature, fall into three categories. The cases in the first category all involve reasons that almost any moral theorist would hold to be *derived*, and so fail to serve as part of an adequate inductive basis for the claim that *all* reasons for action, be they basic or derived, can change or lose their valence with changes in context. For instance, one of the examples Dancy cites is the following:

> . . . that we did this last time can be a reason for doing the same this time, but sometimes it will be a reason for doing something different. Whether it is so or not will depend on other features of the case.[34]

The example is intended to work as follows: we are supposed to accept that in some circumstances the fact an action would be one we have done before is a reason *in favor of* performing the action (say, if we find it comforting to do the same thing we have always done in the past), and that in other circumstances the fact an action would be one we have done before is a reason *against* performing the action (say, if our former practices have grown tedious and we yearn for something new). However, it should be clear that in either case the action's being one we have done before provides a reason for or against doing that action only because there is some *other* feature that the action possesses in virtue of its being one that we have done before, such as its being tedious or comfortable. But this is just to say that the feature Dancy has fixed on is only *derivatively* reason-giving in either situation, and I have already argued that *everyone* should recognize that some derivatively reason-giving features are only contingently right- or wrong-making. The real issue is whether there are any *basically* reason-giving features that are contingently right- or wrong-making.

The second category of cases avoids this failing, for the examples in this group fix on features involving pain or pleasure and thus succeed in isolating features that some moral theorists claim to be basically reason-giving. One such example is the following, which Dancy gets from David McNaughton:

> Consider the suggestion that we have more reason to have public executions of convicted rapists if the event would give pleasure both to the executioner and to the crowds that would no doubt attend. Surely this pleasure is a reason against rather than a reason for; pleasure at a wrong action compounds the wrong.[35]

Dancy's claim here is that, in such a case, the fact that an action (performing a public execution) would

---

[34] Dancy (1993), p. 61.

[35] Ibid. Note that if the passage is read most literally, the action under consideration would seem to be our *scheduling* of public executions ("*we* have more reason to *have* public executions"), but Dancy's subsequent discussion makes it clear that the action he intends to be assessing the reasons for and against doing is in fact the executioner's *performing* of the public execution.

cause pleasure in some people (the executioner and the spectators) is a reason *against* performing that action, whereas in many other cases the fact that an action would cause pleasure in some people is a reason *in favor of* performing that action. Now if Dancy's interpretation of this case is correct, then he has succeeded in establishing that a feature that some theorists insist to be basically reason-giving is only contingently right-making, for utilitarians, at least, hold that bringing about pleasure in someone is always a basically reason-giving feature of an action.[36] However, utilitarians need not accept Dancy's construal of the case at hand.

To begin with, a utilitarian might challenge Dancy's intuitions about the case. I find that very few people share Dancy's firm intuition that the execution's causing pleasure in the onlookers actually constitutes a reason *against* performing the execution. But even if the utilitarian concedes that many people share Dancy's intuition, the utilitarian has available to her a plausible way of explaining that intuition away. A standard utilitarian move is to distinguish between the rightness or wrongness of an action and the praise- or blameworthiness of the individual performing that action, or more generally, to distinguish between evaluations of actions and evaluations of agents.[37] With this distinction in hand, the utilitarian can reply to Dancy that the fact that the execution would cause pleasure in the executioner and the spectators does, in fact, count as a reason to perform the public execution, albeit one with such little weight in its favor that it is easily overridden by other considerations so as still to yield the result that it would be wrong to carry out the execution. However, the utilitarian might continue, the fact that the executioner and the crowd would derive pleasure from watching the execution is also a reason to judge poorly about their character, and it is this latter reason that Dancy is confusing with a reason against performing the execution. In this way the utilitarian can disarm Dancy's example by denying in a principled way his intuition about the case, and similar considerations can be employed against the particularists' other examples in which the causing of pleasure is alleged to give rise to a reason against (or the causing of pain to a reason for) a given action.

---

[36] Or at least those who accept the generalized weighing framework do.

[37] This move traces at least as far back as Mill (1863/1987), ch. 2.

Thus the examples in the second category would only tell against utilitarian theories of basic reasons, and even then the utilitarian has ample resources available to her with which to disarm those examples. Similar comments apply to the third group of cases, all of which appeal to the sorts of features that only a Rossian would claim provide a basic reason for or against action. Suppose I promised to do something, and performing action X in circumstance C would involve breaking that promise. Then according to most Rossians, the fact that X would involve breaking a promise is a basic reason against so acting. However, suppose the promise was made under duress—what then? Most particularists insist that in that case, the fact that X would involve breaking a promise is neither a reason for nor against performing X, since the promise in question was ill-gotten. Thus we appear to have a feature (being a breaking of a promise) that in some contexts provides a basic reason against action, or at least does so on a Rossian account, but in other contexts provides no reason one way or the other.[38]

However, as with the examples of the second sort, this third kind of case only threatens a single theory's account of which basic features are necessarily right- and wrong-making, and even then, theorists of that ilk have adequate means with which to rebut the alleged counterexample. According to most non-Rossian theories of reasons for action, the fact that an action would involve a breaking of a promise is only a *derived* reason against so acting; for example, on a (hedonistic act) utilitarian account, that fact is only a reason against action in virtue of the long-term pain brought about both by the specific breaking of the promise and by its contribution toward a general erosion to the practice of promise-keeping. Moreover, even theories like the Rossian's which do claim that the reason in question is basic have available to them ways of resisting the alleged counterexample, for Rossians (if they accept the particularists' intuitions about the case) are likely to insist that the real basically reason-giving feature is not *being a breaking of a promise*, but rather *being a breaking of a promise that was not given under duress*. Particularists might counter with claims that the condition of not being given under duress is merely an *enabling condition* for a promise to be practically relevant and is not itself part of the relevant reason against action.[39] But what entitles them to

[38] Dancy leans heavily on this sort of case in his (2004); see pp. 38-43, for instance.
[39] The "enabling condition" terminology is Dancy's. See his (2004), pp. 39-41.

such claims? Are our intuitions really so precise that we can say with confidence that in a normal case in which a promise is not given under duress, it is *the fact that the action would involve breaking a promise* that is a reason against action and not *the fact that the action would involve breaking a promise that was not given under duress*? My inclination is to say that our intuitions are not so precise, and if so, then it is perfectly open to the Rossian to insist that it is the latter fact that is a basic reason against action in all possible contexts.

This move represents a second sort of response to the particularists' examples. With monist theories such as utilitarianism, it makes sense, when faced with an example that threatens to show that what one identifies as a basic reason can switch in valence with changes in context, to stand one's ground and explain away our alleged intuitions about the case. However, with pluralist theories like Ross', there is no reason not to turn around when faced with a putative counterexample and simply build the particularist's enabling condition into one's basic reason; after all, if the pluralist is already positing a wide variety of basically reason-giving features, why not make a few of those features a little more complex? The particularist might in reply propose a new counterexample to the newly proposed basic reason (for example, what if the promise was solicited by fraud?), but then the Rossian, if convinced by the example, can turn around and make her proffered basic reason even more complicated in order to handle that example as well. The particularist will no doubt insist that she can produce such counterexamples indefinitely. But is this really so? After three or four iterations of this dialectic one's ingenuity tends to run out, and although the Rossian might be forced to make her set of basically reason-giving features even more complex and gerrymandered than they were in the beginning (which the Rossian is unlikely to see as a fault in her account, since those attracted to simplicity and unity in their ethical theories tend to avoid Ross-style pluralism), it is far from clear that she needs to make those features *infinitely* complex in order to make them immune to counterexample. So although the third category of examples might seem to threaten the Rossian's account of basic reasons, with some maneuvering she can avoid trouble.

Where does this leave us? I have claimed that every example provided by the particularists in support of holism about reasons for action falls into one of three categories. The examples in the first category are irrelevant: they fix on reasons that everyone would think are derived, and all theorists can

agree that some derived reasons switch in valence from context to context. The examples in the second category threaten the utilitarian account of basic reasons, but the utilitarian has responses available to her. And the examples in the third category threaten the Rossian account of basic reasons, but the Rossian also has responses available to her. Thus even if I am wrong about the utilitarian and the Rossian having adequate responses available, so that the particularists' examples do in fact show that their candidates for basically reason-giving features give rise to reasons of different valences in different contexts, then—ignoring the irrelevant cases in the first category—particularists would appear to be arguing from an inductive basis *of size two* to the claim that *any* given theory's candidates for basically reason-giving features must switch in valence from context to context. However, such a quick inductive argument seems incredibly rash, to say the least, for there remains a whole host of moral theories whose designated basically reason-giving features are untouched by the particularists' examples.

For instance, according to so-called desire-based accounts of reasons for action, all reasons for action are ultimately provided by what the agent wants, or would want under certain conditions.[40] On one such view, an agent has a reason to perform action X in circumstance C if and only if after ideal deliberation the agent would desire that she perform X in C. So on this account, *the fact that an action would bring about pleasure in someone* and *the fact that an action would involve breaking a promise* are only ever *derived* reasons for action in a given context, if they are even reasons at all; the only basic reasons are *the fact that an action would be one that after ideal deliberation the agent would desire that she do* and *the fact that an action would be one that after ideal deliberation the agent would desire that she not do*, and none of the particularists' examples do anything to show that the first of these is not always a reason in favor of action and the second not always a reason against. The same holds for any number of other non-holist theories of reasons for action, from numerous other desire-based accounts, to various non-desire-based views: the particularists' examples do nothing to show that these theories' designated basically reason-giving features are all only contingently right- or wrong-making.[41] The basic problem is that, by doing little more than proliferating putative

---

[40] See Chang (2004) for a useful discussion of the distinction between desire- and value-based accounts of reasons for action.

[41] Particularists might argue on independent grounds that desires cannot provide reasons for action, as Dancy does in ch. 2 of his (2000b), but then they would have only succeeded in eliminating *three* classes of plausible candidates for basically reason-giving

counterexamples of the same general sort, particularists have failed to provide any kind of guiding insight into the normative realm that would lead us to think that *any* purported basic reason must change its valence from possible context to possible context. As such, we have not been given any general reason to conclude that the phenomena the particularists are pointing at will extend to theories more sophisticated than those offered by the utilitarian and the Rossian.

6. The Argument from a General Holism

We have found the argument from cases for particularism about reasons for actions to be wanting: as an argument for the holist half of that thesis it is inconclusive, and as an argument for the non-combinatorialist half it is non-existent. Until recently, the argument from cases was the only positive argument that particularists had for their doctrine about reasons. However, in a recent article Dancy offers a second positive argument for holism about reasons for action that proceeds by arguing for a general holism about all normative reasons.[42] Let us now turn to this second argument and see if it fares better than the first.

According to Dancy, there are two basic kinds of normative[43] reasons: *theoretical reasons* (i.e. reasons for belief) and *practical reasons* (i.e. reasons for action). Within the domain of practical reasons Dancy sees a further division, between *ordinary practical reasons* (i.e. non-moral reasons for action) and *moral reasons* (i.e. distinctively moral reasons for action, if there are any), though he remains uncommitted as to the exact nature of this distinction (indeed, it would only help his argument if there is no such distinction to be found). With this taxonomy in place, the central claim behind Dancy's second argument is this: holism about reasons is a general thesis that holds throughout the domain of normative reasons, including both theoretical and practical reasons—or in other words, not only is holism about reasons *for action* true, but holism about reasons *for belief* is also true.

Once we recognize this (alleged) fact, Dancy believes we have an argument for holism about

---

features rather than two, and the main point would remain.

[42] Dancy (2000a), pp. 131-137; repeated with minor changes in his (2004), pp. 73-78, and in a more condensed form in his (2001), §3. Little briefly alludes to the argument in her (2000), p. 281-282, and in her (2001a), p. 34.

[43] "Normative" because some philosophers draw a contrast between *normative reasons* (which underwrite the normative requirements on an agent) and *motivating reasons* (which explain an agent's having acted in a certain way): see Smith (1994), ch. 4.

reasons for action available to us. To begin with, Dancy insists that the truth of holism about reasons for belief is utterly uncontroversial: "it seems to me that nobody has ever thought of denying [the thesis]."[44] Similarly, Dancy holds that holism about ordinary (i.e. non-moral) reasons for action is a universally held view whose truth "nobody has ever really debated." But then it would seem we have good reason to believe that holism most be true of moral reasons for action as well: wouldn't it be odd, Dancy asks, if holism were true of reasons for belief and of ordinary reasons for action, but were not true of moral reasons for action? Isn't it more plausible that the logic of reasons is a unified kind that is the same for all types of reasons, so that all normative reasons function holistically? Thus we should conclude that holism is true of moral reasons for action as well, and as we supposedly already know that holism is true of ordinary reasons for action, we can establish that holism is true about all reasons for action, both ordinary and moral.

In other words, Dancy appears to be arguing for holism about reasons for action as follows:

1. Holism about reasons for belief is true. [*premise*]
2. Holism about ordinary reasons for action is true. [*premise*]
3. If holism about reasons for belief and holism about ordinary reasons for action are both true, then holism about moral reasons for action must also be true. [*premise*]
4. So, holism about moral reasons for action is true. [*follows from 1-3*]
5. So, holism about *all* reasons for action is true. [*follows from 2, 4*]

This constitutes the particularists' second positive argument for the holist half of particularism about reasons for action. Unfortunately for particularists, though, there is good reason to doubt the truth of all three premises, and certainly the case Dancy makes for each is unsatisfactory. For example, Dancy's case for premise 2 rests on examples that are just as problematic as the ones he appealed to during his argument by example for holism about reasons for action, and his case for premise 3 involves an unmotivated assumption that there is a presumption in favor of every type of reason behaving similarly. However, the most serious problem with the argument occurs at premise 1, so in what follows I focus on that portion of the argument.

---

[44] Dancy (2000a), p. 132.

According to premise 1, holism about reasons for belief is true. Presumably holism about reasons for belief should be formulated as follows:

> *holism about reasons for belief*: If the non-normative fact that p is a reason for (or against) a subject believing that q in one possible context, then there is another possible context in which the fact that p either is a reason against (or for) the subject believing that q, or else is no reason for the subject to believe one way or the other.[45]

Or in other words, a given fact can only at best be *contingently* a reason for (or against) a given belief. Dancy's argument for the truth of holism about reasons for belief is exceedingly brief: he just takes it to be blindingly obvious that the thesis is true, and he provides a single example to illustrate his point.

> For instance, suppose that it currently seems to me that something before me is red. Normally, one might say, that is a reason (*some* reason, that is, not necessarily sufficient reason) for me to believe that there is something red before me. But in a case where I also believe that I have recently taken a drug that makes blue things look red and red things look blue, the appearance of a red-looking thing before me is reason for me to believe that there is a blue, not a red, thing before me. It is not as if it is some reason for me to believe that there is something red before me, but that as such a reason it is overwhelmed by contrary reasons. It is no longer *any reason at all* to believe there is something red before me; indeed it is a reason for believing the opposite.[46]

The example is supposed to work as follows. The fact that I seem to see something red before me is in most situations a reason for me to believe that there is something red before me. However, if I (justifiably) believe that I have recently taken a drug that makes blue things look red and red things look blue, then the fact that I seem to see something red before me is actually a reason for me *not* to believe that there is something red before me. Thus a feature which gives rise to a reason for belief in one possible context gives rise to a reason against belief in another—that is, the fact that I seem to see something red before me is only *contingently* a reason for me to believe that there is something red before me.

Now the first thing to note about Dancy's argument here is that, this time at least, he is not arguing by induction from cases. Dancy takes holism about reasons to be generally accepted by everyone, and he offers his example for purely illustrative purposes, not to serve as a single-element inductive basis.

---

[45] In keeping with my previous regimentation of terminology (cf. n. 9), I am taking reasons for belief to be facts, which some might dispute, particularly since there is a venerable (though to my mind mistaken) tradition in epistemology according to which only *beliefs* can be reasons for belief. However, as before nothing turns on my particular choice for the ontology of reasons of belief, and one can easily reformulate the examples to follow so that the reasons in question are beliefs, or propositions, or whatever other ontological category one prefers.

[46] Dancy (2000a), p. 132. This one example has become quite the particularist warhorse: Dancy produces this same, solitary example when recounting the argument from a general holism in his (2001), §3, and his (2004), p. 74, and Little does the same when alluding to the argument in her (2000), p. 281 and her (2001a), p. 34.

But it is far from clear that Dancy's sociological claim about the general acceptance of holism in the domain of reasons for belief holds true. The type of phenomenon that Dancy is pointing to with his example is usually called a *defeater* in the epistemology literature: the idea is that my (justified) belief about having ingested a red/blue-switching pill *defeats* the reason which I would otherwise have to believe, on the basis of how things appear to me, that there is something red before me. However, determining whether there is widespread agreement with Dancy's interpretation of the case at hand is complicated by two issues. First, most contemporary epistemologists formulate their theories in terms of *justification* (or near synonyms such as "warrant" or "entitlement") rather than in terms of *reasons for belief*, and it is not always a trivial matter to translate claims made in one language into claims made in the other. Second, and more importantly, discussions of defeaters in the epistemology literature primarily focus on defeaters *at the overall level* rather than defeaters *at the contributory level*: the issue epistemologists tend to concentrate on is whether, at the overall level, one's sufficient reason to believe that *p* (or one's adequate justification for believing that *p*) is defeated in some particular situation, not whether, at the contributory level, one of one's *pro tanto* reasons to believe that *p* (or one of one's *pro tanto* justifications for believing that *p*) is defeated.[47]

It is customary to distinguish between two types of defeaters at the overall level: rebutting defeaters and undercutting defeaters.[48] Formulated in terms of reasons for belief, a *rebutting defeater* is a *pro tanto* reason to believe that not-*p* which is strong enough that, although one *would* have had sufficient reason to believe that *p* were that reason not present, as things are one *does not* have sufficient reason to believe that *p*.[49] So in the above example, if instead of remembering that I had taken a red/blue-switching

---

[47] Talk of defeaters and defeasibility in epistemology began with Roderick Chisholm's (1964), in which he develops a formalized version of a Ross-style ethics of *prima facie* duties and then notes certain parallels between how ethical requirements may be overridden or defeated and how evidence may be overridden or defeated. Since then, appeals to notions of defeat in the epistemology literature has centered around attempts to solve the Gettier problem by proposing a "defeasibility condition" on the justification required for knowledge; for an introduction to defeasibility accounts of knowledge, see the appendix to Pollock (1986), and for a more thorough investigation of the various accounts that have been proposed, see Shope (1983), ch. 2.

[48] The rebutting defeater vs. undercutting defeater terminology is John Pollock's, who is usually credited with making this distinction; see his (1986), p. 38, among other places (though be aware of Pollock's idiosyncratic definition of a reason in terms of overall justification on p. 36). Terminologies for the distinction vary: for example, Casullo (2003) prefers to call them overriding and undermining defeaters, respectively, whereas Sosa (1985) uses the terms "opposing overrider" and "disabling overrider."

[49] It is customary when characterizing types of defeaters to appeal to subjunctive conditionals concerning what reason or justification one *would* have *were* certain factors not present, but often the subjunctive conditionals appealed to lead to unintended consequences for the proposed definitions: see Feldman (2003), p. 33-34. However, for the purposes of introducing the general rebutting/undercutting defeater distinction, we can overlook such difficulties.

pill I had been told by an extremely reliable source that there is nothing red in front of me, the reason provided by that piece of testimony would have been a rebutting defeater of the reason to believe that there is something red before me provided by how things appear to me. A rebutting defeater of one's sufficient reason to believe that *p* thus attacks the conclusion that *p*. An *undercutting defeater*, on the other hand, attacks the connection between one's reasons and the conclusion that *p*, with the result that one no longer has sufficient reason at the overall level to believe that *p*. However, undercutting defeaters can do this in various ways: they can *disable* one of one's *pro tanto* reasons so that what would have been a *pro tanto* reason to believe that *p* is now no such reason (perhaps because it is now not a reason at all, perhaps because its valence has been switched so that it is now a *pro tanto* reason to believe the opposite), or they can *diminish* the strength of one of one's *pro tanto* reasons to believe that *p* to such a degree that one now no longer has overall sufficient reason to believe that *p*, or they can leave the valence and strength of one's *pro tanto* reasons untouched but change what is required for one's overall reason to believe that *p* to constitute *sufficient* overall reason to believe that *p*. Thus in order for his argument to work, Dancy needs epistemologists to agree that the red/blue-switching case is not just an example of an undercutting defeater, but rather an example of an undercutting defeater of the first variety, in which one of one's would-be *pro tanto* reasons is completely disabled. In fact, he needs more than that: since ultimately he is advocating a strong version of holism according to which *all* reasons can switch or lose their valence from context to context, Dancy also needs there to be general agreement among epistemologists that *every* reason for belief can be undercut in the way described.

Some epistemologists would simply deny that the red/blue-switching case is an example of an undercutting defeater that completely disables the *pro tanto* reason to believe that there is something red before me. One particularly clear example of how one might deny this is provided by the version of modest foundationalism defended by James Pryor in his (2000). In that article Pryor argues that we always have *prima facie* (i.e. *pro tanto*) justification to believe that the world is as our experiences present it to be. Translated into talk of reasons for belief, Pryor's claim is that an experience as of its being the case that *p* necessarily provides a *pro tanto* reason to believe that *p*—so in particular, its seeming to me that there is

something red before me provides a reason to believe that there is something red before me in every possible situation, even one in which I believe that I have ingested one of those pesky red/blue-switching pills. In such a case, what Pryor would say is that I also have a very good reason to believe that if I seem to see something red before me, then there is something blue before me (and not something red). Since in addition I have good reason to believe that I seem to see something red before me, simple *modus ponens* reasoning gives me good reason to believe that there is not something red before me, and this reason might *outweigh* the one to believe that there is something red before me provided directly by my experience. So Pryor would claim that Dancy's example is at most an example of a *rebutting* defeater, and other theorists might agree that this is not an example of the needed type of undercutting defeater, perhaps because they think that the *pro tanto* reason provided by my experience as of something red is only *diminished* in strength, or perhaps because they think its strength remains the same but what is required for my overall reason to qualify as *sufficient* overall reason to believe that there is something red before me has been changed.[50]

However, let us suppose that Dancy is correct about the case at hand, and that once one believes (or perhaps justifiably believes) that one has taken a red/blue-switching pill, the fact that one seems to see something red before one is no longer a *pro tanto* reason to believe that there is something red before one. The problem is that then Dancy would have provided only one compelling example of a fact that is merely contingently a reason to believe a given conclusion, and there are any number of examples of other facts that it is plausible to think are necessarily a reason to believe some conclusion, in a way that is not undermined by the type of phenomenon present in the red/blue-switching example. For example, consider our reasons for believing mathematical truths. The fact that 19 times 23 equals 437 seems like a good reason to believe that 437 is not a prime number; indeed, it would seem to be a *conclusive* reason in favor of such a belief. However, it is rather difficult to conceive of any circumstances in which the fact that

---

[50] Another way of rejecting Dancy's interpretation of the example would be to deny that how things seem to me *ever* provides a (basic) *pro tanto* reason for belief, let alone one whose sufficiency can be defeated; for example, this would probably be the reaction of most externalists about justification, since they usually hold that that it is "external" facts such as those concerning the reliability of the process by which a belief was formed that provide justification or give one reasons, not "internal" facts about what one seems to see. (For more on externalist theories of justification of the reliabilist variety, see chapter 2 of this dissertation.)

19 times 23 equals 437 would be neither a reason for nor against believing that 437 is not a prime number, much less one in which that fact would somehow be a reason *against* believing that 437 is not a prime number. Or take Descartes' *cogito* argument: the fact that I am now thinking seems to be a good reason to believe that I exist. (The exact nature of this "I" that exists is another matter, but surely it at least gives me reason to believe *that* I exist.) Yet is there any context in which I am now thinking but that gives me no reason for or against believing that I exist, or any context in which I am now thinking and that (rather incredibly) gives me reason *against* believing I exist? Finally, we can turn Dancy's own example against him: there may indeed be some strange contexts in which the fact that I seem to see something red before me is not a reason for me to believe *that there is something red before me,* but are there any possible contexts in which the fact that I seem to see something red before me is not a reason for me to believe *that something is causing me to seem to see something red before me*? In effect, what Dancy needs to claim is that undercutting defeaters of the relevant disabling variety exist for each of these reasons—a claim which is *incredibly* controversial, and to the best of my knowledge not endorsed by *any* currently practicing epistemologist.

Which is not to say that such undercutting defeaters might not exist. Indeed, even if every single contemporary epistemologist were united in denying the existence of such defeaters (and I suspect they very well might be), they could all be wrong. Definitively settling the truth or falsity of holism about reasons for belief would take us too far adrift. But it should be clear that Dancy's insistence that "nobody has ever thought of denying" the view is blatantly untrue; indeed, if there is any lack of controversy over the status of holism about reasons for belief, it would seem to be a lack of controversy among epistemologists in agreeing, contra Dancy, that the thesis is *false*. So Dancy cannot avoid making a substantive argument for the view by merely gesturing at its (alleged) universal acceptance. Rather, if he plans to establish holism about reasons for action by first establishing holism about reasons for belief, he is going to need some sort of positive argument in favor of holism about reasons for belief. And if he proceeds by appealing to examples, he will face the same problems he faced when offering his argument from cases for holism about reasons for action: some theorists will, like Pryor, deny his interpretation of

the cases provided (compare the response of the utilitarian to the argument from cases); other theorists will accept his interpretation of the cases but insist that the real basic reason for belief is a more complicated fact that avoids the counterexample, such the fact that one seems to see something red when one's faculty of vision is working correctly (compare the response of the Rossian to the argument from cases); and almost every theorist will deny that his examples serve as an inadequate inductive basis for the claim that *all* reasons for belief can switch or lose their valence from context to context (compare the inadequacy of the inductive basis in the argument from cases). So Dancy's argument from a general holism for holism about reasons for action will ultimately be just as convincing—or rather, just as *un*convincing—as his original argument from cases for that thesis.

7. Whither Non-Combinatorialism?

Thus I conclude that the two positive arguments originating in Dancy's work for holism about reasons for action both fail to establish the needed version of holism in a convincing manner. However, the attentive reader will notice that something seems to have been left by the wayside here. Wasn't particularism about reasons for action supposed to be a *two-part* view, of which holism about reasons for action was only one half? What, one might ask, about non-combinatorialism about reasons for action? Do particularists produce any positive arguments for *that* portion of their creed? And if so, are those arguments convincing?

The answer is that although the most full-blooded particularists such as Dancy and Little endorse both holism and non-combinatorialism, and although sometimes they act as if the argument from examples and the argument from a general holism also establish non-combinatorialism, Dancy and Little rarely, if ever, explicitly argue for non-combinatorialism. In fact, matters are even worse than that. It is not as if they merely forgot to extend their arguments to non-combinatorialism. On the contrary: I will now argue that attempting to extend those arguments to non-combinatorialism about reasons for action is hopeless once one has already embraced holism about reasons for action.

Consider first the argument from a general holism. Recall that the idea behind this argument was that, because holism about reasons for belief and about ordinary reasons for action are (supposedly)

universally accepted views, there is pressure toward accepting holism about reasons for action, lest one endorse an unattractively bifurcated account of the logic of reasons. So one might ask: can we argue in a similar way that almost everyone accepts non-combinatorialism about reasons for belief and about ordinary reasons for action, so that we should find it natural to also accept non-combinatorialism about moral reasons for action? As with the argument for holism about reasons for action, the most pressing problem with this argument occurs at the first step, where one insists that non-combinatorialism about reasons for belief is a universally held view. If we take the relevant overall epistemic verdict about a given belief that *p* to be "one has sufficient reason to believe that *p*," then non-combinatorialism about reasons for belief can be formulated as follows:

> *non-combinatorialism about reasons for belief*: The combinatorial function that determines whether one has sufficient reason to hold a given belief from the valence and weight of the reasons for and against that belief is not finitely expressible (and so in particular, not additive).

However, it should be clear that this thesis is quite controversial, so that one can hardly gesture at its (alleged) universal acceptance as a way of avoiding having to provide a substantive arguments in its favor. Indeed, the very notion of a defeater, which we discussed in light of Dancy's red/blue-switching case, was introduced into epistemology precisely as a way of capturing in finite terms what is required for one to have sufficient reason to hold a given belief in a given context.

Non-combinatorialism about reasons for belief is controversial; but when we conjoin that view with holism about reasons for belief (as particularists must if they want to use the argument from a general holism/non-combinatorialism to secure both portions of particularism about reasons for action), then the result is more controversial still. Particularists such as Dancy and Little think that the conjunction of holism and non-combinatorialism about reasons for action directly entails particularism about moral principles as well as the particularist moral epistemology. But then it would seem that these particularists must also think that the conjunction of holism and non-combinatorialism about reasons *for belief* directly entails a kind of particularism about *epistemic* principles as well as a kind of *general* particularist epistemology. Such a particularism about epistemic principles would hold that there are no substantial, finite, exceptionless epistemic principles, while a general particularist epistemology would imply that the

only way we can come to know what we ought to believe in a given situation is by "directly perceiving" the epistemically relevant features of that situation without any recourse to general epistemic principles of any kind. However, very few practicing philosophers adhere to such radical views about the nature of either epistemic principles or our access to what we ought to believe: witness the widespread acceptance of such substantial, finite, and seemingly exceptionless epistemic principles as that knowledge implies belief or that knowledge implies truth, or the complete and utter lack of any contemporary defenders of a general particularist epistemology. So once one has endorsed holism about all normative reasons, non-combinatorialism about reasons for belief can hardly serve as an uncontroversial assumption from which one can go on to prove other claims, but rather must be *argued for*, perhaps by providing intuitive examples in support of it.[51]

Which brings us to Dancy's other positive argument for holism, namely the argument from examples. But here the chances of successfully extending the argument to one in support of non-combinatorialism are even worse than they were for the argument from a general holism. It is tempting to think that examples can be found which support non-combinatorialism, and perhaps some can. However, such examples (if, indeed, they exist) will be utterly useless if one's ultimate goal is to establish *both* halves of particularism about reasons for action, for the following reason: once one accepts holism about reasons for action, it then becomes virtually impossible to argue for non-combinatorialism about reasons for action via an appeal to examples.

The idea behind this claim is as follows. Suppose we wanted to establish that there is no additive combinatorial function specifying how the valences and weights of the reasons for and against action present in any possible context combine to yield the rightness and wrongness of the actions available to the agent in that context. In order to do this, we would need to establish that what it would be right or wrong to do in a given situation cannot be determined by simply adding up the weights of the reasons for

---

[51] Some particularists might respond by insisting that although particularism about reasons for action leads to particularism about moral principles and the particularist moral epistemology, particularism about reasons for belief does not in fact lead to a particularism about epistemic principles or to a general particularist epistemology. However, doing so would undermine the case for premise 3 of the argument from a general holism, for once one has highlighted such a disanalogy between the epistemic and moral realms, one can hardly argue that there is a strong presumption in favor of reasons for action behaving just as reasons for belief do.

and against each available action and then assigning a verdict ("right," "wrong," or "neither") to each action based on the comparative values of those totals. The standard method of showing that this is not the case by means of examples would be to find two situations in which all the relevant reasons but one are "held constant" (i.e. have the same valence and weight), and then to argue that the dependence between the one varying reason and what it would be right or wrong to do in the two situations is not strictly additive.[52] However, if holism about reasons for action is true, the behavior of a given reason for action potentially swings and shifts with every change in context. But then it seems we can never be assured that all the reasons other than our one selected "independent variable" are being held constant between the two cases, so this entire method is hopeless.[53] And matters are even worse when we attempt to argue from cases that there is no finitely expressible combinatorial function of a non-additive form. Conclusion: the prospects for securing non-combinatorialism about reasons for action through an appeal to cases while at the same time endorsing holism about reasons for action are dim at best.

8. Do Particularists Have a Coherent Notion of a Reason for Action?

So far I have restricted myself to criticisms of the positive arguments that have (or could be) offered by particularists in support of their particularism about reasons for action. But the results of the previous section point the way to a deeper problem for the view. There I argued that it is impossible to use either of the main particularist arguments to establish *both* portions of the twin-doctrine that is particularism about reasons for action. The argument from a general holism/non-combinatorialism cannot be used to establish both portions of the view: since the combination of holism and non-combinatorialism about reasons for belief has extremely radical consequences, one can hardly say that no one has ever thought of the denying their combination, and from there argue that we should hold analogous views about the nature of reasons for action on the grounds that there is a presumption in favor of reasons for belief and reasons for action

---

52 For example, this is how Kagan argues in his (1988) that there is no additive combinatorial function.

53 A particularist might reply that she can ensure that all but one reason are "held constant" in the two situations by intuiting that the valence and weight of all but one of the reasons are the same in both situations. However, while it might be conceded to the particularist that our moral intuitions reliably track the valence of reasons from context to context and perhaps can provide a general approximation to their weight, it seems rather excessive to claim that our powers of intuition are so great as to be able to discern the *exact* weight of a given reason in a given context, which is what would be needed for the argument to go through.

behaving in a similar way. And the argument from cases cannot be used to establish both portions of the view, either: once holism about reasons for action is in place, it is virtually impossible to show that the combinatorial function for rightness and wrongness is not of a given form by comparing two situations in which all relevant moral factors but one are held constant, so it becomes excessively difficult to support non-combinatorialism through an appeal to examples. Thus, while the particularists' two main arguments might have had some chance (however slim) of establishing one half of particularism about reasons for action, once that half is in place, the chance of those arguments establishing the other half all but vanishes. All of this strongly suggests that there is a kind of internal tension between the holist and non-combinatorialist portions of the particularists' views about the nature of reasons for action. And the source of this tension, I believe, is the following: once they have embraced both holism and non-combinatorialism about reasons for action, *then it is no longer clear that particularists have available to them a coherent notion of a reason for or against action.*

Recall the three-level generalized weighing framework within which particularists formulate their claims about how reasons work: according to that framework, certain non-normative features of the actions available to an agent provide reasons of various valences and weights for and against performing those actions (the dependence of the contributory level on the underlying level), and the overall moral status of the available actions is determined in virtue of the valences and weights of those reasons (the dependence of the overall level on the contributory level). Thus within this framework, the contribution of an individual reason for action to the entire system is exhausted by two roles it plays: (i) some non-normative feature of an action gives rise to that reason, and (ii) that reason counts one way or the other toward the rightness or wrongness (as well as other overall moral verdicts) of the action bearing that feature. However, particularists insist that each of these defining roles is inscrutably context-dependent. As particularists see it, reasons for action are context-dependent both "from below" and "from above": context-dependent "from below" since whether a given non-normative feature gives rise to a reason for or against action varies from context to context (holism about reasons for action), and context-dependent "from above" since how the reasons present in a given circumstance combine to determine the overall

moral status of the actions available in that circumstance varies from context to context (non-combinatorialism about reasons for action). But then particularistic reasons for action would appear to be free-floating cogs in the normative machinery, and it becomes difficult to understand what particularists even mean when they call something a "reason for action" or a "reason against action."

According to one common conception, a reason for action is a consideration that would decisively count in favor of a given action were no other reasons present. So on this view, which we might call the *isolation conception of a reason for action*, the fact that action X would have feature F in circumstance C qualifies as a reason *for* performing X in C if and only if, in any possible situation in which F is the only morally relevant feature of the actions available to the agent, the actions possessing F are the *right* thing to do: when no other moral considerations are present, a reason for action "carries the day." Similarly, on this conception, the fact that action Y would have feature G in circumstance D qualifies as a reason *against* performing Y in D if and only if, in any possible situation in which feature G is the only morally relevant feature of the actions available to the agent, the actions possessing G are the *wrong* thing to do. However, it just falls out of this conception of what it is to be a reason for or against action that the holist half of particularism about reasons for action is false. Suppose the fact that a given action would be a telling of a lie is a reason against performing that action in some circumstance C. Then given the isolation conception, it follows that in any situation in which the property of being a telling of a lie is the only morally relevant feature of the actions available in that situation, those actions that involve telling a lie are wrong. But this in turn implies that the fact that an action would be a telling of a lie is a reason against performing that action in *every* circumstance, not just C, for in every circumstance it is now the case that, were the property of being a telling of a lie the only morally relevant feature of the actions available to the agent, the actions possessing that property would be wrong. Thus the isolation conception entails, contra the claims of holism, that whenever some feature provides a reason against action in one possible context, that feature must provide a reason against action in *every* possible context.[54] For this reason, particularists

[54] Note that, as stated, the proposed definition also implies the falsity of *weak* holism about reasons for action, which in §4 I argued is trivially true. Therefore if a non-particularist wishes to endorse the isolation conception, the proposed definition should be revised so as to apply only to *basic* reasons for or against action (and similarly for the other conceptions to be discussed).

who endorse holism cannot accept the isolation conception of a reason for action.

According to a second possible conception, which we might call the *removal conception of a reason for action*, a reason for action is a consideration whose removal would make the action in question less right, and a reason against action is a consideration whose removal would make the action less wrong. More precisely, on this conception the fact that action X would have feature F in circumstance C is a reason *for* (or *against*) action X if and only if, for any sufficiently similar circumstance C′ in which X, if performed, would have all the same morally relevant features except F and no additional morally relevant features, X is less *right* (or less *wrong*) in C′ than in C. In a sense, then, this proposal is the converse of the previous one: in order to determine whether a given feature of an action gives rise to a reason for or against performing that action, we remove *only the feature in question* and see how the rightness or wrongness of the action varies, rather than remove *every other* morally relevant feature. Now there are difficulties making precise how all of this should go. For example, the proposal implicitly assumes that rightness and wrongness come in degrees (which one might contest), and something needs to be said about how in general one is to strip away the feature being evaluated while holding all other morally relevant features constant. (It is clear enough what one is to do when determining whether the property of, say, producing pleasure gives rise to a reason in favor of action—simply make the action less pleasurable and see if the rightness of the action in turn diminishes—but in most other cases it is far less clear how to "cleanly excise" the property being assessed.) However, we can sidestep these issues, because regardless of how they are resolved, the particularists' holism prohibits them from being able to make use of the removal conception.

The whole point of holism about reasons for action is that the valence of the reason provided by a given feature of an action is determined by other features *of the case at hand*, not by features of the action (such as its comparative rightness or wrongness) in certain counterfactual situations. Holism holds that in a counterfactual situation in which we remove only the feature being scrutinized, this change might lead to any number of other changes in the valences and strengths of the reasons provided by other features of the available actions; thus, given holism, the method invoked by the removal conception will be unable to isolate the individual contribution made by the specific feature being considered. The following example is

used by Dancy to illustrate this very point:

> [C]onsider a case in which I am thinking of doing something for a friend. My action, were I to do it, would be good, and partly good because it is an expression of our friendship. But now, if I were to be doing the action and not doing it for a friend, I would presumably be doing it for someone who is not a friend, and it might be that doing it for someone who is not a friend is even better than doing it for a friend. . . . [O]ur friendship seems to be a reason to do the action even though if we were not friends I would have even more reason to do it.[55]

The example is intended to work as follows: in one situation the fact that an action would be one done for a friend might be a reason to perform that action, even though one has more overall reason to perform the action (and hence, the action is more right) in some sufficiently similar situation whose only relevant difference is that the person in question is not one's friend. Now Dancy's interpretation of this particular pair of cases is controversial, but even if one rejects that interpretation the main point remains: holists about reasons for action will want to allow for the possibility that pairs of cases might exist that have the general structure Dancy alleges these two cases to have; the removal conception rules out the possibility of such cases from the outset; so holists about reasons for action cannot avail themselves of this conception of a reason for action.

According to a third possible conception, a reason for action is a consideration that counts in favor of a certain course of action being the right thing to do, and a reason against action a consideration that counts against. So on this story, which we might call the *right-making conception of a reason for action*, a reason *for* performing action X contributes *positively* toward X being right in the given circumstance, and a reason *against* performing action X contributes *negatively* toward X being right in the given circumstance (that is, contributes *positively* toward X being *wrong* in that circumstance). However, particularists cannot hold this conception of a reason for action, either. The problem is that talk of a reason being a consideration counting in favor of or against a certain action being the right thing to do makes the most sense when there is an *additive* combinatorial function; the more we deviate from an additive combinatorial function, the more obscure such talk becomes.

In order to show why this is so, it will help if we get a little more formal. Suppose there are only two non-normative features of actions that give rise to reasons for or against action: F_1 and F_2. If X is an

---

[55] Dancy (2004), p. 20.

action available to an agent in circumstance C, let r_1(X, C) be a real number representing the valence and weight of the reason, if any, provided by X's possessing feature F_1 in the following manner:

- if X would possess F_1 if performed in C, and that fact is a reason *for* performing X in that context, then r_1(X, C) > 0 and its absolute value represents the weight of the reason so provided;
- if X would possess F_1 if performed in C, and that fact is a reason *against* performing X in that context, then r_1(X, C) < 0 and its absolute value represents the weight of the reason so provided;
- otherwise, r_1(X, C) = 0.

Moreover, let r_2(X, C) represent in a similar way the valence and weight of the reason, if any, provided by X's possessing feature F_2 if performed in C. Then if there is an additive combinatorial function, we can represent the *total reason* in favor of action X in circumstance C as follows:

$$t(X, C) = r_1(X, C) + r_2(X, C).$$

Finally, let the actions available to the agent in circumstance C that maximize this function be the *right* ones to do in that circumstance, and let all other available actions be the *wrong* ones to do.[56]

When we determine the rightness or wrongness of actions from the valences and weights of the relevant reasons in this way, it is readily transparent why reasons for a given action are considerations that count in favor of that action's being right and reasons against are considerations that count in favor of it's being wrong. Since r_i(X, C) is always positive when a reason for action is provided by X's possessing F_i, reasons in favor of performing X always *add* to the total reason in favor of X in C. Since r_i(X, C) is always negative when a reason against action is provided by X's possessing F_i, reasons against performing X always *subtract* from the total reason in favor of X in C. So when the combinatorial function is additive, it is perfectly natural to talk about a reason in favor of some action being an individual contribution toward its rightness, and a reason against that action being an individual contribution toward its wrongness.

For some ways of deviating from a strictly additive combinatorial function, the naturalness of this

---

[56] Thus, if I is an index set that ranges over the indices for all the actions X_i available to the agent in circumstance C, the *combinatorial function* can be represented in its full glory as follows:

$$f(X_i,C)=\begin{cases} right & \text{if } (\forall j \in I)\left[t(X_i,C) \geq t(X_j,C)\right] \\ wrong & \text{otherwise} \end{cases}$$

way of talking is preserved. For example, suppose there are still only two non-normative features, F_1 and F_2, that ever give rise to reasons for or against action, but the total reason in favor of action X in context C is instead determined as follows:

$$t'(X, C) = [r_1(X, C)]^3 + r_2(X, C).$$

As before, the right actions are those that maximize the total reason function, and the wrong actions those that do not. And since $[r_1(X, C)]^3$ is always positive when $r_1(X, C)$ is positive and always negative when $r_1(X, C)$ is negative, it still makes sense to talk about the reasons of positive valence provided by feature F_1 counting in favor of the rightness of the action possessing that feature and the reasons of negative valence provided by feature F_1 counting against the action's rightness. However, suppose instead that the total reason function is determined like so:

$$t''(X, C) = [r_1(X, C)]^2 + r_2(X, C).$$

Now we have trouble. If the fact that X would have feature F_1 is a reason against performing X in C, then although $r_1(X, C)$ is negative, $[r_1(X, C)]^2$ is positive, so that fact *adds* to the total reason in favor of X and hence counts *in favor of*, not *against*, X being right in the given circumstance. In other words, given this way of combining reasons to determine overall rightness and wrongness, whenever feature F_1 provides a reason against action in a given context, that feature is actually *right-making* in that context. For this reason we are no longer operating with the right-making conception of a reason for action. And a similar point applies if instead the total reason function is a multiplicative function like the following:

$$t'''(X, C) = r_1(X, C)\, r_2(X, C).$$

Suppose X would have feature F_1 if performed in C, and that fact is a reason for performing X in C. Then $r_1(X, C)$ is positive, but whether this positively or negatively impacts the total reason in favor of X in C now depends on the sign of $r_2(X, C)$, and hence depends on the valence of the reason (if any) provided by the *other* feature, F_2. In this case, talk of a reason for action as being a consideration that always counts in favor of the rightness of that action has become strained at best.[57]

---

[57] Kagan (1988), p. 17, makes much the same point when he notes that without what he calls the "additive assumption" (in effect, the assumption that there is an additive combinatorial function), any talk of "contributions" made by individual morally relevant factors seems out of place.

I take these examples to show that one cannot hold the right-making conception of a reason for action while allowing that the combinatorial function determining rightness and wrongness can have any old form whatsoever. In order to make sense of the idea that reasons for action always contribute toward an action's rightness and reasons against always contribute toward its wrongness, we need the combinatorial function to be such that (i) individual reasons always make discernible individual contributions to the overall rightness or wrongness of a given action, and (ii) the individual contribution made by a reason of positive valence always *positively* impacts the total reason in favor of the action in question, and the individual contribution made by a reason of negative valence always *negatively* impacts the total reason in favor of the action. Let us call the combinatorial function *quasi-additive* if, like the combinatorial function constructed from the total reason function $t'(X, C)$ above, it satisfies these two criteria.[58] It is important to notice that the vast majority of possible combinatorial functions are neither additive nor quasi-additive.[59] However, those who endorse the right-making conception must hold, on pains of having an incoherent notion of a reason for action, that the combinatorial function is, if not additive, then at least quasi-additive. It follows from the particularists' non-combinatorialism that the combinatorial function is not additive. Moreover, it is not clear what grounds particularists have to insist that, although we don't know enough about the combinatorial function to be able to write it down in finite form, we do know enough about it to know that it is quasi-additive. This puts particularists who endorse both non-combinatorialism and the right-making conception in a precarious position: the

---

[58] Note that on this way of characterizing what makes a combinatorial function quasi-additive, all additive combinatorial functions also count as quasi-additive.

The definition of quasi-additivity provided here is admittedly vague, but it will do for present purposes. Indeed, I suspect it is not possible to be any more precise in characterizing quasi-additivity without making various assumptions about the combinatorial function that would not be accepted by all moral theorists. For example, if we can represent the valence and weight of every distinct reason by a variable r_i in the way described in the text, and if the combinatorial function can be calculated in terms of a total reason function, $t(r_1, r_2, \ldots r_n)$, that is a function of the reason-variables, and if that total reason function is differentiable at every point with respect to each of those variables, then I put forward that the combinatorial function is quasi-additive if and only if the following obtains for each i such that $1 \leq i \leq n$:

$$\frac{\partial t}{\partial r_i} \geq 0 \text{ for all values of } r_1, r_2, \ldots r_n.$$

However, this precise a formulation of quasi-additivity comes at the cost of some fairly substantial—and controversial—assumptions about the nature of the combinatorial function. (It is interesting to note that, given this formulation of what it takes for the combinatorial function to be quasi-additive, the requirement that the combinatorial function be quasi-additive is equivalent to the requirement that reasons for action satisfy the removal conception.)

[59] This point is only bolstered if we do not assume, as I did in the examples of the past few paragraphs, that the weight of a given reason can be represented by a real number.

coherence of their very notion of a reason for action depends on there being a quasi-additive combinatorial function, but it is difficult to see what means they have of establishing that there is one without conceding that the combinatorial function might, after all, be finitely expressible. Conclusion: because of their commitment to non-combinatorialism, particularists are not entitled to hold the right-making conception of a reason of action.[60]

Could particularists resist this line of argument by insisting that it is an analytic truth that the combinatorial function is quasi-additive? I doubt it. Even if we ignore Quinean concerns about analyticity, the purported analytic truth is far more recherché than the sorts of things that are usually held to be true merely in virtue of the meaning of the terms involved.[61] Moreover, if we allow that we can know by studying the meaning of the relevant terms that the combinatorial function satisfies one constraint (that of being quasi-additive), this raises the question: why can't we also know by studying the

---

[60] *Objection*: It is possible to construct a combinatorial function that is quasi-additive but not finitely expressible by splicing together several quasi-additive combinatorial functions in an uncodifiable manner. For example, particularists could insist that in some contexts the total reason function is given by $t(X,C)$, that in other contexts the total reason function is given by $t'(X, C)$, and that there is no cashing out in finite terms when one of these two total reason functions applies. As each of these total reason functions satisfies the conditions for quasi-additivity, it follows that the resulting combinatorial function will be quasi-additive but not finitely expressible.

Reply: It is a mistake to assume that just because the combinatorial function constructed from total reason function $t_1(X,C)$ is quasi-additive and the combinatorial function constructed from total reason function $t_2(X,C)$ is quasi-additive, therefore the combinatorial function constructed from the following total reason function must be quasi-additive:

$$t_3(X,C) = \begin{cases} t_1(X,C) & \text{if such-and-such conditions obtain} \\ t_2(X,C) & \text{otherwise} \end{cases}$$

For example, in the way already explained, we can construct an additive (and hence quasi-additive) combinatorial function from the following total reason function:

$t(X, C) = r_1(X, C) + r_2(X, C)$.

Moreover, in a similar way we can construct an additive (hence quasi-additive) combinatorial function from the following total reason function:

$t''''(X, C) = r_1(X, C) + r_2(X, C) - 500$.

However, a combinatorial function constructed in the same way from the following total reason function is *not* quasi-additive:

$$t^*(X,C) = \begin{cases} r_1(X,C) + r_2(X,C) - 500 & \text{if } r_1(X,C) = 10 \\ r_1(X,C) + r_2(X,C) & \text{otherwise} \end{cases}$$

After all, $t^*(X,C)$ is the same function, in the mathematical sense (see n. 22), as the following non-quasi-additive total reason function:

$$t^{**}(X,C) = \begin{cases} -49\, r_1(X,C) + r_2(X,C) & \text{if } r_1(X,C) = 10 \\ r_1(X,C) + r_2(X,C) & \text{otherwise} \end{cases}$$

Therefore patching together several quasi-additive combinatorial functions does not always yield a quasi-additive combinatorial function, and so the objection fails.

[61] After all, recall (n. 58) how difficult it was to even characterize quasi-additivity. If it is not even clear how to precisely formulate quasi-additivity without making contentious assumptions about the nature of the combinatorial function, how could we possibly derive analytic platitudes merely from the meaning of the term "quasi-additive"?

meaning of the relevant terms that the combinatorial function satisfies various other constraints, enough of which might reduce it to finite form? Recall that the majority of possible combinatorial functions are not quasi-additive. So to require that the combinatorial function be quasi-additive is a fairly demanding restriction—a fairly sizeable way of cutting down the space of possible combinatorial functions. Why, then, can't we appeal to more analytic truths to cut down the space of possible combinatorial functions even further? And why couldn't doing so eventually result in a finitely expressible combinatorial function? To insist that we can't reduce the combinatorial function to finite form but can know that it must be quasi-additive is like insisting, of some number with an infinite, non-repeating decimal expansion, that we can't reduce that number to a finite formula, but we can know that a "1" never directly follows a "0" in its infinitely long decimal expansion.[62] Or closer to home, it is like insisting that, although we can know from the meaning of the terms involved that some principle of impartiality holds, that is the only non-trivial, exceptionless constraint we can know about rightness and wrongness.[63] *How* we come to know that the combinatorial function must be quasi-additive is beside the point; what matters is that *if* we accept the right-making conception and its subsequent requirement that the combinatorial function be quasi-additive (whether through an appeal to analytic truths or on some other grounds), *then* this puts pressure on the particularists' adherence to non-combinatorialism.

According to a fourth and final conception of what it is to be a reason for action, which we might call the *favoring conception*, a reason for action is a consideration that counts in favor of a given action, and a reason against action one that counts against.[64] The difference between this and the previous conception is that, whereas with the right-making conception we said that a reason for performing action X is a consideration *that counts in favor of X being the right thing to do*, here we are saying that a reason for performing X is a consideration *that counts in favor of X* (full stop). It is tempting to think that at the end of the day there

---

[62] Of course, we can prove similar things about some numbers (like π and e) that have infinite, non-repeating decimal expansions, but in those cases there is always some way of expressing the number in question in a finite formula.

[63] Note I am not claiming that it is plausible that a principle of impartiality is an analytic truth; rather, what I am claiming is that if such a principle *were* an analytic truth, then it would be odd if there were no other analytic moral truths of that sort.

[64] Raz (1990), p. 186: "The reasons for an action are considerations which count in favor of that action"; Scanlon (1998), p. 17: "Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favor."

is little difference between these two conceptions of a reason for action. However, some philosophers—including many particularists—make a distinction between the *favoring relation* and the *right-making relation*. Dancy makes the point as follows (although in this passage he focuses on the alleged difference between the favoring relation and the ought-making relation, the same reasoning, if it works, can establish a difference between the favoring relation and the right-making relation):

> [I]t seems to me that we are in fact dealing with two normative relations rather than one. The first is the relation between reasons and ought-judgments; we specify the reasons, and pass to the judgments that we ought to act. The second is a relation between reasons and action which is not necessarily mediated by any ought at all; it is the one that is in play when we engage in the sort of practical reasoning whose "conclusion" is an action. I don't always think, "There is this reason for jumping; so I ought to jump"; sometimes I just think, "There is this reason for jumping; so I'll jump." Crucially, the relation between reason and ought-judgment is different from the relation between reason and action. And it is really the latter that we are after when we try to understand the notion of a reason for action—a practical reason.[65]

Switching from talk of ought-making to talk of right-making, the idea here is that the favoring relation is a relation that holds between a reason and an action, whereas the right-making relation is a different normative relation that holds between a reason (or more accurately, a reason-giving feature) and an action's rightness or wrongness. However, note that once one endorses the generalized weighing model of morality (as particularists do), then feature F is a right-making feature of action X in circumstance C if and only if the fact that X would have F if performed in C is a reason for performing X—and so, on the favoring conception, counts in favor of X. Thus, even if technically the favoring and right-making relations are different relations, within the generalized weighing framework the favoring and right-making conceptions of a reason for action amount to much the same thing.

However, suppose I am wrong here, so that the favoring conception is, in fact, a distinct conception of what it is to be a reason for action; even then the particularists' non-combinatorialism bars them from being able to hold the favoring conception. The idea behind this claim is as follows. Let us suppose there is a distinct type of practical reasoning that ends in an action, rather than a theoretical conclusion about what it would be right to do (or what one ought to do). Then, given the generalized weighing framework, which action is the correct conclusion of that piece of practical reasoning is

---

[65] Dancy (2004), p. 20. Dancy later makes it clear (pp. 78-79) that he intends there to be an analogous difference between the favoring relation and the right-making relation. This idea that there is a distinctive type of practical reasoning whose conclusion is an action, though Aristotelian in origin, has its modern roots in Anscombe (1957); for criticism of the idea, see Jarvis (1962).

determined by the weight and valence of the reasons that are (or should be) appealed to in that bit of reasoning. So, just as there is a combinatorial function that takes as input the valences and weights of the reasons present in a given situation and outputs the rightness and wrongness of the actions available in that situation, there is a combinatorial function that takes as input the valences and weights of the reasons that are (or should be) appealed to in a given piece of practical reasoning and outputs the action (or actions) that are the correct conclusions of that reasoning. And just as particularists hold that the first of these combinatorial functions is not finitely expressible, they no doubt would also hold that the second of these combinatorial functions is not finitely expressible—and in particular not additive in form. So we can run an argument exactly parallel to the one used to show that particularists cannot hold the right-making conception in order to establish that they cannot hold the favoring conception, either.

All of which raises the question: what exactly do particularists think reasons for action are? Their holism about reasons for action blocks them from being able to hold the isolation or removal conceptions of a reason for action, and their non-combinatorialism blocks them from being able to hold the right-making or favoring conceptions. But then the particularists' notion of a reason for action is mysterious indeed.[66]

9. A Newtonian Analogy

An analogy will help illustrate just how mysterious the particularists' notion of a reason for action really is. We can think of the model of reasons that the particularists oppose as being akin to the following simplified version of Newtonian classical mechanics. Suppose we have a number of massive point particles in empty space that are interacting with each other purely through the classical law of gravitation. Then each particle exerts a force on every other particle that is determined solely by the masses of the two

[66] Does the argument just given assume that one must be able to give a *reductive* account of reasons for action? No, it does not. I was intentionally silent about whether the proposed conceptions involved reducing reasons to more fundamental notions, or whether they allowed for the possibility that some of the terms in the account of a reason might not be definable except in terms of a reason. Thus even if, following Scanlon, one endorses the favoring conception of a reason but insists that the favoring relation can only be cashed out in terms of the being-a-reason-for relation (so that one is a primitivist about both reasons and the favoring relation), it still follows that one is not entitled to adhere to the favoring conception unless the combinatorial function is quasi-additive, and that is all I needed in order to argue that particularists cannot hold the favoring conception.

particles and their relative positions.[67] So in *any* situation in which a particle of mass m_1 is at coordinates (x_1, y_1, z_1) and a second particle of mass m_2 is at coordinates (x_2, y_2, z_2), the *individual force* exerted on the first particle by the second is the same in both magnitude and direction, regardless of the mass and position of any other particles that may or may not be present in that situation. This is analogous to the claim, denied by holism about reasons for action, that there are certain features of actions that always give rise to an *individual reason* for or against performing the action bearing that feature, regardless of what other features may or may not be present. Moreover, in our Newtonian model, the *total force* acting on a given particle is determined by a vector sum of the individual forces acting on that particle due to every other particle in the situation. This is analogous to the claim, denied by non-combinatorialism about reasons for action, that we can determine the *total reason* in favor of each action by adding up the weights of the individual reasons for and against performing that action.

We can easily imagine what a particularistic version of this Newtonian model would look like. As before, we have various massive point particles moving around in an empty three-dimensional space. However, for any given configuration of particles, the individual force acting on one particle due to another is not given by any general formula that holds regardless of the positions and masses of the other particles. Suppose that in one configuration there is a certain individual force acting on a particle of mass m_1 at coordinates (x_1, y_1, z_1) due to a second particle of mass m_2 at coordinates (x_2, y_2, z_2). Then there is no assurance that in any other configuration in which those two particles have the same mass and are at the same position, the individual force acting on the first particle due to the second is the same: depending on the positions and masses of the various other particles in the configuration, the first might have *no* force exerted on it by the second, or it might even have a force in the *opposite* direction exerted on it. This is the holist element in the model. But we can also build in a non-combinatorialist element, so that there is no finitely expressible formula, additive or otherwise, for determining how the individual forces being exerted on a given particle combine to yield the total force that the particle is subject to. In some configurations

---

[67] More precisely: the force acting on some particle 1 due to another particle 2 always points in the direction of particle 2 and always has a magnitude of $F = Gm_1m_2/r^2$, where m_1 is the mass of particle 1, m_2 is the mass of particle 2, r is the distance between the two particles, and G is the universal gravitational constant.

the total force acting on a particle might be a vector sum of the individual forces acting on it, but in other configurations the total force might be a cross product of the individual forces acting on the particle, or a cross product added to another cross product, or something even more complicated, depending on the intricacies of the case at hand.

However, it should be clear that, once we include both the holist and the non-combinatorialist ingredients in our Newtonian model, we begin to lose sight of what this notion of an individual force even amounts to. What does it mean to say that, in this model, one particle exerts an individual force in a given direction on a second particle? It doesn't mean that if no other forces were acting on the second particle, then it would accelerate in the direction of the individual force, because (due to the holist element in the model) the nature of the individual force acting on the second particle is dependent upon the other forces acting on it as well. It doesn't mean that if the individual force were not present, then the second particle would tend to accelerate in the direction opposite to the individual force, because (also due to the holist element in the model) the other individual forces acting on the second particle could potentially change if the individual force due to the first were removed. And it doesn't mean that there is a contribution in the individual force's direction to the total force acting on the second particle (a contribution that "counts in favor" of the total force being in that direction, as we might put it), because (due to the non-combinatorialist element in the model) the total force acting on a particle is not always a vector sum of the individual forces acting on it. So what, then, does this notion of an individual force being exerted on a particle amount to? And similarly, in case of particularism about reasons for action, what does this notion of a reason for (or against) a given action amount to, if it doesn't mean that were no other morally relevant considerations present it would be right (or wrong) of the agent to act in that way, and if it doesn't mean that were the consideration in question removed then the given action would be less right (or less wrong), and if it doesn't mean that the individual reason is a contribution counting in favor of (or against) acting in that way, or in favor of (or against) that action's being the right one to do?

Thus once we build in the holist and non-combinatorialist elements into our Newtonian model, we lose our grip on what is meant by an individual force exerted on one particle by another. Similarly, I

maintain, once particularists build in their holist and non-combinatorialist elements into the generalized weighing framework for how reasons for action function, we lose our grip on what they mean by a reason for or against action. This problem is particularly serious for particularists, for both of their positive arguments for particularism about reasons for action rely on our having intuitions about whether or not certain reasons, be they for action or for belief, are present in a given situation. However, if we have no idea what these mysterious entities which particularists are calling "reasons" are, how can we settle whether they are present in a given situation by an appeal to our intuitions? So this worry that particularists do not have available to them a coherent notion of a reason for action is especially troubling given their general argumentative strategy.

Finally, not only are particularists in danger of being left with no coherent notion of a reason for action once they have taken on board both their holism and their non-combinatorialism, but moreover, there is a worry that even if particularists could provide a satisfying account of their notion of a reason for action, there would never be any point in appealing to these things particularists call "reasons for action" anyway. To see this, let us return to the particularistic version of the simplified Newtonian model. Suppose this model were a true description of the actual world, so that that the physical world were entirely composed of massive point particles such that (1) the individual force each particle exerts on another is dependent on the position and mass of every single other particle in the universe (holism), and (2) the total force acting on each particle cannot be determined from the individual forces acting on it according to some finitely expressible formula, whether additive or otherwise (non-combinatorialism). In addition, suppose it were possible (perhaps *per impossibile*) for us to construct a device for measuring both the individual and total forces acting on a given particle. If this were the case, then it should be clear that given the particularistic Newtonian framework for the way the laws of nature work, the first use of the measuring device would be absolutely useless. Why would one ever bother measuring the individual force exerted on one particle by some other particle? Because of the holist element in the laws of nature, there is no assurance that the same individual force will be exerted on the first particle by the second in all situations in which the particles have exactly the same mass, position, velocity, and so on. And because of

the non-combinatorialist element in the laws of nature, it is not possible for us to calculate the total force acting on a particle from the values of the individual forces acting on it. So what would be the point in ever measuring the individual forces acting on a given particle?

On the particularists' conception of how reasons for action behave, the moral realm operates just as the physical realm does in the particularist version of the Newtonian model, and the morally astute person is precisely such a device for measuring both the individual and total reasons for action present in a given circumstance. But then what use is the morally astute person's knowledge of the individual reasons for and against action that are present in a given situation? It is not as if knowledge of those individual reasons is a useful stepping stone to working out what it would be right and wrong to do in that situation, for according to the particularists' non-combinatorialism, the individual reasons do not combine to yield our overall moral obligations in a finitely expressible manner, additive or otherwise. And it is not as if the knowledge of which reasons are present in the case at hand could be usefully applied to other cases, for according to the particularists' holism the functioning of a reason in one situation cannot be predicted from its functioning in another. So why even bother making note of the individual reasons for and against action? By positing individual reasons to be so irremediably context-dependent, the particularist seems to have given up the very possibility of practical reasoning and to have rendered all knowledge of reasons for action otiose.[68]

10. Giving Up Non-Combinatorialism? Or Moving Beyond Weighing?

I have just argued that if particularists accept both holism and non-combinatorialism, they are left with no coherent notion of a reason for action. One natural line of response to this argument is as follows: "According to you, particularists can't endorse holism and non-combinatorialism at the same time. But who says they *need* to endorse both theses? Indeed, it seems that particularists can get most of what they

[68] Some particularists might try to deny that the particularistic Newtonian model described in this section is really analogous to a particularistic theory of reasons for action. However, all parties should agree that a *non-particularistic* theory of reasons for action is analogous to the *non-particularistic* Newtonian model. Moreover, I put forward that the ways in which a particularistic theory of reasons for action deviates from a non-particularistic one are exactly analogous to the ways in which the particularistic Newtonian model deviates from a non-particularistic one. It follows that particularists must concede that their theory of reasons is analogous to the particularistic Newtonian model.

want out of their position by accepting *one* of those two theses while denying the other. So in particular, since particularists spend a good deal more time discussing holism than they do discussing non-combinatorialism, why can't particularists maintain their allegiance to holism, but give up any commitment to non-combinatorialism?"

It is indeed open to particularists to adopt such a position, but I doubt that many true particularists would. One of the main forces driving some philosophers to become particularists is a general suspicion about the usefulness and value of the traditional quest for true and exceptionless moral principles, and if holism is true but non-combinatorialism is not, then there remains a significant place in ethics for at least one portion of that traditional quest. Given the three-level generalized weighing model of morality to which particularists are committed, there are three basic categories of moral principles explaining how the facts at one of the model's three levels are determined in general by the facts at one of the more basic levels:

> *underlying-to-contributory principles*: principles specifying how in general the non-normative facts at the underlying level determine the facts about the valences and weights of individual reasons at the contributory level;
>
> *contributory-to-overall principles*: principles specifying how in general the facts about the valences and weights of individual reasons at the contributory level determine the facts about the rightness and wrongness of the available actions at the overall level;
>
> *underlying-to-overall principles*: principles specifying how in general the non-normative facts at the underlying level determine the facts about the rightness and wrongness of the available actions at the overall level.

Thus if particularists give up their non-combinatorialism, they must admit the existence of true and exceptionless *contributory-to-overall principles*. Moreover, debates about which contributory-to-overall principles are the correct ones will be instances of precisely the sort of moral theorizing that particularists are loath to engage in.

For example, suppose particularists concede that there is an additive combinatorial function. That does not yet settle the question of what the proper contributory-to-overall principles are, for there are many different additive combinatorial functions—many different ways of assigning rightness and wrongness to the actions available in a given context on the basis of comparing the total sum of reason in

favor of each. This leaves us with questions like the following. Is the true additive combinatorial function a *maximizing* combinatorial function (so that the right actions in a given context are those with the greatest sum of reason in their favor), or is it a *satisficing* combinatorial function (so that the right actions are those with a suitable amount of total reason in their favor)? If the true combinatorial function involves satisficing, does it involve *absolute-level satisficing* (so that how much total reason in favor of an action makes for rightness is the same in every context), or *comparative satisficing* (so that an action is right in a given context if and only if the total reason in its favor is greater than the total reason in favor of a reasonable percentage of the other available actions)? If the true combinatorial function involves maximizing, how does it handle situations in which the total reason function has no maximum (so that for every action that the agent might perform, there is another available action with a greater sum of reason in its favor)? Should we be worried that a maximizing additive combinatorial function would make morality too demanding (since for almost every action that any of us actually perform, there exists an alternate action with a slightly greater sum of reason in its favor that we could have performed instead)? And so on: these types of questions should look familiar, for they are exactly the sorts of issues that face *consequentialists* when they try to settle, for a given set of available actions, the connection between the total intrinsic value of each action's consequences and the overall rightness and wrongness of those actions.[69] Moreover, just as particularists are dubious of the analogous debates in the consequentialist literature, so too are they likely to be deeply suspicious of this sort of back and forth over the most plausible version of an additive combinatorial function.

Matters are not helped if instead particularists insist that there exists a finitely expressible combinatorial function that is not additive in structure. To see this, consider two examples of works that spend some time searching for a non-additive combinatorial function. In the penultimate chapter of *The Possibility of Altruism*, Thomas Nagel goes through great effort to show that the theory of reasons defended in his book need not be a version of consequentialism, and he does this by proposing various non-additive

[69] On maximizing versus satisficing versions of consequentialism, see Slote (1985), ch. 3. On absolute-level versus comparative satisficing, see Hurka (1990). On the worry that a maximizing version of consequentialism might make morality too demanding, see Kagan (1989).

"combinatorial principles" that specify how the reasons present in a given situation determine one's overall duty—or in other words, does so by proposing various non-additive, finitely expressible combinatorial functions. And in *What We Owe to Each Other*, T. M. Scanlon develops an elaborate contractualist theory of how rightness and wrongness is determined by the reasons present in a given circumstance—or in other words, develops a distinctively *contractualist* combinatorial function. But again, these works are instances of exactly the kind of moral theorizing that particularists are supposed to be suspicious of. What particularists really want is to insist that there is something misguided about Nagel and Scanlon's quest for true and illuminating contributory-to-overall principles, not to engage in a debate with Nagel and Scanlon over the particular merits and faults of their respective positions.

For this reason, I doubt that many true particularists would be content to simply renege on their non-combinatorialism. What recourse does this leave them, then, if they accept what I have argued here? I think a better option—though ultimately a far more radical one—would be for particularists to turn their back on the generalized weighing framework all together. What underwrote that framework was a certain analogy, either with a literal weighing of elements on a scale, or with physical forces acting on particles. Particularists want to move as far from those two analogies as possible while still working within the generalized weighing framework, but if my arguments are correct, there is a certain sense in which those analogies are *essential* to the weighing framework, a certain sense in which talk of reasons combining to determine the overall status of actions *commits* one to preserving those analogies, at least to some degree. Deviate too far from a strict analogy with weighing or with physical forces, and the framework falls apart. So perhaps particularists, if they want to preserve their general skepticism about our ability to limn the true and ultimate structure of the moral realm, should abandon talk of individual reasons, or at least should give up the idea that reasons are the fundamental normative units that determine all other normative properties and relations.

Indeed, there is something truly bizarre about the picture of the moral realm that particularists wind up endorsing at the end of the day. On their picture, we can be completely confident that the proper metaphysics of morals has exactly three layers to it: the underlying level consisting in the non-normative

facts; the contributory level consisting in the facts about reasons and their valences and weights; and the overall level consisting in the facts about whatever normative properties and relations (such as rightness and wrongness) are determined by the interplay of those reasons. Also on this picture, we can be completely confident that we will never be able to fully discern the way in which the second level depends on the first, and the third level on the second. But why should there only be *three* layers to the picture? Why couldn't there be four levels, so that the values of the being-a-reason-for relation at the second level determine the values of some other normative relation at a third level, which in turn determine the overall moral status of the available actions at the final level? And if four, why not five, or six, or seven levels? And once we have started adding new levels to the picture, why aren't we on the road toward explicating the moral world, in the way in which natural science explains molecules in terms of atoms, and atoms in terms of quarks, and quarks in terms of . . . ?

Thus I think it would be more in keeping with the general sentiment behind particularism to cut out the middleman and drop all talk of individual reasons, or at least drop the claim that properties of reasons always determine the rightness and wrongness of actions. If one really wants to be a *holist*, in the true sense of that term, about normativity and claim that what normative properties a given feature has depends on every other feature of the situation at hand (on the *whole* situation, as we might put it), then any attempt to isolate individual normative elements—call them "reasons"—in that situation will be artificial at best; really it is the *entire situation*, not any subportion of it, that gives a particular action the overall normative status that it has. (Recall the claims of holists about confirmation that it is an entire system of belief that is confirmed or disconfirmed, not individual portions of that system.) So perhaps what particularists really want, rather than the sort of three-level view they defend, is a *two-level view* according to which we have first the underlying level consisting of the non-normative facts, and then a second level consisting of whatever normative facts there may be, which obtain in virtue of the facts at the underlying level in an uncodifiable manner.

The problem with adopting this line, though, is that it deprives particularists of the existing arguments for their view—if we give up the generalized weighing framework and its assumptions that

there are such things as individual reasons that determine other normative properties like rightness and wrongness, we lose the particularists' way of establishing (or at least attempting to establish) that there are no true and exceptionless moral principles. But this, I am sometimes inclined to think, is the fundamental problem with the particularist program. The guiding thought behind particularism is that ultimately the normative is incapable of being codified in an illuminating way; but how can we learn enough about the normative to know this without codifying the normative in at least one way? What particularists need is to provide enough structure to the moral realm for their arguments to get a grip, without thereby undermining their eventual conclusion that finite minds such as ours can never completely discern the true nature of that realm, and how to achieve that balancing trick is not easy to see. Thus in the end I suspect that problems analogous to those raised in this chapter will face any attempt to give a positive argument for a view that is true to the spirit of particularism. That, by itself, does not mean that such a view must be false; but it does suggest that if the real position particularists are after is true, there is no way we could ever establish its truth.

Chapter 2: General Processes, Specific Maxims

1. Introduction

Consider the generality problem for process reliabilism. In its crudest form, process reliabilism holds that a belief is justified if and only if the process through which it was formed is reliable (in the sense of being truth-conducive). Now suppose I look out my window one night and form the belief that it is snowing. What is the relevant process through which that belief was formed? Is it the process of forming a belief on the basis of perception? The process of forming a belief about the weather on the basis of visual perception in bad lighting conditions? The process of forming a belief that it is snowing on the basis of such-and-such retinal stimulations at 8:02 PM on Tuesday, February 13, 2007? Depending on which process we choose as the relevant one to test for reliability, process reliabilism yields different answers as to whether my belief is justified. And the more general worry for process reliabilism is that every non-ad-hoc way of specifying the relevant belief-forming process for a given belief leads to counterintuitive results: either some beliefs that intuitively count as justified are deemed to be unjustified, or some beliefs that intuitively count as unjustified are deemed to be justified.

There is a familiar problem for Kantian ethics that is strikingly similar in form to the generality problem for process reliabilism. According to the first formulation of Kant's Categorical Imperative, one should act only in accordance with that maxim which one can at the same time will to be a universal law. Now suppose an axe-murderer comes to my door and says, "I wish to murder your friend—is he here in your house?" If I respond by lying about my friend's whereabouts, what is the relevant maxim in accordance with which I have acted? Is it the maxim "I will tell an untruth in order to deceive my interlocutor"? The maxim "I will tell an untruth about my friend's whereabouts in order to save his life"? The maxim "I will tell an untruth about the whereabouts of Seth Yalcin in order to save him from being axe-murdered at 8:02 PM on Tuesday, February 13, 2007"? Depending on which maxim we choose as the relevant one to test for universalizability, the categorical imperative yields different answers as to whether my action is morally permissible. And the more general worry for Kantian ethics is that every non-ad-hoc way of specifying the relevant maxim for a given action leads to counterintuitive results: either some

actions that intuitively count as permissible are deemed to be impermissible, or some actions that intuitively count as impermissible are deemed to be permissible.

Thus we have two theories, one a theory of epistemic justification, the other a theory of moral permissibility, both of which face a similar problem. And once one notes the generality problem as it arises for reliabilist theories in epistemology and Kantian theories in ethics, one begins to see similar problems cropping up throughout philosophy. In epistemology, it is well known that most (if not all) externalist theories of knowledge and justification face some version of the generality problem; more recently, though, some epistemologists have argued that something like the generality problem arises even for internalist theories of epistemic justification.[1] In ethics, one finds versions of the generality problem besetting crude forms of rule-consequentialism, several versions of contractualism, and some varieties of virtue ethics. Outside of ethics and epistemology, the generality problem for reliabilism bears a striking resemblance to the so-called problem of the reference class for frequentist interpretations of probability. And even Hume's constant-conjunction theory of causation faces a kind of generality problem: if we think that event e_1 causes event e_2 if and only if e_1 is a member of a group of events that are constantly conjoined with members of a second group of events that includes e_2, how do we fix on the relevant groups?

Though I am interested in each of these variants of the generality problem, in this chapter I will be restricting my discussion to the generality problem as it arises for process reliabilism and for Kantian ethics. I have chosen to focus on these two theories because, despite their marked difference in content (if anything, process reliabilism is the epistemic analogue of a kind of rule-utilitarianism, not of Kantianism), both clearly face a version of the generality problem. For this reason, examining these two quite dissimilar theories side by side will help bring out what is really doing the work in generating a generality problem for each. And once we have a better handle on what gives rise to the generality problem in these two cases, we can bring this insight to bear on the other manifestations of the problem.

---

[1] I have heard this line argued in conversation by Ralph Wedgwood, Thomas Kelly, and Juan Comesaña, among others. Wedgwood also briefly argues the point in his (2002), pp. 286-287.

2. The Generality Problem for Reliabilism: An Initial Characterization

Our first order of business is to get clearer on what the generality problem is for both process reliabilism and Kantian ethics. Let us start with process reliabilism (which I hereafter refer to simply as "reliabilism").

The guiding thought behind reliabilism is that, much as a thermometer is a device for determining the temperature in some region, our belief-forming mechanisms are cognitive devices for acquiring true beliefs and avoiding false ones. So, just as a thermometer qualifies as reliable when it tends, in a suitable range of (actual and counterfactual) circumstances, to give the correct temperature, a belief-forming process is taken to count as reliable when it tends, in a suitable range of (actual and counterfactual) circumstances, to yield a high ratio of true beliefs to false ones. But it is here that the generality problem comes creeping in. To yield a ratio of true to false beliefs that is other than 0 or 1, a belief-forming process must be *repeatable*: it must be possible for it to occur on more than one occasion. The specific process *token* by which a given belief *token* is formed is a sequence of datable events—event *tokens*, that is—which either does or does not result in a true belief. That, surely, is not what reliabilists mean to be assessing the reliability of. Rather, what they mean to be evaluating is whether (the instances of) a certain belief-forming process *type* on average result in a high ratio of true to false beliefs in a suitable range of circumstances. But then the generality problem is thrust upon us: when on a given occasion one comes to believe that *p*, the process token by which that belief was formed is an instance of ever-so-many process types; which of these is the relevant process type whose reliability we are to evaluate when determining whether one's belief is justified?

With these distinctions in mind, we can formulate reliabilism as follows:

> *reliabilism*: S's token belief that *p* is epistemically justified if and only if the process token by which it was formed is a process token whose relevant process type is reliable (i.e. results in a sufficiently high ratio of true to false beliefs in a suitable range of circumstances).[2]

The *generality problem for reliabilism*, then, is to specify in a non-ad-hoc manner the relevant process type to

---

[2] Actually, this formulation of reliabilism is not entirely accurate: a more complete formulation would make a distinction between *belief-dependent processes* that take as input at least some other belief tokens, and *belief-independent processes* that do not; see Goldman (1979), p. 117. However, distinguishing between belief-dependent and -independent processes would bring in distracting complications that are largely orthogonal to the main issues raised by the generality problem, so throughout this chapter I will implicitly assume that all beliefs under discussion are formed via belief-independent processes.

test for reliability when assessing whether a given belief token is justified.

It is standard to distinguish the generality problem from two related, but distinct, problems for reliabilism.[3] The first of these we can call the *problem of extent.* This is a problem that arises in determining the boundaries of the process token that results in one's coming to believe a given proposition: how far back in the causal ancestry of the belief should we go? Consider again the case in which, as a result of looking out my window, I form a belief that it is snowing. What is the extent of the relevant process token by which that belief token was formed? Does it begin with the event of the snow falling? With the event of certain photons bombarding my sensory organs? With the event of its seeming to me as if there is snow falling outside my window? Settling on an answer to this question is no trivial matter, but for the purposes of this chapter I will simply assume that reliabilists have an adequate response to it; my main concern lies elsewhere.

A second problem that is standardly distinguished from the generality problem we can call the *problem of the scope of assessment.* Suppose that—having solved the problem of extent—we fix on the relevant process token by which a given belief token is formed, and that—having solved the generality problem—we fix on the relevant process type of which that process token is an instance. Now we wish to assess whether, on average, instances of that process type yield a high preponderance of true beliefs. What range of cases do we consider in order to determine this average? Do we consider every possible situation in which the process type is (or would be) instantiated? Or do we restrict ourselves only to some suitable range of cases, and if so, how do we determine the restriction? As with the problem of extent, settling this question is no trivial matter, but once again I will simply assume that reliabilists have an adequate response to it.[4]

---

[3] In fact, it is not entirely clear that these two problems are distinct from the generality problem, rather than two further manifestations of it. However, even if they are additional aspects of the generality problem, I am going to follow the general practice in the literature of treating them as distinct problems, and set them aside for now. (For more on this issue, see n. 24.)

[4] There is a related problem that is worth mentioning, if only because it is not often stressed. It seems likely that any solution to the problem of the scope of assessment will result in there being an *uncountably infinite* number of situations in which the token instance of the relevant process type results in a true belief, and an *uncountably infinite* number of situations in which the token instance of the relevant process type results in a false belief—even if, intuitively, there are "many more" of the former type of case than the latter. The additional problem, then, is to figure out how to determine a precise ratio when we are comparing infinities in this manner. (In their casual talk of the ratio of cases in which a process yields a true belief to those in which it yields a false one, reliabilists seem to be assuming that the space of possible worlds is a *measure space*, but this is a highly nontrivial assumption.)

To get a feel for how difficult it is for reliabilists to solve the generality problem, let us consider one attempt at a solution. Taking a hint from the naturalistic aspirations of reliabilism, one might propose that the relevant type for any belief-forming process token is the natural kind to which it belongs. However, this will not do. As Earl Conee and Richard Feldman point out, "every belief-forming process token is categorized in multiple ways by laws in each of several sciences," so each process token is an instance of an abundance of natural kinds.[5] For example, in the case in which I come to believe that it is snowing, the process token by which my belief is formed is plausibly an instance of the following natural kind process types: physical process, electrochemical process, perceptual process, visual process resulting in a belief, visual process resulting in a belief characterizing the environment around me, and so on. There is no single process type that we can single out as *the* natural kind to which the process token belongs, and hence appealing to natural kinds provides no solution to the generality problem.

There have been a multitude of alternate proposals for how to solve the generality problem for reliabilism, but in each case the following pattern has held: the solution is proposed in the literature, and then someone (usually Conee and Feldman) comes along and points out either that the proposal does not isolate a single type for each process token, or else that it does isolate a single type but the type it fixes on yields counterintuitive results. Often the advocates of the solution do not agree that their proposal has these consequences, but I side with Conee, Feldman, and the other critics of reliabilism in holding that, at least so far, no successful solution to the generality problem has been offered.[6] If this is correct, then reliabilism at this point in time is crucially incomplete: until a (defensible) way of fixing on the relevant process type for a given process token is filled in, it is not even clear what the implications of the theory are. And the deeper worry is that, as William Alston insightfully puts it, "the assignment of each token to a unique type has been rigged to fit an antecedent decision as to the epistemic status of the belief, thereby giving rise to the suspicion that reliability is not the most basic determinant of justification after all."[7]

---

[5] Conee & Feldman (1998), p. 10.

[6] For effective criticism of nearly every solution to the generality problem on offer in the literature, see Feldman (1985), Conee & Feldman (1998), Feldman & Conee (2002), and the afterword to the reprint of Conee & Feldman (1998) in their (2004).

[7] Alston (1995), p. 6.

3. The Generality Problem for Kantian Ethics: An Initial Characterization

Such is the generality problem for reliabilism; what, exactly, is the analogous problem for Kantian ethics? My discussion here will focus only on the problem as it arises for one formulation of Kant's Categorical Imperative, namely the Formula of Universal Law, which reads: "act only in accordance with that maxim through which you can at the same time will that it become a universal law."[8] Moreover, in order to avoid getting bogged down in issues of Kant exegesis, I will concentrate my attention on a recently influential reading of Kant due to John Rawls and his students Onora (Nell) O'Neill, Barbara Herman, and Christine Korsgaard, a group I will occasionally refer to as the neo-Kantians.

One of the central tenets of the neo-Kantian interpretation is that, for Kant, *all action is purposive*: action is not mere behavior—not merely a string of observable events in the external world—but rather essentially involves an agent willing that some end or purpose be achieved.[9] A second central neo-Kantian tenet is that *all action is in accordance with maxims*: whenever one acts, one acts in accordance with a subjective principle of volition called a "maxim" that specifies what one is doing, why one is doing it (the end or purpose that one hopes to further through the action), and in what circumstances one would do what one is doing. Thus, on the neo-Kantian interpretation, all maxims can be expressed in the form, "When in circumstances C, I will do act A in order to achieve end E."[10] The term "act" here refers to a piece of mere behavior; following Korsgaard, it will be helpful to reserve the term "action" for the compound consisting of an act, done for the sake of some end, in some circumstances.[11] One need not

---

[8] Kant (1785/1997), AK 4:441 (emphasis suppressed). All references to Kant's work will use the standard practice of referring to the page numbers of the Germany Academy (AK) edition of Kant's complete works.

Some might question my decision to focus on the Formula of Universal Law, since several recent commentators have argued that the Formula of Universal Law is merely a provisional formulation of Kant's supreme moral law that yields indefensible conclusions about moral permissibility. (I am thinking here of Allen Wood in particular; see his (1999), ch. 3.) However, not only do I believe that the Formula of Universal Law plays a more crucial role in Kantian ethical theory—and, in particular, in the *positive arguments* for Kantian ethical theory—than these commentators think, but, moreover, if the object of assessment for both the Formula of Humanity and the Formula of the Kingdom of Ends is the maxim of one's action (as surely must be the case), then every formulation of the Categorical Imperative faces a version of the generality problem.

[9] See Nell (1975), p. 37; Herman (1990), pp. 15-19; Korsgaard (1989a), p. 57; and Korsgaard (1989b), p. 176.

[10] See Rawls (2000), p. 168; Nell (1975), pp. 34-37; Herman (1993c), p. 134; and Korsgaard (1997a), p. xviii. Rawls gives the form of a maxim as "I am to do X in circumstances C in order to bring about Y unless Z," but the "unless Z" clause can be combined with the "in circumstances C" clause to make a more complex circumstances clause "in circumstances C-unless-Z." Korsgaard usually omits the "in circumstances C" clause in her formulation of a maxim, but that clause is needed to avoid the absurd result that, when one imagines the world of the universalized maxim while running the CI-procedure, one must imagine everyone in that world attempting *at every instant of time* to act on the maxim being tested.

[11] Korsgaard (forthcoming-a), lect. 1, p. 11 (although, as mentioned in the previous footnote, Korsgaard usually omits the

explicitly rehearse a given maxim in one's head before acting, but whenever one performs an action—as opposed to a mere act—there is always, on this interpretation of Kant, some maxim in accordance with which one acts.[12]

A final central tenet of the neo-Kantian interpretation is that the object of moral assessment for the Formula of Universal Law is the maxim of one's action: on this reading, the Formula of Universal Law provides a test for the moral permissibility of an action by evaluating that action's maxim in a certain way. When deliberating about whether to do something, one is to run through the following five-step procedure, which Rawls and Herman call the *CI-procedure*:

1. Formulate the *maxim* M of the proposed action. The maxim can be expressed in the form "When in circumstances C, I will do act A in order to achieve end E."

2. Imagine what Korsgaard calls the *world of the universalized maxim* (WUM): a world in which everyone acts according to M. Also imagine yourself in that world, attempting to act on M.

3. If either WUM is inconceivable (the logical contradiction interpretation of this step), or your attempt to act on M in WUM is bound to be frustrated (the practical contradiction interpretation of this step), then M fails the *contradiction-in-conception test.*[13]

4. If there are ends that you are bound to have simply in virtue of being a finite, rational agent but that would necessarily be frustrated in WUM, then M fails the *contradiction-in-the-will test.*

5. If M fails either test, then the proposed action is *morally impermissible*; otherwise the proposed action is *morally permissible.*[14]

An example will help us get a feel both for the CI-procedure and for how the generality problem arises for it. Earlier I mentioned the case of the axe-murderer at the door for familiarity's sake, but we can also apply the CI-procedure to less well-worn examples.

Suppose I would like to own a copy of the latest release by my favorite band. However, being stingy by nature, I am unwilling to buy the album through legitimate means. My friend tells me of an

---

reference to a circumstance). Ross (1930), p. 7, makes a similar distinction between "actions" and "acts." I bypass here a difficulty about whether an action should be considered a distinct entity from its act or merely the same entity under a different description.

[12] Could there be, for a single action, *multiple* maxims in accordance with which one acts? Throughout this chapter I will assume—as Kant's frequent talk of *the* maxim of an action strongly suggests—that there is always a single maxim for each action.

[13] I have formulated the contradiction-in-conception test disjunctively to gloss over an internecine dispute among the neo-Kantians about the proper interpretation of that test: Herman (1993c), for example, advocates the logical contradiction interpretation, whereas Nell (1975) and Korsgaard (1985) advocate the practical contradiction interpretation. It will not matter for our purposes which interpretation is correct.

[14] For similar summaries of the steps in the CI-procedure, see Rawls (2000), pp. 167-169; Nell (1975), pp. 59-63; Herman (1984), pp. 46-47; Korsgaard (1997a), pp. xviii-xxi; and Millgram (2003), pp. 526-527.

online computer server from which I can download, for free, a pirated copy of the album. I ask myself, "Would it be morally permissible for me to do so?" Running through the CI-procedure, I first formulate the maxim M of my proposed action: "When I don't feel like spending money on a music album that I want, I will download a pirated copy of that album in order to obtain a copy of it to listen to." Next I imagine the world of the universalized maxim: a world in which (i) everyone, when they wish to obtain a music album but don't feel like buying it, downloads a pirated copy of the album, and in which (ii) I am also in that world, attempting to act on M. Let us suppose that, in such a world, downloading of pirated music is so rife that the entire music industry collapses.[15] Then either we have a logical contradiction, since it is inconceivable that there be an album for me to attempt to download when the music industry has collapsed, or we have a practical contradiction, since my intent to download the album would of necessity be frustrated if there is no album to download; thus on either interpretation, our maxim M fails the contraction-in-conception test. So it appears that, according to Kant's Formula of Universal Law, it is impermissible for me to download a pirated copy of the album.[16]

Not so fast, though—who says that the maxim of my action is M? For example, my maxim might really be the more specific maxim M′: "When I don't feel like spending money on music album X, I will download a pirated copy of X in order to obtain a copy of it to listen to." Presumably no contradiction, logical or practical, exists in the world of this universalized maxim: even if we assume that most people in that world want to hear the album in question and would rather not purchase it through legitimate means, the widespread downloading of a single album would hardly overthrow the entire music industry. So while perhaps it might take me a bit longer to download that specific album since so many others are downloading it as well, it seems safe to assume that no logical or practical contradiction arises in a world in which M′ is a universal law of nature. And presumably a similar results applies if the maxim of my action is instead the slightly less specific maxim M″: "When I don't feel like spending money on a music

---

[15] Note that in making this assumption, we face another vexed issue in Kant interpretation, namely to what degree it is acceptable (and perhaps necessary) to make empirical assumptions when determining whether there is a contradiction of one form or another in the world of the universalized maxim.

[16] Almost every step of the preceding application of the CI-procedure requires additional defense and/or clarification. However, since I am more concerned with *what we input into the CI-procedure* than with *how the CI-procedure works once we have an input*, I will for the most part be ignoring such interpretative details.

album of musical genre G that I want, I will download a pirated copy of that album in order to obtain a copy of it to listen to." Moreover, let us assume (as seems plausible) that there are no ends that I am bound *qua* finite, rational agent to have and that would be frustrated by everyone's downloading that specific album, or albums of that specific genre, so that neither M′ nor M″ fails the contradiction-in-the-will test, either. Then depending on which maxim (M, M′, M″, or some other) is the true maxim of my action, we get different results as to whether it is morally permissible for me to download the album in question.[17]

Earlier I appealed to a distinction between process tokens and process types in formulating the generality problem for reliabilism, and we can do the same when formulating the analogous problem for Kant's Formula of Universal Law. When deliberating about whether to download the album, what I am pondering is whether, in a given *token* circumstance, to perform a given *token* act for a given *token* end. Maxims, on the other hand, must advert to circumstance, act, and end *types*, or else it could not be the case that more than one person acts on the same maxim in the world of the universalized maxim: really a maxim is of the form, "When in circumstance type C, I will do act type A in order to achieve end type E." But now the generality problem is thrust upon us: a specific circumstance, act, or end token is an instance of ever-so-many circumstance, act, or end types; which of these types are the relevant ones to put in the maxim whose universalizability we are to be assessing when running through the CI-procedure?

4. The Need for a Deeper Characterization of Each Problem

The similarities between the two problems facing reliabilists and Kantians are striking, but almost as striking are the differences in their respective attitudes toward their respective versions of the problem. Reliabilists are terrified of the generality problem; probably the only solution to it in the literature generally thought to have much chance of succeeding is a proposal of William Alston's, but as we shall see, Alston's proposal faces grave difficulties. On the other hand, Kantians—particularly those of the neo-Kantian persuasion—tend to think of their variant of the generality problem as already having been solved. During the early and middle parts of the 20th century, it was widely held that there are

[17] Onora O'Neill has called this the "problem of relevant descriptions" (1975, p. 12), but I will continue to call it the generality problem to emphasize its similarity to the analogous problem for reliabilism.

insurmountable problems with taking Kant's Formula of Universal Law to be a workable principle that yields plausible moral verdicts.[18] However, thanks largely to the work of neo-Kantians such as O'Neill, Herman, and Korsgaard, many contemporary Kantians take these problems for the Formula of Universal Law to have been addressed, including—it is usually assumed—the generality problem. Given how much effort, in particular, was spent in the early work of O'Neill and Herman attempting to answer the generality problem, mention of the problem tends to elicit an exasperated tone from neo-Kantians. "Not this again," they seem to say. "Wasn't this problem solved years ago?"

But despite the neo-Kantians' confidence that a solution to their version of the generality problem has already been found, in my experience the vast majority of non-Kantians are convinced that the Kantian generality problem has not been solved—indeed, are convinced that the Kantian generality problem *could not* be solved, at least not on the Kantian's terms. So even if the neo-Kantians are correct that a solution to their problem is on offer, they still need to diagnose why so many intelligent individuals resist accepting that solution.

In fact, I think that both sides to both debates are in need of a deeper characterization of the generality problem if they are to fully make their case. *Critics of reliabilism* need a deeper characterization because, so far, their case has largely consisted in finding problems with every existing solution to the generality problem. As such, all they have shown is that a solution *has not yet been found*, not that a solution *could not be found*, and one is left with the lingering feeling that maybe reliabilists have simply not been clever enough in their attempts at solving the generality problem. To forestall such suspicions, the critics of reliabilism need to do a better job than they have explaining *why*, exactly, the problem arises. *Defenders of reliabilism* need a deeper characterization of the generality problem because they seem fresh out of ideas for how to answer it, and no doubt the best route to a solution to a philosophical problem is a deeper understanding of what gives rise to the problem in the first place. *Critics of Kantianism* need a deeper characterization of the Kantian version of the generality problem if they are to convince the neo-Kantians

[18] Besides the generality problem, the problems for the Formula of Universal Law that I have in mind here are the charge that it is an empty formalism and the charge that it is overly rigoristic. For discussion of these latter two charges, see Korsgaard (1985), O'Neill (1984), and O'Neill (1985).

that a solution to that problem has not, in fact, been found—and more importantly, if they are to convince the neo-Kantians that a solution to that problem *cannot* be found. And *defenders of Kantianism* need a deeper characterization of the problem to help persuade their critics that they have indeed solved the generality problem, and to help diagnose why acceptance of their favored solution has faced such resistance.

The main purpose of this chapter will be to attempt such a deeper characterization—of what the generality problem for each theory is, of why it arises, and of whether there is any hope for a solution to it. My guiding idea is that by looking at these two variants of the generality problem in parallel, we can gain better insight into each. So in particular, the chapter's next few sections (§§5-6) will be devoted to reformulating the reliabilist and Kantian versions of the generality problem so as to emphasize the similarities between the two and iron out irrelevant differences of detail. I will do this by a somewhat circuitous route: first (§5) I will consider Alston's proposed solution to the generality problem for reliabilism, both because I do not think its faults have been adequately addressed in the literature, and because Alston's way of casting the problem will provide a clue for how to bring our formulations of the problem for reliabilism and Kantianism closer to one another, which I will attempt to do in the section that follows (§6). Then I will use this way of characterizing what is common to the problem besetting both theories to argue that the three main strategies utilized by the neo-Kantians to resist the generality problem do not succeed (§§7-10). Finally, lessons learned from this discussion will allow me to close (§11) by positing a general theory as to why reliabilism and Kantianism both face a generality problem, and to speculate about what hope they each have of ever solving it.

5. Alston's Attempted Solution to the Generality Problem for Reliabilism

As mentioned earlier, William Alston's proposed solution to reliabilism's generality problem in his article "How to Think about Reliability" is widely regarded as the most successful attempt at solving that problem. Alston's proposal has two basic ingredients. First, following a suggestion of Goldman's, Alston recommends that we think of each belief-forming process type as a *function* (in the mathematical sense)

mapping certain token states—the "inputs"—to certain other token states—the "outputs." So in particular, each output of such a function consists in the token state of one's believing some proposition. Second, Alston insists that it follows from an assumption he calls "psychological realism" that whenever one comes to hold a given belief token, there is a unique "psychologically real" function that is activated in the formation of that belief token. This gives us a nice, tidy solution to the generality problem: the psychologically real function determines the relevant process type, and we can evaluate whether that type is reliable by seeing whether, in a suitable range of (actual and counterfactual) cases, the ratio of cases in which the function's output is a true belief token to those in which its output is a false belief token is sufficiently close to 1.

Several simplifying assumptions will make it easier to both understand and assess Alston's proposal. First, let us ignore cases in which the input states to a given function include belief states—such functions correspond to what Goldman calls "belief-dependent processes" and give rise to additional complications that need not concern us. Second, let us follow Alston in assuming a particular solution to the problem of extent. In particular, let us follow Alston in assuming that when one forms a given (token) belief via perception (such as: a belief in the proposition *that there is a tree to one's right*), the relevant input state is one's having a certain (token) experience (such as: an experience as of a tree being to one's right).[19]

Finally, let us make one last assumption so that we can overlook a slight mistake in Alston's proposal. Alston's use of the term "function" in the mathematical sense ignores a crucial element in the desired process type: once a given domain and range are fixed, mathematical functions are individuated solely by <input, output> ordered pairs, so if some function F maps experiential input x to belief output b,

[19] Throughout this chapter, I will talk about *experiences as of certain things being the case*, but in doing so, I do not intend to commit myself to any particular theory concerning the nature of perceptual experience, or to any particular view about the admissible contents of such experience. For ease of exposition, I will usually assume that fairly robust contents such as *that there is a tree to my right* can be part of the content of experience, but if one disagrees with this assumption, all of my examples could (with a bit of work) be reformulated in terms of much thinner contents. Similarly, I do not intend to commit myself to any particular view about which perceptual processes are belief-dependent and which belief-independent. For ease of exposition, I will usually assume that processes like one's coming to believe via perception that there is a tree to one's right are belief-independent, but if one disagrees with this assumption, all of my examples could (with a bit of work) be reformulated in terms of whatever the true belief-independent processes end up being. (Note that process reliabilists are committed to there being at least *some* belief-independent processes.)

It is also worth noting that Alston's solution to the problem of extent requires making some potentially contentious assumptions about the connection between experience and belief. In particular, his solution commits him to the claim that perceptual experiences *cause* perceptual beliefs about the content of those experiences, but one might doubt that perceptual beliefs and experiences are, in this way, entirely distinct existences. For more on this subject, see §6 of chapter 3 of this dissertation.

then *any* process token that begins with the event token of one's coming to have x and ends (by whatever deviant means) with the event token of one's coming to have b qualifies as an instance of the type associated with function F. However, we might think that only *certain ways* of getting from the having of x to the having of b should count as instances of the relevant process type. In effect, then, Alston's proposal ignores anything that might occur in the *middle* of a given process. So let us follow Alston in making the assumption that what happens at the middle stages of a process type is irrelevant to its reliability; conceding this to Alston will not matter much, for (as we shall soon see) Alston has a difficult enough time pinning down what happens at the beginning and end of a given process type.[20]

With these assumptions in place, Alston's proposal can be formulated in the abstract as follows. Suppose one comes to believe a certain proposition on the basis of a certain perceptual experience that one has. This gives us an ordered pair of state tokens $<x, b>$ = *<a certain token experience, a certain token belief>* which (since we are ignoring what happens in the middle of a given process type or token) we can identify with the process token by which b was formed. Alston's suggestion is that the assumption of psychological realism gives us a unique function that is activated in the formation of b. We can identify this function with a set F of <experience token, belief token> ordered pairs such that (i) no two ordered pairs share the same first member (i.e. the function maps every input to a unique output), and (ii) $<x, b> \in$ F (i.e. the function maps x to b). Let X be the function's *domain*: the set of experience tokens such that one of the ordered pairs in F has that experience token as its first member. And let B be the function's *range*: the set of belief tokens such that one of the ordered pairs in F has that belief token as its second member. Then we can assess the reliability of F by seeing whether, in a suitable range of circumstances in which one is in token states x^* and b^* such that $x^* \in$ X, $b^* \in$ B, and $<x^*, b^*> \in$ F, the ratio of cases in which b^* is a belief token with a true content to those in which b^* is belief token with a false content is suitably close to 1. If this is so, then the original belief token b counts as justified; otherwise, it counts as unjustified.

---

[20] How to amend Alston's proposal once one discharges this assumption is not entirely clear. The natural suggestion is to consider a string of functions composed with one another, but this proposal raises a host of worries. For example, how do we determine how many—and which—intermediate stages to take into account when formulating the string of functions? But more importantly, adding additional intermediate stages does not address the main worry, since for each intermediate function we will still need some way of ruling out deviant ways of getting from a given input to a given output of that intermediate function.

However, to phrase Alston's proposal in this way is to immediately see what is wrong with it. If we are somehow given a domain X of relevant token experiences and a range B of relevant token beliefs, we can all agree with Alston that his thesis of psychological realism fixes the map F from elements of X to elements of B: given that one's having experience token x causes one to have belief token b, it might well be true that one's psychological mechanisms fix the fact that if one were to have somewhat similar experience token x^*, one would come to have somewhat similar belief token $b^* = F(x^*)$; that if one were to have somewhat similar experience token x^{**}, one would come to have somewhat similar belief token $b^{**} = F(x^{**})$; and so on. The problem, however, is: *nothing in Alston's proposal gives us any guidelines for how to fix the relevant domain X and range B.* What is the extent of the somewhat similar <experience token, belief token> ordered pairs ($<x^*, b^*>$, $<x^{**}, b^{**}>$, etc.) that have a bearing on the reliability of the process by which b was formed? By itself, Alston's proposal gives us no way to answer this question.

For example, suppose I am playing with a collection of colored blocks. I have an experience as of a red block being to my right, and as a result I form a belief that a red block is to my right. Thus we have an ordered pair of state tokens $<x, b>$ = <*my token experience as of a red block being to my right, my token belief that a red block is to my right*> that we can identify with the token process by which b is formed. What is the relevant function F whose reliability we are to assess when judging whether b is justified? Suppose that, given the perceptual and cognitive mechanisms I happen to have, the following facts all hold true:

- if instead I were to have token experience x_1 = *a token experience as of a red block being slightly more to my right*, I would have had a different token belief with the same content, namely b_1 = *a token belief that a red block is to my right*;
- if instead I were to have token experience x_2 = *a token experience as of a reddish-orange block being to my right*, I would have had a different token belief with the same content, namely b_2 = *a token belief that a red block is to my right*;
- if instead I were to have token experience x_3 = *a token experience as of a red block being to my left*, I would have had a different token belief with a different content, namely b_3 = *a token belief that a red block is to my left*; and
- if instead I were to have token experience x_4 = *a token experience as of a green block being to my right*, I would have had a different token belief with a different content, namely b_4 = *a token belief that a green block is to my right.*

Given these facts, it seems plausible that *if* x_1, x_2, x_3, and x_4 are all members of F's domain X, and *if* b_1, b_2,

b_3, and b_4 are all members of F's range B, *then* $\langle x_1, b_1\rangle$, $\langle x_2, b_2\rangle$, $\langle x_3, b_3\rangle$, and $\langle x_4, b_4\rangle$ are all members of F. However, how do we determine whether x_1, x_2, x_3, and x_4 are all members of X? And how do we determine whether b_1, b_2, b_3, and b_4 are all members of B? When evaluating whether my perceptual belief in the proposition *that a red block is to my right* is the outcome of a reliable process, do we count instances in which I seem to see a block in a different position, or seem to see a block of a different hue, as instances of the same process, and if so, how great a difference in position or hue is allowable? It should be clear that to answer these questions *just is* to answer the generality problem for reliabilism, and thus an appeal to "psychologically real" functions cannot be the complete solution to that problem.

One reply on Alston's behalf would be to claim that, at least with regards to the selection of range B, the choice is obvious: we pick the set of all token beliefs that share the same content with our original belief token *b*. An initial problem with this reply is that, as Conee and Feldman point out, if we use this as a general strategy for how to select the relevant function for every belief-forming process, we are saddled with the absurd result that any process resulting in a belief in a necessary truth or a self-confirming belief (such as the belief *that someone has beliefs*) is perfectly reliable, since on the proposed account it is impossible for the function associated with that process to yield an output state whose content is false.[21] A bigger worry for this reply, though, is that even if it helps us fix the range of our function, we are still left with the issue of how to select the proper domain. For example, in the case just mentioned, do we include experience tokens x_1 and x_2 in the domain? If we do, then each would be mapped to a belief token with the same content as *b*, but it is far from clear what constraints are in play to rule them in or out.

At this point it is extremely tempting to insist that, in each case, there is a certain "natural" domain that is fixed by the way the cognitive and psychological mechanisms in question work; indeed, there is good evidence that this is what Alston intended to propose from the outset. However, to fill in Alston's proposal in this way is just to slide back into an appeal to natural kinds as a way of attempting to solve the generality problem: the suggestion, in effect, is that there are certain joints in nature that help us

[21] Conee & Feldman (1998), p. 14. Comesaña (2006)'s recent attempt at solving the generality problem for reliabilism also has this absurd consequence.

fix the relevant domain—certain natural ways of carving up sets of experiential input states into the right groups. The problem, though, is that there are *too many* joints in nature for this suggestion to be of any use. In the example we have been considering, should the domain X be the set of all token experiences as of something red being to one's right? The set of all token experiences as of something colored being somewhere in one's field of vision? The set of all token experiences as of something exactly such-and-such shade of red being in exactly such-and-such spot in one's field of vision? Each of these sets seems as natural as the rest.

Thus I conclude that Alston's suggested solution to the generality problem for reliabilism fails. His central mistake is that he has conflated the *input/output* of a function with its *domain/range*. It is true that if one's coming to have token experience x causes one to come to have token belief b, then this fixes *one input* and *one corresponding output* of the function that Alston seeks. However, this fact by itself gives us no way of determining *the entire domain of inputs* that the function takes and *the entire range of outputs* that it can yield, which is what we would need to solve the generality problem.[22]

6. Structural Similarities Between the Two Generality Problems

As a solution to the generality problem for reliabilism, Alston's proposal is unsuccessful. However, Alston's way of formulating his would-be solution brings out an important parallel between the generality problem for reliabilism and the analogous problem for Kantian ethics.

In the case of perceptual beliefs formed via belief-independent processes, Alston's proposal is as follows. Belief token b formed on the basis of experience token x is epistemically justified if and only if the function F which is activated in the formation of that belief token is reliable. Let X be F's domain and B be F's range, where $x \in X$ and $b \in B$. Then the problem that Alston's account leaves unsolved is how to get from a given input/output token $<x, b>$ to the relevant ordered pair $<X, B>$. The analogous generality problem for Kantianism can be formulated as follows. According to the Formula of Universal

[22] Adler & Levin (2002)'s recent defense of Alston's proposal against Conee & Feldman (1998)'s criticisms faces a similar problem: Adler & Levin argue that Alston's psychologically realized functions relate *variables*, not *values*, but then provide no way of determining the relevant *range* of these variables.

Law, the (token) action of performing act token *a* for the sake of end token *e* in circumstance token *c* is morally permissible if and only if that action's maxim is universalizable. Moreover, the action's maxim can be represented as an ordered triple <act type A, end type E, circumstance type C>, where token *a* is an instance of type A, token *e* is an instance of type E, and token *c* is an instance of type C. The problem left unsolved is how to get from a given action token <*a*, *e*, *c*> to the relevant maxim <A, E, C>.

The similarities here should be noticeable. However, with several additional assumptions and simplifications, we can bring these formulations even closer to one another. First, let us assume that types are sets of tokens. This involves taking a stand on the metaphysics of types, but for our purposes the assumption is harmless, and it will greatly simplify our notation. Second, we should notice that in the case of Kantianism, it is <act token, end token, circumstance token> ordered triples that are the object of moral evaluation, whereas in the case of reliabilism, it is only the belief token itself—rather than the <experience token, belief token> ordered pair—that is the object of epistemic evaluation. However, this difference does not seem to be one of great consequence. As Goldman stresses in his seminal defense of process reliabilism, reliabilism is a *historical* (as opposed to a *current time-slice*) *theory of epistemic justification* according to which every belief token is the outcome of some (token) causal process, and it is *in virtue of* certain properties of its causal history—that is, *in virtue of* certain properties of the process token by which it was formed—that the belief token is or is not justified.[23] Thus whether we say that the belief token itself is epistemically justified due to its being formed by a process token with certain properties, or whether we instead say that it is the entire process-token-ending-in-the-formation-of-a-belief-token that is epistemically justified, seems largely a matter of terminological convention. So hereafter, let us take it to be the <experience token, belief token> ordered pair (our proxy for the process token) that is the object of epistemic evaluation according to reliabilism.

Third, let us ignore the circumstance portion of maxims, and treat maxims as <act type, end type> ordered pairs rather than <act type, end type, circumstance type> ordered triples. A full solution to the generality problem for Kantianism would include a way of determining the relevant circumstance type

[23] Goldman (1979), p. 117.

for a given action token. However, determining how to select the relevant act and end types is already a difficult enough problem, and presumably whatever technique is used to select the relevant act and end types can also be extended to select the relevant circumstance type.[24]

Finally, let us make one last rather large simplification, and identify the relevant process type for a given process token with the ordered pair <X, B> consisting of the domain X and range B of the function F activated by that process token. This might seem like an egregious over-simplification of our problem, since we are in effect ignoring the actual function mapping inputs to outputs, but I have already conceded to Alston that if we can find a way of selecting the pertinent domain X and range B in a given case, we can appeal to his thesis of psychological realism to fix the relevant function F.[25]

With all of these qualifications in place, let us now restate each theory's version of the generality problem. Suppose belief token b is formed on the basis of experience token x. Then reliabilism holds the following:

<x, b> is epistemically justified iff <X, B> is reliable,

where $x \in X$, $b \in B$, and <X, B> is the epistemically relevant process type for process token <x, b>. The *generality problem for reliabilism*—modulo the simplifications we have made—is to find a defensible, non-ad-hoc way of selecting the relevant process type <X, B> for every process token <x, b>.[26]

---

[24] In many ways, the problem for Kantians of selecting the relevant circumstance type is analogous to what I called the *problem of the scope of assessment* for reliabilism: the problem of selecting the relevant range of actual and counterfactual situations to consider when assessing the reliability of a given process type. Moreover, this parallel should confirm something that many readers no doubt suspected from the outset: the scope of assessment problem *just is* another aspect of the generality problem for reliabilism, not a separate problem that needs an alternate treatment. However, I will continue to neglect the problem of the scope of assessment (and the corresponding problem of circumstance type selection for Kantianism) in order to keep the discussion manageable.

[25] The need for a way of selecting the relevant function F might suggest at least one point of disanalogy between the generality problem as it arises for reliabilism and the generality problem as it arises for Kantian ethics. However, appearances are deceiving: in fact, a full-blown solution to the Kantian generality problem would include a component that plays the same role as function F, in the form of an instance of the *Hypothetical Imperative* that connects the specified means to the specified ends. For a given maxim <act type A, end type E>, it is not true that any action token <act token a^*, end token e^*> such that $a^* \in A$ and $e^* \in E$ counts as a pertinent action to have someone doing in the world of the universalized maxim: that action token will only be relevant if act a^* *really is* a means to end e^*. (For more on the complicated issue of how the Hypothetical Imperative interacts with the Categorical Imperative within Kantian ethics, see Herman (1990), ch. 3, and Korsgaard (1997b).)

[26] This way of formulating the generality problem for reliabilism reveals an important mistake in Conee and Feldman's favored way of arguing against potential solutions to that problem. Conee and Feldman most often argue that a given proposal falls prey to what Feldman (1985) calls the "No-Distinction Problem": the problem of selecting a process type for a given process token which is such that, intuitively, some of the belief tokens formed as a result of an instance of that type are justified, and some of the belief tokens formed as a result of that type are unjustified (so that the proposal fails to make a distinction where intuitively there should be one). However, it is a mistake to assume that just because <X, B> is the relevant process type for process token <x, b>, therefore every process token <x^*, b^*> such that $x^* \in X$ and $b^* \in B$ has <X, B> as its relevant process type.

Suppose act token *a* is performed for the sake of end token *e*. Then Kant's Formula of Universal Law holds the following:

<*a*, *e*> is morally permissible iff <A, E> is universalizable,

where $a \in A$, $e \in E$, and <A, E> is the morally relevant action type (= the maxim) for action token <*a*, *e*>. The *generality problem for Kantian ethics*—modulo the simplifications we have made—is to find a defensible, non-ad-hoc way of selecting the relevant action type <A, E> for every action token <*a*, *e*>.

7. Desiderata for a Successful Solution to Either Generality Problem

Our reformulations and simplifications have revealed a certain structural similarity between the two versions of the generality problem, but we still have further to go before reaching the desired characterization of the problem and what gives rise to it. However, in the next few sections I want to consider one pay-off from our characterization of the problem so far: it will allow us to fairly easily show why the three main neo-Kantian solutions to the generality problem are all unsatisfactory. Our discussion of the failings of these would-be solutions will in turn motivate my attempt at a deeper characterization of the generality problem in the chapter's final section. But before we launch into an assessment of the neo-Kantians' attempted solutions to the generality problem, it will help to have on the table certain desiderata for a successful solution to either theory's version of the generality problem.

In their classic statement of the generality problem as it arises for reliabilism, Conee and Feldman list three conditions for the adequacy of a proposed solution to the generality problem. First, the solution should be principled and not just an ad hoc, case-by-case selection of the relevant process type for a given process token. Second, the solution should remain true to the naturalistic spirit of reliabilism (so, for example, the account should not smuggle in a non-reliabilist term of epistemic evaluation into the specification of the relevant type). And third, the solution should make intuitively plausible epistemic classifications (i.e. it should, for the most part, deem as justified those beliefs which intuitively count as justified, and deem as unjustified those beliefs which intuitively count as unjustified).[27]

---

[27] Conee & Feldman (1998), pp. 4-5.

A fully adequate solution to the generality problem as it arises for Kant's Formula of Universal Law would meet three parallel conditions. First, the solution should not be an ad hoc, case-by-case selection of the relevant maxim for a given action. Second, the solution should remain true to the spirit of Kantianism; so, for example, the account should not appeal to an *independent* standard of moral salience in weeding out extraneous or inapposite bits of detail in the formulation of a given maxim. After all, the Categorical Imperative is supposed to be *what fixes* which items of a given situation are morally pertinent and which are not; to appeal to an independent criterion of moral appropriateness over and above the Categorical Imperative would give rise to the worry that the Categorical Imperative is not, after all, the most basic determinant of moral justification.[28] And third, the solution should make defensible moral classifications, though crucially the Kantian is not as beholden to our intuitions about cases as the reliabilist is. The reliabilist's primary (if not exclusive) method of arguing for her theory is to appeal to our intuitive judgments about whether a given scenario, real or imagined, counts as a genuine case of justified belief, and then to insist that reliabilism does the best job of fitting that data.[29] Thus if reliabilism ends up yielding the verdict that, say, a large number of cases that we intuitively deem to be cases of justified belief are in fact cases of unjustified belief, this raises serious doubts as to whether reliabilism really does do the best job of capturing the intuitive data. Kantianism, on the other hand, has the potential to be much more revisionary about our case-specific intuitions. Although Kant's argument in the *Groundwork of the Metaphysics of Morals* opens with, in effect, an appeal to our intuitions about whether various things are or are not good under certain conditions, by the time he has argued for both the form and existence of the Categorical Imperative by the *Groundwork*'s end, Kant takes himself to have discharged his appeal to those intuitions. For this reason it is perfectly open to the Kantian who faces a recalcitrant moral intuition to stand her ground and simply insist that the case at hand is one in which our intuitions go wrong. So while the results of the CI-procedure should not be wildly implausible, they also need not comport with our

---

[28] For this reason, I do not consider Herman's suggestion in her (1985) that Kantians need to find a place within their moral theory for so-called "rules of moral salience" that "determine which facts it is legitimate to include in a maxim" to constitute an adequate solution to the generality problem for Kantian ethics (p. 75).

[29] For example, this is how Alvin Goldman argues for reliabilism in his reliabilist manifesto, Goldman (1979).

moral intuitions in all cases.[30]

With these desiderata in mind, I now want to consider three different general strategies that have been pursued by neo-Kantians as a way of resisting the generality problem.

8. First Neo-Kantian Strategy: Privileging a Certain Level of Generality

Perhaps the most tempting strategy in responding to the Kantian generality problem is to pick a certain level of generality in action type, and then to insist (for whatever reason) that the maxim of an agent's token action is always at that privileged level of generality. For example, in her early attempt to solve the generality problem for Kantianism (or the "problem of relevant descriptions," as she called it then), Onora O'Neill proposed identifying the relevant maxim to test for universalizability with the agent's specific intention when acting: on this suggestion, the morally pertinent act, end, and circumstance type are the descriptions of the specific act, end, and circumstance tokens under which the agent intends the action.[31] In her later work, O'Neill has rejected this account in favor of privileging a different level of generality: O'Neill now holds that maxims consist in general "underlying or fundamental principles" by which we control our more specific intentions.[32] Similarly, Rüdiger Bittner and Otfried Höffe have argued that maxims should be understood as higher-order *Lebensregeln* ("rules of life") that structure the lower-level particularities of how we live our daily lives.[33] And Barbara Herman has insisted that the CI-procedure should only be applied to what she calls "generic maxims," which are the "most general form of a given kind of maxim."[34]

---

[30] While the Kantians' positive arguments for the Categorical Imperative make this third condition on an adequate solution to the generality problem less demanding, they also give rise to a fourth condition: the solution to the generality problem must not undermine the force of those positive arguments. One cannot take any old theory of maxims and append it to a statement of Kantian ethics; the theory of maxims must be compatible with the Kantian arguments for the form of the Categorical Imperative given in Section II of the *Groundwork*, and for the existence of a Categorical Imperative given in Section III of the *Groundwork* (but later revised in the *Critique of Practical Reason*). However, the exact structure of these arguments is a complicated matter that would take several additional chapters to address properly, so I will not have much more to say, at least in the current work, about this fourth condition for an adequate solution to the generality problem for Kantianism.

[31] Nell (1975), pp. 40-42.

[32] O'Neill (1984), pp. 151-152; (1985), pp. 84-85; and (1989b), pp. 129, 141.

[33] Bittner (1974) and Höffe (1977).

[34] Herman (1989), p. 117; see also her (1993c), p. 147. Herman deviates from the other neo-Kantians in that she does not think that a generic maxim's failing the contradiction-in-conception or contradiction-in-the-will test makes action tokens falling under that generic maxim morally prohibited. Rather, she thinks failing these tests only establishes a *deliberative presumption* against performing action tokens falling under the generic maxim—a presumption that may be *rebutted* in certain cases in which an omission of the action token falls under a more stringent generic maxim that also fails the contradiction-in-conception or

One problem for each of these interpretations is finding textual evidence for the proposed restriction to the possible content of maxims. Kant himself provides examples of maxims at various levels of generality. Some of the maxims he mentions are amusingly specific, such as his proposed maxim for a pleasant dinner party (from *Anthropology from a Pragmatic Point of View*): "not to allow deadly silences to set in, but only momentary pauses in the conversation."[35] Other maxims mentioned by Kant are at a much higher level of abstraction, including the Categorical Imperative itself, which in *Religion within the Limits of Reason Alone* is claimed to be the "supreme" maxim of one's action.[36] So endorsing any of these proposed restrictions of maxim content to a fixed level of generality involves doing a certain amount of damage to at least some of Kant's texts.

Another problem with these proposals is that, despite their intent, they almost always fail to pin down a unique level of generality for maxims to have. For example, intentions—especially intentions construed as particular descriptions that an agent takes an action to fall under—can occur at any number of levels of generality. Similarly, while requiring that maxims tested by the CI-procedure be "underlying principles," "rules for life," or "generic maxims" excludes extremely specific maxims such as "I will download a pirated copy of music album M from computer server S in order to obtain a copy of it to listen to," there are still any number of more fundamental maxims that could fairly be said to be the maxim of my action when I download a particular album on a particular occasion, from the moderately general ("I will download a pirated copy of a music album in order to obtain a copy of it to listen to") to the ultra unspecific ("I will do what it takes to get what I want"). So on all of these proposals, we are still left wondering at which level of generality we should look for the relevant maxim to test for universalizability when assessing the permissibility of a particular action.

However, the major problem with these proposals is not their inability to select a unique level of

---

contradiction-in-the-will test; see Herman (1989), pp. 116-117, and (1993c), pp. 147-157. However, it is far from clear how Herman has the resources to explain why one deliberative presumption trumps another in a given context without greatly deviating from the spirit of Kantianism, so I will not discuss this portion of her view in what follows. (For more on the sorts of substantive assumptions about the structure of morality that underlie talk of one moral consideration "trumping" or "outweighing" another, see chapter 1 of this dissertation.)

[35] Kant (1798/2006), AK 7:281-282; cited in Kitcher (2003), p. 221. Other dinner party maxims recommended by Kant: "to choose topics for conversation that interest everyone and always provide someone with the opportunity to add something appropriate," and "not to change the topic unnecessarily or jump from one subject to another."

[36] Kant (1793/1960), AK 6:48; also cited in Kitcher (2003), p. 221.

generality; rather, the major problem is their assumption that selecting a unique level of generality would constitute a solution to the generality problem. It is not uncommon for people to attempt to summarize the generality problem for Kantianism by asking the question, "How broad or narrow do we make the maxim of one's action?" This way of phrasing the problem makes it seem as if there is a continuum of successively broader maxims that we must somehow choose between. However, things are much more complicated than that: for each action token, we do not have a *linear ordering* of more and more general action types that the action token falls under. Rather, we have a *partial ordering* of action types, some of which are *fully contained* by another candidate action type (in the sense that every instance of the one is also an instance of the other), but others of which *merely intersect* each other. So restricting ourselves to a certain level of generality in action type does not solve the generality problem: for a designated level of generality, there will always be *multiple* types at that level which the action token falls under, and moreover, there is no reason to think that all of these types yield the same result when run through the CI-procedure.

This point is most easily illustrated if we momentarily switch back to discussing the generality problem for reliabilism. Recall my example in which, while playing with some colored blocks, I form *b = a token belief that a red block is to my right* on the basis of *x = a token experience as of a red block being to my right.* Consider the following candidates for process token <*x*, *b*>'s epistemically relevant process type:

$<X_1, B_2>$ = < {experiences with content *that a red block is to my right*},
{beliefs with content *that a C block is to my right*, for some color C}>

$<X_2, B_1>$ = < {experiences with content *that a C block is to my right*, for some color C},
{beliefs with content *that a red block is to my right* }>

$<X_2, B_2>$ = < {experiences with content *that a C block is to my right*, for some color C},
{beliefs with content *that a C block is to my right*, for some color C}>

$<X_3, B_3>$ = < {experiences with content *that a red block is to direction D of me*, for some direction D},
{beliefs with content *that a red block is to direction D of me*, for some direction D}>

Since domain X_1 and range B_1 are at the same level of generality, and domain X_2 and B_2 are also at the same level of generality, process types $<X_1, B_2>$ and $<X_2, B_1>$ are presumably both at the same level of generality, yet neither entirely contains the other. Moreover, we can also generate such intersecting process types even when each type's domain and range are at the same level of generality, as in the case of

process types $<X_2, B_2>$ and $<X_3, B_3>$. So it is simply a mistake to assume that the various process types of which $<x, b>$ is an instance are all *nested* within one another in a linear fashion; instead, what we have is a series of *crisscrossing* process types of various different levels of generality.

A similar result holds when we are trying, for a given action token, to select the appropriate action type on which to run the CI-procedure. Consider a case in which I perform *a = my (token) downloading of a pirated copy of music album M from computer server S* for the sake of *e = my (token) obtaining of music album M so as to listen to it*. The following are both plausible candidates for action token $<a, e>$'s morally relevant action type:

$<A_1, E_1> = <$ {downloadings of a pirated copy of music album M from some computer server}, {obtainings of something to listen to}>

$<A_2, E_1> = <$ {downloadings of a pirated copy of some music album from computer server S}, {obtainings of something to listen to}>

As act types A_1 and A_2 are presumably at the same level of generality, action types $<A_1, E_1>$ and $<A_2, E_1>$ must also be at the same level of generality. So we do not have a series of *nested* action types to choose between; instead we have a series of *crisscrossing* action types at various different levels of generality. For this reason, talk of the level of generality of a maxim is a red herring. If we are to solve the generality problem for Kantian ethics, we need a more precise means of maxim individuation.

9. Second Neo-Kantian Strategy: Counterfactual Tests

Another extremely tempting strategy for fixing on the right action type for a given action token is to appeal to various *counterfactual tests* in order to sift out morally irrelevant features of the specific act and end tokens. Here is Herman's way of formulating such a counterfactual test:

> Extraneous detail is pared from the description in a maxim by asking counterfactually, 'Were this feature of the description not part of the circumstances of the action, would you still act as you proposed?' If the answer is yes, the descriptive element is rejected as extraneous to the agent's conception of his action.[37]

O'Neill calls such an appeal to counterfactuals an "isolation test,"[38] but "removal test" would be a more appropriate label: one is evaluating what would happen were one to *remove* some feature from an action,

---

[37] Herman (1990), p. 257.
[38] O'Neill (1984), p. 152, and (1985), p. 85.

not what would happen were one to *isolate* that feature by itself.[39] We can formulate one version of such a removal test as follows:

> *the removal test for the act component of a maxim*:
>
> If the following is true:
>
> if act token *a* didn't have feature F, the agent would still perform <*a*, *e*>,
>
> then the maxim <A, E> for action token <*a*, *e*> can't be such that for every $a^* \in A$, a^* possesses F.

Clearly we could also formulate an analogous removal test for the end component of a maxim. The suggestion I now want to consider is whether an appeal to removal tests of this sort would be enough to solve the generality problem for Kantian ethics.[40]

The answer, of course, is "No." As is often the case with counterfactual tests, the removal test yields both false negatives and false positives—yields both cases in which the test *rejects* some feature as being part of the act component of an agent's maxim when intuitively it *should not*, and cases in which the test *does not reject* some feature as being part of the act component of an agent's maxim when intuitively it *should*. Here is an example of a false negative:

> *deadly medicine*: Medicine X is a pharmaceutical wonder that completely cures an eye infection within hours when administered directly into the eye. However, Medicine X also causes severe kidney failure in a small percentage of the population, all of whom share a rare gene. Now suppose I am an eye doctor, and a patient who I know possesses that rare gene comes into my office with a minor eye infection. I administer three drops of Medicine X into his eye, thereby curing his eye infection but also irrevocably damaging his kidney. However, if the patient hadn't possessed that rare gene, I still would have administered three drops of Medicine X into his eye.

Applying the removal test, we get the result that the following cannot be the maxim of my (token) action of administering three drops of Medicine X into the eye of my patient in order to cure his eye infection:

> <A, E> = <{administerings of a medicine to someone in whom it will cause kidney failure}, {curings of that person's eye infection}>.

After all, I would have given Medicine X to my patient even if he hadn't possessed the rare gene that

---

[39] This distinction between *isolation* and *removal tests* parallels the distinction between *isolation* and *removal conceptions of reasons for action* discussed in §8 of chapter 1 of this dissertation.

[40] The earliest appeal to removal tests as at least a partial solution to the generality problem occurs in Nell (O'Neill) (1975), pp. 41, 72. Herman also endorses using such tests to help solve the generality problem in her (1990), pp. 64, 67-70, 117-119, 257. However, in their later work both O'Neill and Herman raise worries for the thought that such counterfactual tests can serve as the basis for a solution to the generality problem: see the O'Neill references cited in n. 38, and Herman (1993b), p. 219.

makes him susceptible to Medicine-X-induced kidney failure. But, intuitively, the fact that I administered a medicine to someone in whom it causes kidney failure is precisely the sort of morally relevant consideration we want to build into the act component of my maxim when assessing its universalizability.

Here is an example of a false positive for the removal test:

> *striking resemblance*: I am in a candy store looking at their display of fudge. To my surprise, I notice that one of the pieces of fudge bears an uncanny resemblance to my late uncle. So I steal that piece of fudge in order to show it to my father (the brother of my uncle). However, if the piece of fudge hadn't looked like my late uncle, I wouldn't have stolen it.

In this case I perform token act a = *my token stealing of a piece of fudge that looks like my late uncle* for the sake of token end e = *my token showing of that piece of fudge to my father.* Intuitively, the fact that the piece of candy resembles my late uncle is morally irrelevant, but the removal test does not bar that fact from being part of the content of my action's maxim. And more worryingly, if it *is* part of my action's maxim, then that maxim is almost certain to be universalizable (since cases in which a piece of fudge bear a striking resemblance to one's uncle are so rare), despite the fact that my pilfering of the piece of candy would appear to be a canonical example of an impermissible action. But surely doing what would otherwise be a morally prohibited action for overly specific reasons does not absolve one from wrongdoing!

Thus the proposed counterfactual test is far too blunt a tool to use as a way of solving the Kantian's generality problem. Once again we are left looking for a more precise means of maxim selection.

10. Third Neo-Kantian Strategy: The First-Person Perspective

The third neo-Kantian strategy involves denying the very presupposition that what we are after is a means of maxim *selection*. On this third view, we shouldn't think of the CI-procedure as a way of assessing, after the fact, whether a particular action performed by a particular agent *in the past* was right or wrong; rather, we should think of the CI-procedure as essentially a way, from the first-person perspective, of determining what one should or should not do *in the future*. Moreover, the claim is that once we construe the CI-procedure in this way, the generality problem dissolves, since then there is no issue of selecting the proper maxim for a given action token that has already been performed.

It will help to have a certain distinction on the table before assessing this proposal. We can distinguish between two roles that a moral principle like the Categorical Imperative might play.[41] The first is as a *third-personal standard*: a criterion for the applicability of a moral term such as "right" or "wrong," usually in the form of necessary and sufficient conditions. It is this sort of principle that one uses when making theoretical evaluations of an action from a third-person perspective (including theoretical evaluations of one's own past actions). The second role a moral principle might play is as a *first-personal guide*: a tool for choosing the right or wrong action when, from the first-person perspective, one is deliberating about what to do. A principle that is well suited for one of these roles might be ill suited for the other; most famously, the act-utilitarian principle "An action is right iff it produces at least as much net utility as any other action available at that time" is usually put forward by act-utilitarians as a third-personal standard but not as a first-personal guide. The neo-Kantians' apparent claim is that the reverse holds for the Categorical Imperative (and the CI-procedure by which one implements it): it is meant to be a first-personal guide, not a third-personal standard. As Korsgaard puts it, Kantian ethics always proceeds from "the standpoint of practical reason" (the standpoint from which one reasons about what to do); in O'Neill's terminology, Kantianism is concerned with "the context of action or decision," not "the context of assessment."[42]

Suppose we grant that the Categorical Imperative is essentially a first-personal guide, rather than a third-personal standard; how does this help with the generality problem? Here is O'Neill on the matter:

> [I]n *practical* judgment we are not judging a particular act. The task in practical judgment is to shape action *that is not yet done*. . . . Practical judgment, including ethical judgment, does not encounter the problem of relevant descriptions [i.e. the generality problem] because it is not directed at individual act[ion]-tokens. Whereas [theoretical] judgment aim[s] to *fit the world* or *some possible world*, practical judgment aims in some measure to *shape the world* . . . . The different direction of fit shields practical judgment from the problem of relevant descriptions.[43]

O'Neill's idea is this: when one is deliberating about what to do in a given situation, there is no individual action token <*a*, *e*> about which to ask what its morally relevant action type <A, E> is, *since one has not yet*

---

[41] This sort of distinction is commonplace in contemporary discussions of moral principles; for a slightly older appeal to such a distinction, see Bales (1971), and for a more recent appeal, see McKeever & Ridge (2005), pp. 84-87. Interestingly, reliabilists sometimes draw a parallel distinction between types of *epistemic* principles, usually in order to stress the fact that, for them, epistemic principles are purely third-personal standards, rather than first-personal guides; see, for instance, Goldman (1980).

[42] Korsgaard (1996), p. xi; Nell (O'Neill) (1975), p. 127.

[43] O'Neill (2004), pp. 312-313.

performed any action. Instead, what one directly considers is a given action type <A, E>.[44] But what does it mean to say that one directly considers an action type <A, E> from the first-person perspective? Here is Korsgaard on the matter:

> Roughly speaking, what happens when an agent chooses an action is something like this: the agent is attracted on some occasion to promoting some end or other. The end may be suggested by the occasion, or it may be one he standardly promotes when he can. He reasons about how he might achieve this end, or what he might do in its service, and he arrives at a possible maxim . . . . That is to say, he considers an action [type], and he asks himself whether it is a thing worth doing. And he determines the action [type] to be . . . morally worthy or at least permissible. Kant thinks he makes this determination by subjecting the maxim to a test, the categorical imperative test . . . . Determining the action [type] to be . . . a thing worth doing for its own sake, he does the action [i.e. he performs an action token which has that type as its maxim].[45]

Korsgaard's idea is this: in a given situation, an agent is tempted to promote some end type, E.[46] So the agent faces a practical problem; from the first-person perspective, his question is, "How can I promote E in a morally permissible way?" The agent reasons about possible ways to achieve that end, and proposes various maxims of the form <A, E>, <A*, E>, <A**, E>, etc. He then determines whether any of these maxims are universalizable. If one is, then the Categorical-Imperative-as-first-personal-guide deems that it is morally permissible for him to act on that maxim—or more precisely, deems that it is morally permissible for him to perform an action token which has that action type as its maxim.

On this way of understanding the role of the Categorical Imperative in practical deliberation, the generality problem never arises. Rather than facing the problem of how—as theorists occupying the third-person point-of-view—we are to "move up" from an action token to its relevant action type, instead we start—as agents occupying the first-person point-of-view—with a given action type, and only *after* it has been found to be universalizable do we "move down" to an instance of that type by then acting on it.

But our problem is not yet solved, for the proposal as formulated has blatantly unacceptable moral consequences. Suppose I am in need of $50. So from the first-person perspective, I face a practical problem: "How can I obtain $50 in a morally permissible way?" Suppose, also, that my good friend

---

[44] In making the claim that there is no individual action token about which to deliberate, O'Neill seems to be taking a stand on the metaphysics of future actions; in particular, she seems to be assuming that judgments about entities that do not yet exist must always be general rather than singular. For more on this issue, see William Godfrey-Smith (1978).

[45] Korsgaard (forthcoming-b), pp. 25-26.

[46] If we are to take to heart O'Neill's point about actions that are not yet performed, it must be an end type, rather than an end token, that the agent is tempted to promote.

Ignatz McGillicuddy has $50 in his wallet, and that he is an extremely gullible fellow. Seeking to solve my practical problem, I consider the following maxim: "I will make a lying promise to Ignatz McGillicuddy in order to dupe him out of $50." I test the maxim and find that—due to its excessive specificity—it is fully universalizable. So it appears that, from the first-person perspective, the Categorical Imperative deems it to be morally permissible for me to act on that maxim, and hence deems it to be morally permissible for me to swindle poor Ignatz out of his hard-earned cash. I take it that a moral verdict this wildly implausible constitutes a *reductio* of the current proposal.[47]

The only hope is to emend the proposal in some way. The main neo-Kantian strategy for doing so involves distinguishing between *the maxim that one tests with the CI-procedure* and *the maxim that one is actually tempted to act on*: the idea is that, even if I run the CI-procedure on an overly specific maxim such as "I will make a lying promise to Ignatz McGillicuddy in order to dupe him out of $50," really I am tempted to act on a much more general maxim such as "I will make a lying promise in order to secure some ready cash." Moreover, the neo-Kantian might add, "it is no objection to a moral theory that it has a principle of judgment which can be employed dishonestly."[48] But this variation of the original proposal will not do. First, there might well be a situation in which I *am* genuinely tempted to act on the maxim "I will make a lying promise to Ignatz McGillicuddy in order to dupe him out of $50"; it would be far too convenient a coincidence *if it just happened to turn out* that every sort of maxim that yields counterintuitive results when plugged into the CI-procedure from the first-person perspective is also the sort of maxim that we are never, in fact, tempted to act on. Second, the standard neo-Kantian way of revealing the genuine maxim that one aims to act on is to appeal to counterfactuals ("You can't really be tempted to act on such an overly specific maxim, since you'd still be tempted to make a lying promise to get the money even if you had to make it to Stan Spatz III instead of Ignatz McGillicuddy"), but as we saw earlier, appeals to counterfactual tests in order to determine the true content of a maxim are hopeless: sometimes they rule

---

[47] Note that this sort of counterintuitive result is not restricted to cases in which one contemplates whether to act on an overly specific maxim. In the same situation in which I am on the verge of bilking Ignatz out of his money, I might instead consider the maxim "I will say something to someone in order to suit a purpose of mine." Presumably a maxim this general is fully universalizable, so the Categorical Imperative seems to yield the verdict that, from the standpoint of practical reason, saying what I am about to say to Ignatz McGillicuddy is morally permissible.

[48] Herman (1990), p. 258.

out too much, sometimes they rule out too little, and sometimes they rule out both too little and too much at the same time. Third, and most importantly, this talk of applying the Categorical Imperative "dishonestly"—of running a rigged up, morally permissible maxim through the CI-procedure but then subsequently acting on a different, morally impermissible one—is to go back on the neo-Kantian's original idea that the Categorical Imperative is only a first-personal guide, not a third-personal standard, for now we have slid back into talk of the moral status of a *performed* action, rather than restricting ourselves to talk of the moral status of a *proposed* action. But then, when on a given occasion a given agent has performed a given action token, we are left with the question of which, of all the action types under which that action token falls, is the maxim on which the agent acted, and the generality problem is thrust upon us once again.

Indeed, upon further reflection, the very idea that there might be a first-personal guide without a third-personal standard for the item toward which it guides us is of dubious coherence. What is the point of directing us, from the first-person perspective, to act in certain ways, unless, once we have done so, we will satisfy some third-person requirement? What exactly is the Categorical Imperative guiding us toward, if as soon as we are done using it as a guide, actions cease to have that property (moral impermissibility) that it is supposedly helping us seek? Thus I conclude that Kantians need to find a place for *both* a third-personal standard *and* a first-personal guide within their moral system. Moreover, the candidates for such a standard and guide have been staring us in the face from the outset: the *Categorical Imperative* is the third-personal standard, and the *CI-procedure* is the first-personal guide. It is the Categorical Imperative that gives Kantians a criterion for what makes right actions right and wrong actions wrong. And it is the CI-procedure that gives them a procedure for figuring out, from the first-person perspective, which of the actions available to them are right and which are wrong. But if this is correct, then emphasizing the "standpoint of practical reason" does not dissolve the generality problem; it merely defers it.[49]

---

[49] The typical explanation given by neo-Kantians of why the Categorical Imperative cannot be a third-personal standard appeals to the *opacity of an agent's maxim*: according to Kant, we often are mistaken about what the true maxim of our action is; indeed, Kant holds that the human heart is so "unfathomable" that we cannot tell with certainty whether a single morally permissible action has ever been performed. (See Kant (1785/1997), AK 4:406-407, and Kant (1797/1996), AK 6:446-447.) However, this is a bad reason to reject the claim that the Categorical Imperative is a third-personal standard. It is no objection to

11. Toward a General Hypothesis on the Origins of the Generality Problem

We have found that the neo-Kantians' attempt to stress the practical nature of Kantian ethics does not help solve the generality problem. However, one feature of O'Neill's proposal will help point the way to a more unified explanation of why the generality problem arises for both reliabilism and Kantian ethics.

Recall that, in making out what separates the theoretical from the practical, O'Neill emphasizes the different *direction of fit* that each evinces: the theoretical, she insists, aims to fit the world, whereas the practical aims to shape the world. One way to apply this distinction involves distinguishing between two different "perspectives": from the theoretical perspective, so the story goes, one aims to form beliefs that fit the world, whereas from the practical perspective one aims to perform actions that shape the world. We saw that putting undue weight on these two perspectives proved to be a dead end, at least as far as finding a solution to the generality problem goes. But another way to apply the theoretical/practical distinction is to the *object* of evaluation in each case: the theoretical realm is concerned with *beliefs*, which attempt to fit the world, whereas the practical realm is concerned with *actions*, which attempt to shape the world.

This difference in the direction of fit between belief and action manifests itself in the way that each relates to the *reasons* there are for and against them. There are reasons for belief, and there are reasons for action. Moreover, in both cases, it is possible for there to be a reason for someone to believe/do something, and for that person to believe/do that thing, but not believe/do it *for* that reason. Or to appropriate some useful Kantian terminology, one can act *in accordance with* a given reason for action (that is, one can act as that reason recommends), without so acting *for the sake of* that reason for action. And similarly, one can believe *in accordance with* a given reason for belief, without so believing *on the basis of* that reason for belief. Note the difference, in these two cases, in the orientation of the connection between one's action/belief and the reason for which one is acting/believing. Acting for a reason is *forward-looking*:

---

the criterion "Two lines are parallel if and only if the distance between them is constant" that, given the inexactness of human measuring devices, we can never tell with certainty whether the distance between two actual lines is constant. Similarly, it is no objection to the principle of utility being a third-personal standard that we can never tell with certainty whether a given action does in fact produce at least as much net utility as each of its alternatives, since every action has indefinitely many consequences. Indeed, the *standard reply* to worries about the immeasurability of all of an action's consequences is to insist that the principle of utility is meant to be a third-personal standard, not a first-personal guide. In general, the issue of *what it takes for a standard to apply* is not the same as the issue of *what it takes for us to know that a standard applies*. So the opacity of an agent's maxim is no reason to reject the possibility of a third-person standard for right and wrong within Kantian ethics.

when one acts for a reason, one commences a course of conduct in order to shape the world in some way—or in other words, one acts *for the sake of* achieving some end (so we have an act-to-end direction of fit). Believing for a reason, on the other hand, is *backward-looking*: when one believes for a reason, one forms a belief in order to fit the evidence, experiential and otherwise, that one has about the world—or in other words, one forms a belief *on the basis of* some piece of evidence that one has (so in the case of experiential evidence, we have an experience-to-belief direction of fit).

This should all start to look familiar. Recall that when act token *a* is performed *for the sake of* end token *e*, neo-Kantians deem the compound <*a*, *e*> (they call it an action token) to be morally permissible iff <A, E> is universalizable, where <A, E> is token <*a*, *e*>'s morally relevant type. And when belief token *b* is formed *on the basis of* experience token *x*, reliabilists deem the compound <*x*, *b*> (they call it a process token) to be epistemically justified iff <X, B> is reliable, where <X, B> is token <*x*, *b*>'s epistemically relevant type. But now we can reformulate each theory's proposal in terms of action or belief for a reason. We can view the compound <*a*, *e*> as a neo-Kantian's way of glossing a token instance of an agent acting in some way for a reason. And we can view the compound <*x*, *b*> as a reliabilist's way of glossing a token instance of a subject believing something for a reason. Reformulated in this way, the two proposals now run as follows. Neo-Kantians deem a token instance <*a*, *e*> of an agent acting for a reason to be morally permissible iff <A, E> is universalizable, where <A, E> is token <*a*, *e*>'s morally relevant type. And reliabilists deem a token instance <*x*, *b*> of a subject believing something for a reason to be epistemically justified iff <X, B> is reliable, where <X, B> is token <*x*, *b*>'s epistemically relevant type.

There are three particularly salient similarities between these two proposals once we reformulate them in this manner. In both cases, we have the following:

1. The most basic object of normative appraisal (whether moral or epistemic) is a token instance of acting/believing for a reason.

2. Acting/believing for a reason is not cashed out in normative terms.

3. The means of normative appraisal is in terms of a non-normative property (universalizability, reliability) that can only be applied to types, not tokens.

I find these similarities between the two accounts far too suggestive to chalk up to mere coincidence.

Indeed, I now want to argue—in a somewhat speculative and very much preliminary way—that the second and third of these points of similarity both ultimately stem from the first.

Showing that the first point of similarity entails the second is fairly straightforward. Suppose that a given account of moral (or epistemic) appraisal cashes out the notion of acting (or believing) for a reason in normative terms; for example, maybe we have an independent account of normative reasons for action (or belief), which we then define acting (or believing) for a reason in terms of. But then the most basic object of moral (or epistemic) appraisal cannot be a token instance of acting (or believing) for a reason, since presumably the most basic object of normative appraisal within a given domain must itself be devoid of all normative content (or else it would not be the most basic such object). Thus the denial of point 2 entails the denial of point 1, so by contraposition, point 1 entails point 2.

Showing that the first point of similarity entails the third is a more delicate affair, and at this stage in my thinking I can only gesture at how it might be done. Let us call a property that can only be applied to types a *typal property*, and let us call a property that can be applied to instances of types an *instantial property*. Now suppose that the most basic object of moral appraisal is a token instance $<a, e>$ of someone acting for a reason, and that $<a, e>$ is appraised by means of some non-normative instantial property, F. For example, if our favored term of moral appraisal is moral permissibility, we might have the following:

$<a, e>$ is morally permissible iff $<a, e>$ is F.

But then there is nothing to stop us from defining, in the following manner, a notion of *morally permissible act token*, in addition to our notion of *morally permissible act-for-a-reason token*:

a is morally permissible iff $(\exists e)$(a is performed for the sake of e, and $<a, e>$ is F).

However, once we do this, then it is no longer true that token instances of someone acting for a reason are the uniquely most basic objects of moral appraisal, for act tokens can now be morally appraised without appeal to any additional normative notions beyond those used to appraise acting-for-a-reason tokens. (I intend basicness *qua* object-of-normative-appraisal to be determined by the degree to which normative notions must be appealed to in the explanation of why that object has the normative status that it does.) Thus the assumption that token instances of acting-for-a-reason are the most basic object of moral

appraisal and that those instances are appraised by means of a non-normative instantial property entails a contradiction. That is, point 1 and the denial of point 3 entails a contradiction, so point 1 entails point 3.[50]

A slightly more concrete example will help make this line of argument a bit more perspicuous. Consider a moral theory according to which (i) token instances of someone acting for a reason (in the Kantian sense of performing an act token for the sake of an end token) are the most basic objects of moral evaluation, and in which (ii) those instances are evaluated in terms of the instantial property *being approved of by God* (or more precisely: *being an act-for-a-reason token that is approved of by God*). Then in a case in which I perform the token act of making a lying promise for the token end of obtaining $50, we have the following biconditional: "My token act *a* of making a lying promise for the token end *e* of obtaining $50 is morally permissible iff God approves of my performing *a* for the sake of *e*." But in this case it seems we can easily extend the notion of moral permissibility so that it attaches to my act token in addition to my act-for-an-end token, in the following manner: "My token act *a* of making a lying promise is morally permissible iff God approves of my performing *a* for the sake of whatever end I performed it for." The basic point is this: when we are evaluating token instances of someone acting for a reason in terms of a property that directly applies to such tokens, *there is nothing to stop us from "detaching" the act token from its act-for-the-sake-of-an-end conglomerate and then evaluating that act token by itself.* So if we wish to hold on to the idea that it is the entire act-for-the-sake-of-an-end package that is the most fundamental unit of moral evaluation, we cannot evaluate token instances of that package in terms of an instantial property. And similarly, in the epistemic case, if we wish to hold on to the idea that it is the entire belief-plus-its-causal-history package (the reliabilist's way of explicating belief for a reason) that is the most fundamental unit of epistemic evaluation, we cannot evaluate token instances of that package in terms of an instantial property. Conclusion: our first point of similarity entails the third.

If the argument I have been sketching here is on track, then point 1 is the essential point of

[50] There are a number of places in which this argument—or better yet, argument-sketch—needs refining. More could be said about what distinguishes normative from non-normative accounts of acting/believing for a reason. What makes for basicness *qua* object-of-normative-assessment needs greater clarification. The crucial difference between typal and instantial properties could use further elaboration. And finally, the argument as presented assumes that a person only believes or does a thing for *one* reason, whereas cases in which people believe or do things for multiple reasons are clearly possible. In later work, I hope to fill in these lacunae.

similarity between process reliabilism and Kant's Formula of Universal Law: the essential link between the two is that, in both cases, it is token instances of acting or believing for a reason that are the most basic objects of normative appraisal. This leads me to propose the following hypothesis:

> *a modest proposal*: Any theory that takes token instances of *ϕ-ing for a reason* (where this is construed non-normatively) as the most basic object of normative appraisal within a given domain faces a version of the generality problem.

Why would this follow? If token instances of ϕ-ing for a reason are the most basic unit of normative appraisal, then (if what I have been suggesting is correct) this appraisal must be done by means of a *typal property*. But since typal properties apply to types, not tokens, we are left with the following question: of all the types under which a given token instance of ϕ-ing for a reason falls, which is the relevant type to apply our typal property to when evaluating that token instance of ϕ-ing for a reason? The generality problem is upon us.[51]

Moreover, my modest proposal leads to a less modest one:

> *a less modest proposal*: The generality problem is unsolvable, unless one appeals to an additional normative relation linking a token instance of *ϕ-ing for a reason* to its relevant type.

In other words, the non-normative typal property is not enough: in order to get from the token instance of ϕ-ing for a reason to the relevant (*normatively* relevant, that is) type to which we are to apply our typal property, we need an additional normative relation that itself requires explication in non-normative terms. Thus insofar as one of our desiderata for a successful solution to the generality problem is that additional normative standards over and above those provided by the relevant typal property (such as reliability or universalizability) not be brought in, the generality problem is—I put forward—unsolvable.

But this also means that the generality problem *can* be solved if we relax our standards for what would constitute an adequate solution to it. If we allow ourselves to appeal to an additional normative relation in moving from a token instance of ϕ-ing for a reason to its normatively relevant type, there is nothing to bar us from eventually finding a solution to the generality problem. The problem for

---

[51] Contemporary epistemologists often make a distinction between *propositional justification* (which concerns the degree to which one is justified in believing a given proposition, regardless of whether or not one actually holds that belief) and *doxastic justification* (which concerns the degree to which a belief that one actual holds, for the reasons that one actually holds it, is justified). In terms of this distinction, my modest proposal runs as follows: any epistemic theory that takes doxastic justification to be explanatorily prior to propositional justification faces a version of the generality problem.

reliabilism and Kantianism, however, is that once we allow in this additional normative relation, we lose much of the motivation for holding those views in the first place. One of the primary attractions of Kantian ethics is its positive arguments for the form and existence of the Categorical Imperative, but these positive arguments do not seem to leave any room for an additional normative element over and above universalizability in a specification of the content of the moral law. On other hand, one of the primary attractions of reliabilism is its aspirations toward being a truly naturalistic theory of epistemic evaluation that specifies necessary and sufficient conditions for epistemic justification in wholly non-normative terms. The central reliabilist gambit is that, as it were, we get truth for free: appealing to truth, and other properties built out of truth such as that of being a truth-conducive (i.e. reliable) process, is seen as a naturalistically respectable way of grounding epistemic properties like justification. However, if we now need an additional normative relation in order to evade the generality problem, the reliabilist must come up with a new trick for how to explicate that extra normative relation in wholly non-normative terms, and the worry is that she ultimately will be unable to do so. Thus I think that both the Kantian and the reliabilist should be extremely hesitant to countenance an additional normative relation that moves us from a token instance of acting/believing for a reason to its normatively relevant type. But if what I have argued is correct, they must do so in order to resolve the generality problem.[52]

In this section I have proposed, in an admittedly tentative and provisional way, the following hypothesis: a version of the generality problem arises for any theory that takes its most basic object of normative evaluation to be token instances of someone ϕ-ing for a reason (in a purely non-normatively sense of that notion). Like any hypothesis, this one yields a variety of predictions: it predicts, for example, that any moral theory that takes an agent's acting for a particular reason to be the most basic object of moral appraisal will face a version of the generality problem. And it predicts that some sort of generality

---

[52] Thus there is a certain sense in which I think that Herman is on the right track when she insists in her (1985) on the need for "rules of moral salience" fixing which facts it is appropriate to include in the content of a maxim (see n. 28), though it is important to emphasize that this move represents a major departure from Kant's original theory, and hence threatens to undermine the positive arguments for that theory. Similarly, in her recent work Korsgaard appeals to a robust Aristotelian metaphysics according to which all entities, including human action, have an essential function, partially—I suspect—to combat the generality problem; however, it is far from clear how Korsgaard can include this Aristotelian element in her theory without it swamping the Kantian elements that she wishes to hold onto, and without it undermining the Kantian arguments that she continues to make use of.

problem will beset any epistemic theory that takes a subject's believing something for a particular reason to be the most basic object of epistemic assessment. So the obvious next step in refining and defending this hypothesis is to see how these predictions hold up—to see to what degree my hypothesis succeeds in demarcating at least one fault line between those theories that confront a kind of generality problem and those that do not. Alas, however, this is a project that will have to wait for another day.[53]

I would like to close by noting one interesting feature of the hypothesis I have been putting forward. Neo-Kantians often stress that what is unique to Kantian ethics is its insistence on the act-for-the-sake-of-a-end compound being the fundamental object of moral assessment.[54] And Goldman, in his seminal defense of process reliabilism, argues vigorously that what allows process reliabilism to avoid the problems besetting traditional internalist theories of epistemic justification is its attention to the causal history of a belief (to the *reason* for which it was formed, in a non-normative sense) when determining the fundamental epistemic status of that belief-plus-causal-history pair.[55] So if what I have argued here is correct, the fact that a generality problem arises for both Kantian ethics and process reliabilism is not an accidental artifact of an inessential way in which those two theories have been formulated thus far; rather, it is one of the *defining characteristics* of those theories that generates the generality problem. But if this is so, then the generality problem is not an irritating quibble that requires little more than an additional theory of process individuation or maxim selection to fend off. Quite the contrary: the generality problem for Kantian ethics and process reliabilism is a troubling worry that cuts to the core of what makes each theory unique as an account of moral permissibility or epistemic justification.

---

[53] One initial worry for my hypothesis is that its focus on objects of normative appraisal and on the notion of ϕ-ing for a reason makes it ill suited for diagnosing why something like the generality problem arises for frequentist interpretations of probability and for Hume's constant-conjunction theory of causation, since neither of these two types of philosophical theory seems essentially normative, and since neither appeals in an obvious way to a notion of someone (or something) ϕ-ing for a reason. My hope is that further refinement of the hypothesis broached here will yield a yet-more-abstract characterization of the origins of the generality problem in the case of reliabilism and Kantianism—a characterization that could then be extended to the versions of the generality problem confronted by frequentist interpretations of probability and Hume's constant-conjunction theory of causation.

[54] For example, Herman (1992), p. 94, takes this feature "as defining of Kantian ethics," and throughout her (forthcoming-a), Korsgaard makes a similar claim.

[55] Goldman (1979), pp. 112-113.

Chapter 3: Luminous Conditions

1. Introduction

In the ongoing skirmishes between internalists and externalists in epistemology, one notion that seems particularly resistant to an externalist treatment is that of rationality. Perhaps I can fail to know or even be justified in believing that *p* without having epistemic access to whatever it is that deprives me of knowledge or justification. However, can it be *irrational* for me to believe that *p* when part of what makes that the case is outside my cognitive grasp, when from my first-person perspective everything indicates that it would be perfectly appropriate for me to so believe? Rationality seems to involve doing the best one can with what one has, and to posit an externalist standard of rationality according to which it can be irrational for one to believe a given proposition despite one's not being able, even upon reflection, to realize that this is so at best sounds like a strained use of the word "irrational," and at worst appears to be changing the subject.

Not so, according to Timothy Williamson. In *Knowledge and Its Limits*,[1] Williamson provides a sorites-like argument for the conclusion that *no* plausible standard of rationality meets the constraint that its demands are always accessible to the subject. In Williamsonian parlance, the fact that it would be irrational of one to believe that *p* is not *luminous*, for that fact can obtain without one's being in a position to know that it obtains. So not only is an externalist notion of rationality fully coherent, but it is the only game in town: even if we posit two notions of rationality, one of which (rationality_1) is claimed to be externalist and the other of which (rationality_2) is claimed not to be, the latter will be just as susceptible to Williamson's argument as the former. Therefore those who cling to the idea that rationality is inherently internalistic should cast off their prejudices and learn to love the bomb; sometimes we lack access to the demands of rationality, but such is our lot, and to hope for things to be otherwise is to quixotically yearn after an impossible cognitive standard.

Williamson's anti-luminosity argument has further consequences. If sound, not only would it show that it can be irrational for one to believe a given proposition without one's being in a position to know that it is, but it would also establish that one *can be in pain* without one's being in a position to know that

[1] Williamson (2000). All page references in the text are to this book.

one is; that one *can seem to see* a certain color without one's being in a position to know that one does; that two words *can mean the same thing* in one's idiolect without one's being in a position to know that they do; and that a given proposition *can be part of one's evidence* without one's being in a position to know that it is. In other words, the argument would show that one's current mental life, the meanings of one's words, the extent of one's evidence, and the dictates of rationality are all non-luminous—that each can, at least in principle, be epistemically inaccessible to a given subject. If Williamson is right, then we are (as he puts it) "cognitively homeless": there is no substantive domain of mental or semantic or normative facts to which we have guaranteed access, no subportion of our mental or semantic or normative lives within which everything lies open to view.

However, we need not accept Williamson's argument. Luminosity is a kind of *epistemic* privileged access: it involves a subject's always being in a position to know that a given fact obtains, whenever it does obtain. I will argue that Williamson's argument only succeeds if he assumes that we do not have a kind of *doxastic* privileged access (as we might put it) to the facts in question, for his argument presupposes that there does not exist a certain sort of constitutive connection between the obtaining of the given facts and our *beliefs* about the obtaining of those facts. The exact nature of this connection will depend on the version of Williamson's argument that one is considering—as we shall see, there are two versions of the anti-luminosity argument, depending on whether one defends the argument's crucial premise in terms of "all or nothing" belief or in terms of degrees of belief. But in either case, the thesis at the level of belief that Williamson must deny in order to secure his results at the level of knowledge is one that is independently plausible, and one to which defenders of luminosity will readily help themselves; to simply *assume* its falsity would beg the main question at issue. Only those who follow Williamson in his radical claims about how, even after ideal reflection, our beliefs about our mental lives, the meanings of our words, and the demands of rationality can swing free from the truth of these matters need accept his equally radical claims about our inevitable cognitive homelessness.

2. Williamson's Core Argument

In Williamson's terminology, a condition C is *luminous* if and only if the following holds:

(*) For every case α, if in α condition C obtains, then in α one is in a position to know that C obtains.

What Williamson calls *cases* are what Lewis, following Quine, calls centered possible worlds: possible worlds with a designated subject and a designated time. Williamsonian *conditions*, on the other hand, are in effect centered states-of-affairs: a condition either obtains or fails to obtain in a given case, and each condition can be specified by a that-clause in which the pronoun "one" refers to the case's designated subject and the present tense refers to the case's designated time. So, for example, the condition *that one has hands* obtains in a case α if and only if in α the subject of α has hands at the time of α. The expression "in a position to know" in (*) is potentially obscure, but for our purposes we only need the following: according to Williamson, "if one is in a position to know *p*, and one has done what one is in a position to do to decide whether *p* is true, then one does know *p*" (p. 95). The basic idea of (*), therefore, is that if a luminous condition obtains in a given case, then if one does not already know that the condition obtains, one *could* come to know that it does merely by taking the time to carefully reflect on the matter.

There are many conditions that are uncontroversially non-luminous, such as *that one has hands* (someone who is blind and paralyzed might have hands but not be in a position to know that she does). But there are also many conditions that it is extremely natural to take to be luminous, such as:

that one is in pain;
that one feels cold;
that it appears to one that p;
that one believes that q;
that words X and Y have the same meaning for one;
that one's evidence includes the proposition that r;
that one's evidence appears to include the proposition that s;
that it is rational for one to believe that t;
that it is rational for one to do φ.

Williamson argues that each of these conditions is not luminous; as he sees it, the only conditions that might, perhaps, be luminous are trivial conditions that either obtain in every case (such as the condition *that one exists*) or obtain in none (such as the impossible condition, for which (*) vacuously holds).[2]

---

[2] Although the details of his position have changed over the years, the philosopher who comes closest to explicitly endorsing a set of luminosity claims that play a central role in his epistemology is Roderick Chisholm; see his (1956), pp. 725-731; (1977), pp. 20-23; (1982), pp. 9-13; and (1997), pp. 27-29. However, most epistemological internalists are also committed to luminosity claims of one form or another.

Williamson's central anti-luminosity argument has the following form.[3] Let us fix on the condition *that one feels cold*, which Williamson takes to be as good a candidate for a luminous condition as any; later the argument will be generalized so as to apply to almost any putatively luminous condition. Williamson begins by asking us to "consider a morning on which one feels freezing cold at dawn, very slowly warms up, and feels hot by noon" (p. 96). We can stipulate that during this entire process, one does everything one can to decide whether or not one feels cold. We can also suppose that there is no other relevant change in the situation over time: all one does for the entire morning is sit there focusing on how hot or cold one feels as the temperature slowly gets warmer. Moreover, we can make the plausible assumption that one's feelings of hot and cold change so gradually during the course of the morning that one is not aware of any change in those feelings from one millisecond to the next. Let $t_0, t_1, t_2, \ldots t_n$ be a series of times at one-millisecond intervals from dawn to noon. For each integer i such that $0 \leq i \leq n$, let α_i be the case at time t_i on the morning in question. Finally, let C be the condition *that one feels cold*, let KC be the condition *that one knows that one feels cold*, and let PKC be the condition *that one is in a position to know that one feels cold*. Williamson's argument then proceeds as follows.

First, Williamson has us assume for *reductio* that the condition *that one feels cold* is indeed luminous. It follows that, for each integer i ($0 \leq i \leq n$), if in α_i one feels cold, then in α_i one is in a position to know that one feels cold:

(LUM) $(\forall i, 0 \leq i \leq n)$(C obtains in $\alpha_i \supset$ PKC obtains in α_i).

Second, Williamson notes that since in each α_i one is doing everything one can to decide whether one feels cold, it follows from his stipulation about the meaning of the expression "being in a position to know" that if in α_i one is in a position to know that one feels cold, then in α_i one does in fact know that one feels cold. Thus we have the following:

(POS) $(\forall i, 0 \leq i \leq n)$(PKC obtains in $\alpha_i \supset$ KC obtains in α_i).

Third, Williamson appeals to what is perhaps the single most important assumption in his entire book. According to Williamson, *knowledge requires safety from error*: in order for one to know something, one must

---

[3] See Williamson (2000), ch. 4. An earlier version of the argument can be found in Williamson (1996).

not have been easily wrong in coming to believe it. Much more will be said about Williamson's safety requirement in §§4-5 below, but for now the following comments will suffice. Suppose that in case α_i one believes that one feels cold. Williamson insists that in order for one's belief in α_i to be safe enough to constitute knowledge, one's belief must not be false in any similar case that one cannot discriminate from α_i. Now one such case is α_{i+1}, the case one millisecond later, so it follows that if one *knows* in case α_i that one feels cold, it must still be *true* in α_{i+1} that one feels cold. As this reasoning will work just as well for any integer i such that $0 \leq i < n$, the safety requirement on knowledge plus one's limited discriminatory capabilities give us the following *margin-for-error principle*:

(MAR) $(\forall i, 0 \leq i < n)$(KC obtains in $\alpha_i \supset$ C obtains in α_{i+1}).

Finally, we have two last assumptions that follow merely from the description of the case:

(BEG) C obtains in α_0.

(END) C does not obtain in α_n.

That is: at dawn one feels cold, and at noon one does not feel cold.

However, now contradiction looms. By (LUM), if C obtains in α_0, then PKC obtains in α_0. By (POS), if PKC obtains in α_0, then KC obtains in α_0. By (MAR), if KC obtains in α_0, then C obtains in α_1. Therefore from these three conditionals and (BEG) it follows that C obtains in α_1. Moreover, by a similar chain of reasoning, we may conclude that C obtains in α_2, that C obtains in α_3, and so on, until we reach the conclusion that C obtains in α_n. But this contradicts (END). Thus one of the argument's five premises must be false. Williamson claims that (POS), (MAR), (BEG), and (END) are all unassailable, so he infers that (LUM) is the premise responsible for our reaching a contradiction. Conclusion: the condition *that one feels cold* is not luminous—one can feel cold without being in a position to know that one feels cold.

All the above argument assumes about the condition *that one feels cold* is that there exists a continuum of cases, starting from a case in which that condition obtains and ending with a case in which it does not, such that the underlying basis for the condition's obtaining or not obtaining changes so gradually that one cannot discriminate a change in that basis from one case to the next. Thus we can run a parallel argument on any condition for which such a continuum exists. We could argue that the

condition *that one is in pain* is not luminous by considering a series of cases in which one feels an agonizing pain that gradually subsides until one feels nothing at all. We could argue that the condition *that it appears that there is a computer in front of one* is not luminous by considering a series of cases in which one at first clearly sees a computer but then one's eyesight gradually gets blurrier and blurrier. Perhaps we could even argue that the condition *that words X and Y have the same meaning* is not luminous by considering a series of cases in which "two synonyms . . . gradually diverge in meaning, as a mere difference in tone grows into a difference in application" (p. 106). And so on: although the details of how we construct these continua of cases will vary depending upon the condition in question, it seems plausible that if Williamson's argument succeeds in showing that the condition *that one feels cold* is not luminous, analogous arguments could show that almost any other condition is not luminous. In particular, as a process of gradual change can take one from circumstances in which it is rational for one to believe that *p* to circumstances in which it is irrational for one to believe that *p*, we can establish that, for any sense of "rational," the condition *that it is rational for one to believe that p* is not luminous: it can be rational, in that sense, for one to believe a given proposition without one's being in a position to know that it is.[4]

Williamson's anti-luminosity argument forms the backbone of *Knowledge and Its Limits*: not only does he take it to show that no non-trivial condition is luminous, but he also uses the argument (or variants of it) to reply to an important objection to his claim that knowledge is a mental state (ch. 4); to contest Dummett's argument for an anti-realist theory of meaning (ch. 4); to argue against a version of the KK-principle according to which one is always in a position to know when one knows a given proposition (ch. 5); to provide a solution to the surprise examination paradox (ch. 6); to rebut any argument for skepticism about the external world that assumes that we and the envatted versions of ourselves possess the same evidence (ch. 8); and to buttress his claim that our evidence is all and only what we know (ch. 9). Moreover, similar uses of a safety requirement on knowledge to derive margin-for-error principles play a crucial role in Williamson's theory of vagueness.[5] And more generally, Williamson's anti-luminosity

[4] Williamson insists that such a continuum of cases exists for *it is rational for one to believe that p* at Williamson (2004), p. 315; see also Williamson (2005), pp. 481-482.

[5] It is these margin-for-error principles that allow Williamson to explain why, even though on his epistemicist theory of

argument constitutes a novel way of criticizing a venerable philosophical tradition—a way which Williamson uses to lay the foundation for a radical new theory of (as he puts it) knowledge and its limits.

3. Narrowing the Target

Such is Williamson's core anti-luminosity argument; but is it sound? The argument works by generating a contradiction from the five premises (LUM), (POS), (MAR), (BEG), and (END). The last two of these are undeniable: they follow directly from the set-up of the scenario. (POS), on the other hand, one might doubt: it depends on assuming that in each case one has done everything one can to decide whether one feels cold, and some might insist that doing this takes longer than one millisecond. However, we can easily avoid this worry by extending the amount of time between successive cases while keeping the total number of cases the same (so that the entire process takes longer than a single morning), and once we change the set-up in this way, (POS) seems fine.[6] Thus the crucial question when evaluating Williamson's argument is whether (MAR) is more plausible than (LUM).

Many people, when first encountering Williamson's argument, find (MAR) independently plausible and so not in need of any further justification. However, resting content with the *prima facie* plausibility of (MAR) reduces Williamson's argument to a mere battle of intuitions over which of (LUM) and (MAR) one finds more plausible, and it would then be open to the defender of luminosity who finds (LUM) compelling to simply *tollens* Williamson's *ponens*. Moreover, we should be wary of trusting our intuitions about principles that involve applying vague predicates to very similar cases: after all, the typical sorites premise "One hair can't make the difference between being bald and not bald" is intuitively *extremely* compelling. So if the only grounds we had in support of (MAR) were our bare intuitions about its seeming plausibility, those would be very thin grounds indeed, particularly in light of the wide variety of radical consequences that Williamson takes to follow from his anti-luminosity argument.

---

vagueness each vague predicate has a sharp cut-off point, we can never know where those sharp cut-off points lie. See Williamson (1992a), §5; Williamson (1992b), §6; and Williamson (1994), ch. 8.

[6] Actually, matters are slightly more complicated than that, for as long as one's feeling of cold is continually changing between successive cases, we might worry that one will not be coming to a decision about a constant feeling of cold. So if we want to completely avoid this type of worry, we should make one more change to the set-up and, instead of having one's feeling of cold change *continuously* over a given interval, have it change *in incremental steps*.

What we want is an independent motivation for (MAR). As mentioned before, Williamson defends (MAR) by appealing to a safety requirement on knowledge. A recurring theme in the Williamsonian *oeuvre*, from his early work on indiscriminability and vagueness on through his recent material more directly concerned with epistemological matters, is that in cases in which we have imperfect abilities to discriminate between alternatives, *a safety requirement on knowledge* plus *our limited powers of discrimination* yield *margin-for-error principles*.[7] A margin-for-error principle is any principle that, like (MAR), is of the form "If in case α one knows that p, then in sufficiently similar case β it is true that p." And the safety requirement is Williamson's way of cashing out the idea that reliability is a necessary condition for knowledge: although in earlier work Williamson talks exclusively of a reliability constraint on knowledge without mentioning the word "safety," by *Knowledge and Its Limits* he passes freely back and forth between talk of reliability and talk of safety.[8]

What exactly is Williamson's safety requirement? And how exactly does it lead to the margin-for-error principle (MAR)? Unfortunately, this matter is complicated by the fact that Williamson explicates the safety requirement in two different ways, the first involving a coarse-grained, "all or nothing" conception of belief, the second involving a more fine-grained conception in terms of degrees of confidence. So in order to assess the case Williamson makes for (MAR), we should consider each of these ways of filling out the safety requirement in turn.

4. Coarse-Grained Safety

Williamson often expresses the idea that knowledge requires safety from error in terms of one's not easily being wrong in similar cases: "if one believes p truly in a case α, one must avoid false belief in cases sufficiently similar to α in order to count as reliable enough to know in α" (p. 100); "in case α one is safe from error in believing that C obtains if and only if there is no case close to α in which one falsely believes that C obtains" (pp. 126-127); "if one knows, one could not easily have been wrong in a similar case. In

---

[7] See Williamson (1990), pp. 104-106; (1992a), pp. 223-226; (1994), pp. 226-230; and (2000), pp. 17-18.

[8] Note that Williamson's identification of reliability with safety is very different from the process reliabilist's identification of reliability with truth-conducivity. For more on the process reliabilist's notion of reliability and some of its pitfalls, see chapter 2 of this dissertation.

that sense, one's belief is safely true" (p. 147). Passages such as these strongly suggest that the safety requirement involves the following necessary condition on knowledge:

(C-SAFETY) In case α one knows that p only if, in all sufficiently similar cases in which one believes that p, it is true that p.[9]

When so construed, the safety requirement has an undeniable air of plausibility: if it easily could have been the case that one falsely believes that p—if in an extremely similar way the world might have been, one believes that p and this belief is false—then one's actual belief that p, even if true, does not seem secure enough to constitute knowledge. Suppose I ask you to think of a number between 1 and 10 and correctly guess that you are thinking of 6; my true belief about which number you are thinking of does not count as knowledge. Why? Plausibly because there is a very similar case in which I still guess that you are thinking of 6, but my belief is wrong since you are really thinking of, say, 4. Suppose I decide, out of mere pessimism, that the lottery ticket I just purchased is not a winning ticket; even if I end up being right, my belief that the ticket is not a winner does not count as knowledge. Why? Plausibly because there is a very similar case in which I still believe out of pessimism that the ticket is not a winner, but the balls determining the winning ticket number bounce slightly differently so as to make my ticket a winner. Of course, how one fixes the similarity relation between cases and what determines the threshold beyond which two cases do not count as sufficiently similar will no doubt be murky matters, but there seems little point in denying that (C-SAFETY) has much to be said in its favor.[10]

It is crucial to notice that (C-SAFETY) is not itself a margin-for-error principle. According to a given margin-for-error principle, one does not know that p in some case if, in a certain sufficiently similar

---

[9] This coarse-grained version of a safety requirement on knowledge is roughly equivalent to one that Ernest Sosa has endorsed in a recent series of articles; see Sosa (1996), (1999a), (1999b), (2000), (2002), and (2004). Note, however, that Sosa's formulation of the safety requirement involves ascribing non-standard truth conditions to the subjunctive conditional "S would believe that p only if it were the case that p"—truth conditions that Williamson does not necessarily endorse: see Williamson (2000), p. 149.

[10] Several recent articles have attempted to provide counterexamples to (C-SAFETY) as a way of resisting Williamson's anti-luminosity argument; see Brueckner & Fiocco (2002), Neta & Rohrbaugh (2004), Comesaña (2005), and Conee (2005). For Williamson's reply, see §I of his (forthcoming). I agree with Williamson's assessment that the alleged counterexamples offered by these authors are not convincing: each aims to describe two sufficiently similar cases such that one knows a given proposition in the first case and falsely believes that same proposition in the second, but for each pair of cases, either it is far from clear that the first case is a genuine case of knowledge, or it is far from clear that the two cases are sufficiently similar in the relevant respects. (This is all the more true when the safety requirement is modified so that one must have sufficiently similar bases of belief in the two cases: see n. 15.)

case, it is false that *p*. According to (C-SAFETY), one does not know that *p* in some case if, in a sufficiently similar case, it is false that *p and* one believes that *p*. For all (C-SAFETY) says, one might know that *p* in some case α despite its being false that *p* in an extremely similar case α^*, provided that one does not believe that *p* in α^*. It is only nearby *false belief* that, according to (C-SAFETY), blocks one from having knowledge, not nearby *falsity of what is actually believed.*

Recall that the margin-for-error principle needed for Williamson's anti-luminosity argument to go through is as follows:

(MAR) $(\forall i, 0 \leq i < n)$(KC obtains in $\alpha_i \supset$ C obtains in α_{i+1}).

How can one derive (MAR) from (C-SAFETY)? Let us concede to Williamson that each case α_i is sufficiently similar to the case α_{i+1} one millisecond later. Then if BC is the condition *that one believes that one feels cold*, straightforward substitution into (C-SAFETY) yields the following:

(SAF) $(\forall i, 0 \leq i < n)$[KC obtains in $\alpha_i \supset$ (BC obtains in $\alpha_{i+1} \supset$ C obtains in α_{i+1})].

So in order for us to be able to use the coarse-grained version of the safety requirement to justify the crucial premise of Williamson's anti-luminosity argument, we must somehow get from (SAF) to (MAR).

How to do so is not difficult to see. Williamson's guiding thought is that for cases in which we have limited discriminatory capabilities, the safety requirement on knowledge gives rise to a margin-for-error principle. So what we need is some premise encapsulating our subject's inability to discriminate the cases α_i from one millisecond to the next. The most obvious candidate is as follows:

(BEL) $(\forall i, 0 \leq i < n)$(BC obtains in $\alpha_i \supset$ BC obtains in α_{i+1}).

The basic idea behind (BEL) is that because the change from t_i to t_{i+1} in the underlying basis for one's belief that one feels cold is beyond the threshold of one's discriminatory capacities, one's belief at t_{i+1} as to whether one feels cold will be the same as one's belief at t_i. (BEL) is a natural way of articulating the idea that one cannot discriminate case α_i from case α_{i+1} with regards to how cold one feels—that, as Williamson puts it, there is "limited discrimination in the belief-forming process" (p. 127). Moreover, once we have (BEL), (MAR) follows from (SAF), given the additional assumption that knowledge implies belief.

So have we managed to adequately justify (MAR)? No, we have not—*for the crucial premise (BEL) is a*

sorites premise. Indeed, (BEL) by itself is enough to generate a sorites paradox from the undeniable assumptions that (i) one believes that one feels cold in case α_0 (i.e. at dawn), and (ii) one does *not* believe that one feels cold in case α_n (i.e. at noon): with n uses of *modus ponens* on an instance of (BEL), we can get from the first of these assumptions to the negation of the second. So we should reject any argument that appeals to (BEL); depending on one's theory of vagueness, the principle is either less than perfectly true or outright false. And if Williamson's anti-luminosity argument does indeed implicitly appeal to (BEL), then that argument is open to the charge of illicitly exploiting the vagueness of the term "believes," in much the same way as a typical sorites argument illicitly exploits the vagueness of a term such as "bald" or "heap."[11]

However, might not Williamson derive (MAR) from (SAF) by means of some principle other than (BEL)? Another obvious candidate is

(KNO) $(\forall i, 0 \leq i < n)$(KC obtains in $\alpha_i \supset$ BC obtains in α_{i+1}).

(KNO) and (SAF) together yield (MAR), without our even having to assume that knowledge implies belief. But what is the independent motivation for (KNO), other than its being what Williamson needs in order to derive (MAR)? It's not as if (KNO) encapsulates some dictum about the subject's inability to discriminate between successive cases—or if it does, it only does so derivatively, in virtue of being implied by (BEL) and

---

[11] A similar objection applies to attempts to save Williamson's argument by appealing to a modalized version of (BEL). It might be insisted that if one believes that one feels cold in case α_i, then even if one does not believe that one feels cold in the case α_{i+1} one millisecond later, there must exist at least one other possible case β similar to α_i in which one has the same qualitative feeling of cold as in α_{i+1} and in which one believes that one feels cold; in symbols:

(BEL′) $(\forall i, 0 \leq i < n)$(BC obtains in $\alpha_i \supset (\exists \beta$ similar to $\alpha_i)[Q(\beta)=Q(\alpha_{i+1}) \wedge$ BC obtains in $\beta]$),

where $Q(\alpha)=Q(\beta)$ signifies that one's qualitative feeling of cold is the same in cases α and β. (Cf. Williamson (2000), p. 127.) Moreover, it is extremely plausible that whether one feels cold in a given case is determined by one's qualitative feeling of cold, so if one has the same qualitative feeling of cold in two cases, one feels cold in one of those cases if and only if one feels cold in the other:

(QUAL) $(\forall \alpha, \beta)[Q(\alpha)=Q(\beta) \supset$ (C obtains in α iff C obtains in β)].

From (BEL′), (QUAL), (C-SAFETY), and the assumption that knowledge implies belief, we can derive (MAR). However, just as repeatedly iterating (BEL) leads to unacceptable consequences, repeatedly iterating (BEL′) leads to unacceptable—or at least highly controversial—consequences. Given the undeniable assumption that one believes that one feels cold in α_0, repeated applications of (BEL′) yields the result that there exists some case β in which one feels as hot as one does in α_n (the case at noon in our original scenario), and yet one nonetheless believes that one feels cold. Indeed, if we tweak our original scenario and stipulate that in the final case things have heated up to such a degree that one's qualitative feeling of hot is the same as it would be were one at the center of the sun, then the defender of (BEL′) is saddled with the result that there is a possible case β such that *one feels as if one were in the center of the sun, and yet one believes that one feels cold.* Of course, since the similarity relation is not transitive, β will be very distant from any case in the actual world. But I think we should have serious doubts that such a case is even possible—serious doubts that there could exist a being who counts as having beliefs and experiences, and yet whose beliefs and experiences are as wildly at odds with one another as they would be in β. To think otherwise is to think that the cognitive and phenomenal realms can come apart from each other to an unacceptable degree. (Compare the discussion of Sosa's thought experiment in §6 below.)

the claim that knowledge implies belief. And it doesn't seem plausible that (KNO) follows from some more general claim that knowing that *p* in some case requires that one believe that *p* in all sufficiently similar cases. Suppose I know a given proposition in the actual case; in a very similar case I might be suspending judgment on the matter, or irrationally clinging to a belief in the proposition's negation—why should *that* bar me from knowing in the actual case? Though knowledge might indeed require a protective belt of cases in which one does not falsely believe, it is extremely implausible to suppose that, in addition, knowledge requires a protective belt of cases in which one believes.

But more importantly, and independent of the specific failings of (BEL) and (KNO), the more general point is this: if one knows that *p* in some case, (C-SAFETY) has *nothing to say* about similar cases in which one does not believe that *p*; at some point during the morning one will stop believing that one feels cold; so (C-SAFETY) has *nothing to say* about whether one really does feel cold from that point on. In particular, as it is incontestable that BC does not obtain in case α_n, (SAF)—and hence (C-SAFETY)—will be completely useless in deriving the conditional "KC obtains in $\alpha_{n-1} \supset$ C obtains in α_n," which is one of the instances of (MAR). The basic purpose of (MAR) in Williamson's anti-luminosity argument is as a *bridge principle* between cases. From (LUM) and (POS) it only follows that if some condition obtains in a given case, then some other condition obtains *in that same case*; with (MAR), on the other hand, we can deduce that because a certain condition obtains in case α_i, a certain other condition must obtain in the successive case α_{i+1}. However, (C-SAFETY) will be unable to fully undergird (MAR), since (C-SAFETY) can act as a bridge principle between successive cases α_i and α_{i+1} only if one believes that one feels cold in both; as this will not be true for all integers i such that $0 \leq i < n$, we will need some *other* bridge principle to secure (MAR) in those cases, and I claim that whatever this additional principle is, it will be implausible.

One might reply on Williamson's behalf: all we need in order to run the *reductio* is a single case in which one falsely believes that one feels cold, so we don't need to go all the way to α_n. Suppose there is some integer j such that: (i) for every non-negative integer $i < j$, both C and BC obtain in α_i, and (ii) BC obtains in α_j but C does not. Then we could use (BEG), (LUM), (POS), and (SAF) to generate the contradictory result that C both does and does not obtain in α_j. However, who is to say that such a j

exists—that as one gradually gets warmer and warmer during the course of the morning while carefully attending to how cold one feels, one stops feeling cold *before* one stops believing that one feels cold? Williamson appears willing to grant to his opponent that there might be "a constitutive connection between the obtaining of the condition [that one feels cold] and one's judging it to obtain" (p. 100), and some candidates for such a constitutive connection rule out the possibility that our subject stops feeling cold before she stops believing that she feels cold.[12] The weakest version of a constitutive connection that has this result is the following:

(CON) If one has done everything one can to decide whether one feels cold, then one believes that one feels cold only if one feels cold.

Since on the morning in question our subject has done everything she can to decide whether she feels cold, it would follow from (CON) that she never believes that she feels cold on that morning without in fact feeling cold. Of course one might doubt that there exists a constitutive connection of this form between feeling cold and believing that one feels cold, but then we need some independent argument against the possibility of such a connection, which the anti-luminosity argument by itself does not provide. Indeed, as Williamson himself notes (p. 100), typically it is *precisely because* they think that there is a tight connection between certain mental states and beliefs about those states that some philosophers claim the mental states in question to be luminous.[13] So to simply assume that (CON) is false would beg the question against the defender of luminosity.[14]

I conclude that (C-SAFETY) is unable by itself to motivate the margin-for-error principle (MAR). However, at times Williamson talks as if (C-SAFETY) is only a first approximation to the proper coarse-grained version of the safety requirement. "A more elaborate account on such lines," he writes, "would

---

[12] A number of authors in the philosophy of mind literature have recently defended accounts according to which there is a constitutive connection between experience and (the exercise of) certain so-called phenomenal concepts, which in turn gives rise to a corresponding constitutive connection between experience and certain phenomenal beliefs that employ those concepts: see Papineau (2002), ch. 4; Chalmers (2003); Block (2006); and Balog (forthcoming).

[13] See, for example, Chisholm (1956), p. 725, and (1982), p. 10.

[14] In his (2004), Brian Weatherson reaches a similar conclusion that a coarse-grained version of Williamson's anti-luminosity argument would be blocked if on the morning in question one believes that one feels cold only if one feels cold. Weatherson makes the point by considering the possibility that *the very same brain state* might constitute the state of one's feeling cold and the state of one's believing that one feels cold, but we need not make so strong an assumption in order to block the argument: all we need is the much weaker claim (CON). Moreover, Weatherson fails to consider how this insight can be extended to Williamson's official defense of (MAR) in terms of a fine-grained version of the safety requirement.

qualify '[one] believes *p*' in the conditional to exclude cases in which [one] believes *p* on a quite different basis from the basis on which [one] believes *p* in the case in which [one] putatively knows *p*" (p. 149). Williamson seems to be suggesting that a more accurate version of the coarse-grained safety requirement would read as follows:

> (C-SAFETY′) In case α one knows that *p* on basis *b* only if, in any sufficiently similar case α* in which one believes that *p* on a sufficiently similar basis *b**, it is true that *p*.

Suppose that in some case α one believes that *p* on a given basis, and in a sufficiently similar case β one believes that *p* on a very different basis (perhaps one believes that *p* by perception in case α and by testimony in case β). Then (C-SAFETY′) captures the very natural idea that if one's belief is true in α and false in β, one's false belief in β should not impugn the reliability (and hence the status as knowledge) of one's true belief in α—after all, one came to believe that *p* on a very different basis in each case.

Thus the shift from (C-SAFETY) to (C-SAFETY′) lends more plausibility to the claim that safety is indeed a necessary condition for knowledge.[15] However, once we have made this move, we are now owed an account of what makes one basis for belief sufficiently similar to another. Suppose I come to believe that there is a cup on the table on the basis of visual perception in dim lighting. Does every sufficiently similar case in which I believe the same thing on the basis of visual perception count as a case in which I believe that proposition on a sufficiently similar basis? Or only cases in which the retinal image on my eye is approximately the same? Or only cases in which the *source* of the retinal image is approximately the same? Or only cases in which the lighting conditions are approximately the same? Depending on how we answer these questions, the safety of my belief—and hence its status as knowledge—will vary. And as readers of chapter 2 of this dissertation should be well aware, this issue is nothing more than the dreaded *generality problem* for reliabilist theories of knowledge and justification, which is just as much a problem for Williamson's safety requirement as it is for any other reliabilist view. Williamson concedes that his appeal to the safety requirement opens him to the generality problem, but he seems to be confused about what

---

[15] The shift to (C-SAFETY′) also provides one more tool in resisting the would-be counterexamples to the safety requirement mentioned in n. 10, since many of those examples fix on two cases in which the bases of one's belief are not sufficiently similar (especially if we individuate bases of belief externally, so that phenomenologically indistinguishable bases of belief need not count as sufficiently similar).

implications this has for his argument. Williamson writes that if the generality problem is insoluble, then the upshot is merely that "the concept of reliability cannot be defined in independent terms" (p. 100); however, the upshot is much more dire than that: if the generality problem is genuinely insoluble, then every reliabilist theory—Williamson's included—either is unacceptably *ad hoc*, or else gives rise to wildly counterintuitive results about what counts as knowledge.[16]

That reliabilist theories of knowledge face the generality problem is nothing new. Thus I propose that we set this worry aside and pretend that Williamson has a response to it. I also propose that we grant Williamson the truth of (C-SAFETY′). There is enough flexibility in what counts as both a sufficiently similar case and a sufficiently similar basis of belief that one can probably avoid any putative counterexample to the principle (indeed, it is this very flexibility that gives rise to the generality problem). Moreover, as mentioned earlier, there is something undeniably attractive about this way of cashing out the safety requirement. So let us concede to Williamson that (C-SAFETY′) holds. I shall now argue that, even if (C-SAFETY′) is true, it provides no help with the criticisms raised earlier against the possibility of deriving the needed margin-for-error principle (MAR) from a coarse-grained version of the safety requirement.

In order to use (C-SAFETY′) to derive (MAR), we need it to be the case that if in some case α_i on the morning in question one knows that one feels cold on a given basis for belief, then in the case α_{i+1} one millisecond later one believes that one feels cold on a sufficiently similar basis. But if one believes on a sufficiently similar basis that one feels cold, then *a fortiori* one believes *simpliciter* that one feels cold. Thus, just as in our attempt to derive (MAR) using (C-SAFETY) as the only bridge principle between cases, in order to derive (MAR) using (C-SAFETY′) as the only bridge principle it must be true that one believes that one feels cold in each successive case. And as this is not true for the morning in question, it will be impossible to use (C-SAFETY′) to motivate (MAR) without appealing to some sorites-premise-like principle to the effect that if one believes that one feels cold on a given basis in case α_i, then one believes that one feels cold on a sufficiently similar basis in the subsequent case α_{i+1}. But we should reject any argument

[16] See chapter 2 of this dissertation for further discussion.

that appeals to such a principle. So moving to (C-SAFETY′) is of no help in attempting to derive (MAR) from the safety requirement.[17, 18]

5. Fine-Grained Safety

For most of *Knowledge and Its Limits*, when Williamson discusses the safety requirement it is the coarse-grained version of the safety requirement that is explicitly mentioned. And when Williamson provides a formal model of how luminosity might fail (pp. 127-130), his model invokes a conception of safety formulated in terms of coarse-grained, "all or nothing" belief. However, during Williamson's official defense of (MAR), he instead appeals to a more fine-grained notion of safety—as Williamson puts it, his

---

[17] At one point Williamson mentions in passing a third version of a coarse-grained safety requirement, one that provides even less support for (MAR) than the other two. Williamson writes, "If at time t on basis b one knows p, and at a time t^* close enough t on a basis b^* close enough to b one believes a proposition p^* close enough to p, then p^* should be true" (p. 102). This quotation suggests the following version of a coarse-grained safety requirement:

(C-SAFETY″) In case α one knows that p on basis b only if, in any sufficiently similar case α^* in which one believes a sufficiently similar proposition that p^* on a sufficiently similar basis b^*, it is true that p^*.

(C-SAFETY″) avoids one problem that potentially faces (C-SAFETY) and (C-SAFETY′): we might want to use the safety requirement to rule out lucky guesses about *necessary truths* from being knowledge, but if I correctly guess, say, that 853 is prime, there will be no sufficiently similar cases in which I falsely believe that 853 is prime, since what I believe is true in *every* possible case. (Cf. Sainsbury (1995), p. 595; Weatherson (2004), p. 378; Williamson (2000), pp. 181-182; Williamson (2005), p. 472.) However, as a means of motivating (MAR), (C-SAFETY″) fares even worse than (C-SAFETY) and (C-SAFETY′). As one gradually feels warmer and warmer on the morning in question, eventually one will stop believing that one feels cold and instead believe a proposition that one might express by saying, "I feel coldish." So (C-SAFETY″), plus various background assumptions about the sufficient similarity of both the propositions believed and the bases on which one believes them, will yield the result that if in the previous case one knew that one felt cold, then in the present case it must be true that one feels cold*ish*. And repeated appeals to (C-SAFETY″) in this way will eventually yield the result that in α_n one feels hot—hardly an absurd conclusion!

[18] The recurring problem in these attempts to motivate (MAR) by means of a coarse-grained version of the safety requirement is that, in each case, we need some way of ensuring that one continues to believe that one feels cold from one millisecond to the next on the morning in question. Thus we might ask: can we abandon these attempts to derive (MAR) from a coarse-grained safety requirement, and instead attempt to derive it from a safety-like requirement that makes no reference to belief? The safety-like requirement that readily springs to mind is the following:

(SIM) In case α one knows that p only if, in all sufficiently similar cases, it is true that p.

However, (SIM) is extremely implausible, for reasons similar to those given against (KNO): why should its not being the case that p in nearby cases obstruct my knowing that p in the actual case if I don't believe that p, or even engage in any belief-forming process vaguely similar to that which I used in forming or retaining my belief that p, in those nearby cases? Indeed, unless we severely restrict what counts as a sufficiently similar case, (SIM) would appear to block almost any belief from counting as knowledge. For example, suppose I am gazing at a leaf on a nearby tree and truly believe that I am seeing a green leaf. Presumably a case whose only difference from the actual one consists in the leaf in question being brown, or in its not being in front of my gaze since it has already fallen to the ground, is similar in nearly all respects to the actual one; yet if we count such cases as sufficiently similar, we are saddled with the absurd consequence that I do not know that I am seeing a green leaf in the actual case. Of course, we could deem these cases as *not* being sufficiently similar, but then we lose the ability of (SIM) to explain the examples used to motivate a safety-like requirement in the first place—for instance, in the lottery example there are any number of differences between the actual case in which my lottery ticket is a winner and the counterfactual case in which my lottery ticket is a loser that are far greater than a difference in the color or position of one mere leaf. As such we would need some new motivation for taking (SIM) to be a genuine necessary condition on knowledge, and how we provide such a motivation is not readily apparent. But more pressingly, the defender of (SIM) who counts the counterfactual leaf cases as not sufficiently similar to the actual case presumably does so partially in virtue of the differences in one's *beliefs* between the actual and counterfactual cases, which would imply that a counterfactual case only counts as sufficiently similar if one has all the same beliefs in that case as in the actual case; however, this makes (SIM) a mere notational variant of (C-SAFETY), and thus all the same problems will arise as before when one attempts to use (SIM) to derive (MAR).

argument here "depends on applying reliability considerations in a subtler way to degrees of confidence" (p. 127).[19] I suspect that Williamson constructs the official argument in terms of a fine-grained version of the safety requirement specifically because he wants the argument to hold even if something like (CON) is true: the real anti-luminosity argument is supposed to be one that even the staunchest defender of constitutive connections must accept. The problem, however, is that whereas the coarse-grained conception of safety is at least somewhat compelling as a way of articulating the general idea that knowledge requires reliably true belief, the fine-grained version has no such appeal. Or so I shall argue.

It will help if I quote Williamson's initial justification of (MAR) in terms of degrees of confidence in its entirety (p. 97):

> Consider a time t_i between t_0 and t_n, and suppose that at t_i one knows that one feels cold. Thus one is at least reasonably confident that one feels cold, for otherwise one would not know. Moreover, this confidence must be reliably based, for otherwise one would still not *know* that one feels cold. Now at t_{i+1} one is almost equally confident that one feels cold, by the description of the case. So if one does not feel cold at t_{i+1}, then one's confidence at t_i that one feels cold is not reliably based, for one's almost equal confidence on a similar basis a millisecond earlier that one felt cold was mistaken. In picturesque terms, that large portion of one's confidence at t_i that one still has at t_{i+1} is misplaced. Even if one's confidence at t_i was just enough to count as belief, while one's confidence at t_{i+1} falls just short of belief, what constituted that belief at t_i was largely misplaced confidence; the belief fell short of knowledge. One's confidence at t_i was reliably based in the way required for knowledge only if one feels cold at t_{i+1}.

In this argument, Williamson appears to be implicitly appealing to the following principle:[20]

(F-SAFETY) In case α one's belief that p with degree of confidence c is reliably based in the way required for knowledge only if, in any sufficiently similar case α^* in which one has an at-most-slightly-lower degree of confidence c^* that p, it is true that p.

Moreover, merely in virtue of the description of the scenario, the following holds for the cases α_i under discussion:

(CONF) For every integer i ($0 \leq i < n$), if in α_i one has degree of confidence c that one feels cold, then in α_{i+1} one has an at-most-slightly-lower degree of confidence c^* that one feels cold.

---

[19] Williamson has a very peculiar notion of degrees of confidence: although one's degree of confidence in a given proposition is that which, when there is enough of it, constitutes outright belief in that proposition, Williamson insists that degrees of confidence "should not be equated with subjective probabilities as measured by one's betting behavior" (p. 98). A better indicator of one's degree of confidence in a given proposition, he claims, is the degree to which one is willing to use that proposition as a premise in practical reasoning (p. 99). Nothing I say in what follows turns on the distinction between Williamsonian degrees of confidence/belief and degrees of confidence/belief more traditionally construed.

[20] More precisely, Williamson appears to be appealing to a principle according to which one's belief that p with a certain degree of confidence constitutes knowledge in a given case only if, in any sufficiently similar case in which one believes that p with a sufficiently similar degree of confidence *and on a sufficiently similar basis*, it is true that p. (Compare the discussion of (C-SAFETY′) at the end of §4.) However, for ease of exposition I shall ignore this complication in what follows, since nothing in my criticism of Williamson's argument depends on issues concerning the basis of one's belief.

(F-SAFETY) and (CONF), together with our usual assumption that each case α_i is sufficiently similar to its subsequent case α_{i+1}, imply the desired margin-for-error principle, (MAR):

(MAR) For every integer i ($0 \leq i < n$), if in α_i one knows that one feels cold, then in α_{i+1} one feels cold.

Unlike (BEL) or (KNO), (CONF) seems indisputable, given the description of the situation at hand. However, why should we believe (F-SAFETY)?

The crucial step in Williamson's justification of (F-SAFETY) is his insistence that if in case α one has a degree of confidence that p just barely enough to constitute full-fledged belief, then that belief is not safe/reliable enough to constitute knowledge whenever it is false that p in some sufficiently similar case α^* in which one has a slightly less degree of confidence that p, *even if one's degree of confidence that p in* α^* *is not enough to constitute full-fledged belief.* It is this feature that allows (F-SAFETY) to be a bridge principle where (C-SAFETY) could not be—that allows us to continue to conclude that one feels cold in successive cases α_i even after one's belief that one feels cold has given out. But why should we withhold the honorific "reliable" in the kinds of cases Williamson describes? What if one's degree of confidence in its being the case that p perfectly tracks the underlying basis for its being the case that p, so that one's degree of confidence that p falls just short of belief at the precise point at which things fall just short of making it the case that p? Why would *that* be a situation in which one's initial belief that p is not reliable enough to constitute knowledge? Or slightly more realistically—since Williamson seems to be making the contentious assumption that there is a precise cut-off point above which one's degree of confidence always constitutes full-fledged belief and below which one's degree of confidence always does not constitute full-fledged belief—what if one's belief that p tapers off (as it were) just as its being the case that p tapers off, and in precisely the same way? In such a situation, why should one's lower degree of confidence that p when it is not the case that p in any way impugn the reliability of one's slightly higher degree of confidence that p when it *is* the case that p? Of course, how one cashes out this "tapering off" metaphor will depend upon one's theory of vagueness, but the main point remains: (F-SAFETY) deems as unreliable belief-forming mechanisms that appear to be as reliable as they could possibly be.

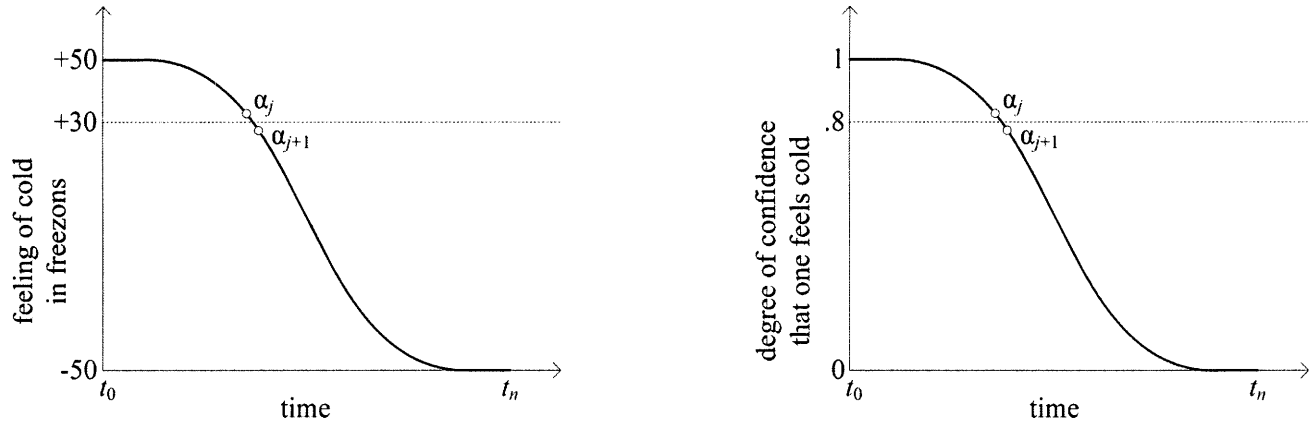

figure 1

To illustrate this point, we can use a slightly altered version of Williamson's own example of a subject gradually feeling warmer on a given morning. The basic idea is that if we are going to go fine-grained with respect to belief, we should also go fine-grained with respect to one's feelings of hot or cold. So let us suppose it were possible to measure the intensity of one's subjective feeling of cold using some set of units—call these units "freezons." To fix on some numbers, let us say that, on the given morning, one's feeling of hot or cold is at a level of 50 freezons at time t_0 (dawn) and a level of −50 freezons at time t_n (noon). Let us also suppose that the following is true of our subject: at any time one's degree of confidence that one feels cold on that morning directly correlates to one's subjective feeling of cold as measured in freezons (see figure 1). If we wanted to be precise, we could encapsulate this correlation with the following equation: if $c(t_i)$ is one's degree of confidence at time t_i that one feels cold (measured on a scale from 0 to 1), and if $f(t_i)$ is one's feelings of hot or cold at t_i as measured in freezons, then $f(t_i) = 100\ c(t_i) - 50$ freezons.[21] Maybe human subjects who are carefully considering how cold they feel would have this sort of a correlation in their freezon/degree-of-confidence levels; maybe they wouldn't. For the purposes of my example it doesn't matter—all I need is for it to be *possible* that there could exist a being with such a correlation between its freezon and degrees-of-confidence levels, which surely is the case.

Finally, let us make one last supposition about the subject of our scenario. Let our subject's

[21] This assumes that one's degree of confidence that one feels cold is 1 at dawn and 0 at noon, but we could easily adjust the equation so that one's degree of confidence that one feels cold is *nearly* 1 at dawn and *nearly* 0 at noon.

confidence in how cold she feels that morning be so well-attuned that the following *penumbral connection*[22] obtains between the vague expressions "believes" and "feels cold" whenever "one" refers to our subject on that morning: one believes that one feels cold if and only if one feels cold. Then if we sharpen "believes" and "feels cold" such that there is a precise cut-off point between the cases in which one does and does not feel cold and a precise cut-off point between the cases in which one does and does not *believe* that one feels cold, the penumbral connection ensures that those cut-off points are the same. Again, it seems evident that there could exist a being for whom this is the case.

But now the crucial point comes: if our subject is such that all of the above is true of her, then according to (F-SAFETY), at some point during the morning her belief that she feels cold is too unreliable to constitute knowledge. However, this just seems wrong: our subject's beliefs about whether she feels cold appear to be as reliable as they possibly could be. This point is most easily illustrated if we sharpen the terms "believes" and "feels cold" so that they do, in fact, have sharp cut-off points.[23] Suppose that the sharp cut-off for belief is at 0.8 degrees of confidence: whenever one has a degree of confidence that p greater than 0.8, one counts as believing that p, and whenever one has a degree of confidence that p less than or equal to 0.8, one counts as not believing that p. It follows from our penumbral connection between "believes" and "feels cold" that the sharp cut-off for feeling cold is 30 freezons. Now let α_j be the last case during the course of the morning such that, on our sharpening of "believes," one counts as believing that one feels cold. Then in α_j, one's degree of confidence that one feels cold is $0.8 + \delta$ (for some small real number $\delta > 0$), and one's feeling of cold is at a level of $30 + \varepsilon$ freezons (for some small real number $\varepsilon > 0$). In case α_{j+1} one millisecond later, one's degree of confidence that one feels cold is $0.8 - \delta'$ (for some small real number $\delta' \geq 0$), and one's feeling of cold is at a level of $30 - \varepsilon'$ freezons (for some small real number $\varepsilon' \geq 0$). So (assuming that α_j is sufficiently similar to α_{j+1} and that $0.8 + \delta$ and $0.8 - \delta'$

---

[22] See Fine (1975), p. 124.

[23] Does doing so stack the deck against Williamson? No, it does not. First, as Williamson himself points out (p. 103), if the argument for (MAR) is no longer cogent when we sharpen the relevant vague expressions, then that gives us reason to suspect that (MAR) only seems plausible because it exploits in an illicit manner the vagueness of its key terms (just as the fact that a typical sorites premise becomes obviously false when "bald" or "heap" is sharpened reveals that the premise is only plausible because it exploits in an illicit manner the vagueness of "bald" or "heap"). And second, on Williamson's own epistemicist theory of vagueness each of the relevant terms already has a sharp cut-off point, so there is no need to sharpen.

are sufficiently similar degrees of confidence), Williamson's fine-grained safety requirement (F-SAFETY) implies that one's degree of confidence that one feels cold in α_j is *too unreliable* to constitute knowledge, since in α_{j+1} one's level of cold as measured in freezons slips just below the threshold of what counts as feeling cold (so that it will be *false* in case α_{j+1} that one feels cold). However, this charge of unreliability seems daft: in α_{j+1} one's level of cold as measured in freezons does indeed slip just below the threshold of what counts as feeling cold, but *at precisely the same point* one's degree of confidence that one feels cold slips just below the threshold of what counts as *believing* that one feels cold. Should we then follow Williamson in saying that, "in picturesque terms," the large portion of one's confidence at t_j that one still has at t_{j+1} is misplaced? I think not.

Now one might have legitimate qualms about the notion of freezons that I have helped myself to in this example. How do we measure these freezons? How do we fix their value both inter-personally, and intra-personally across time? And what entitles us to move, from the familiar fact that some feelings of cold are more intense than others, to the more substantive claim that there is a linear ordering of feelings of cold by their intensity, let alone a linear ordering that has the same structure as an interval on the real line? However, we should note that people can—and have—raised these exact same worries about *degrees of belief*. Freezons are problematic; but so too, in my opinion, are degrees of belief/confidence.

A better objection would be to insist that, once we have sharpened "believes" and "feels cold," the ideal epistemic state for our subject is not one in which (as in figure 1) one's degree of confidence that one feels cold directly correlates to one's feeling of cold as measured in freezons, but rather one in which (as in figure 2) one is absolutely certain that one feels cold until the first case in which one stops feeling cold, at which point one's degree of confidence suddenly drops to 0 and stays that way for the remainder of the morning. But then, the objection continues, since the degree-of-confidence profile given in figure 1 fails to come close enough to the ideally reliable degree-of-confidence profile given in figure 2, it is perfectly acceptable to insist that if one's degree-of-confidence profile were as in figure 1, one's degree of confidence in case α_j that one feels cold would not, in fact, be reliable enough to constitute knowledge. In this way the counterexample to (F-SAFETY) can be avoided.

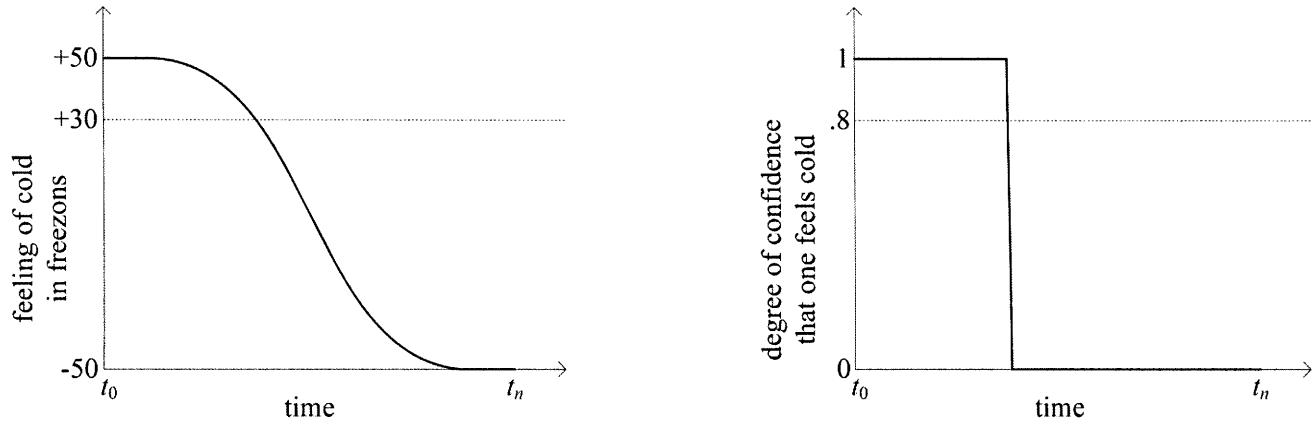

figure 2

It is not clear to me that, even after sharpening the relevant terms, the degree-of-confidence profile given in figure 2 represents the ideal epistemic state that one might have on that morning.[24] However, let us grant to the objector that it does; even then we have not managed to save (F-SAFETY).

The objector concedes that if one's degree-of-confidence profile during the course of the morning were as in figure 2, then in every case in which one's degree of confidence that one feels cold is above the threshold for believing that one feels cold, that degree of confidence would be reliable enough to constitute knowledge. The reason that this possibility is not a problem for Williamson's argument is that, given (CONF), we know that human beings could not have such a degree-of-confidence profile, for that would involve having drastically different degrees of confidence in the same proposition in two successive cases. However, this leaves open the possibility that *other* conceivable beings could have degree-of-confidence profiles that, while not as perfectly accurate as that in figure 2, still are reliable enough to present a problem for (F-SAFETY). For surely it is not the case that *any deviation whatsoever* from the degree-of-confidence profile given in figure 2 results in there being at least one case on the given morning in which one believes that one feels with a certain degree of confidence, but that degree of confidence is not reliably enough based to constitute knowledge. To insist upon *that* would be to insist not just that *reliability* is required for knowledge, but moreover that *perfect reliability* is required, and that way skepticism lies. So

[24] Whether this is so depends on issues outside the scope of this chapter, such as whether truth is both the aim of *belief* and the aim of *degree of belief*, or whether instead there is something else (degree of truth? objective chance?) that stands to degree of belief/confidence as truth stands to belief.

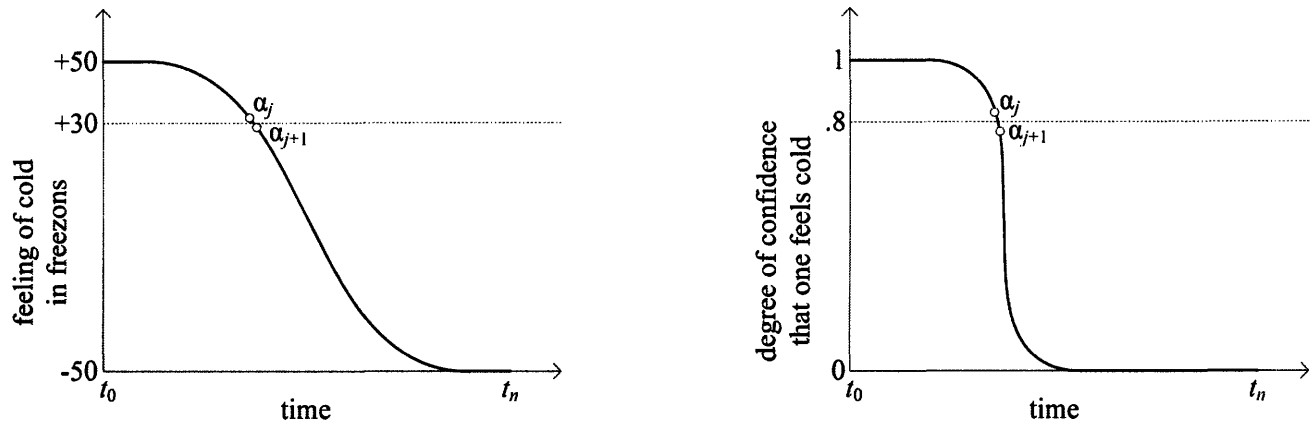

figure 3

there must be *some* degree-of-confidence profiles differing from that of figure 2 such that, for every case in which one has sufficient degree of confidence that one feels cold to count as believing that one feels cold, that degree of confidence is sufficiently reliable to qualify as knowledge.

Another salient feature of the degree-of-confidence profile in figure 2, other than the fact that it depicts a perfectly reliable belief-forming mechanism, is the fact that the profile is *discontinuous*: at the point at which one goes from feeling cold to not feeling cold, there is a sudden, discontinuous jump from one's having degree of confidence 1 that one feels cold to one's having degree of confidence 0 that one feels cold. Non-idealized physical systems rarely—if ever—exhibit discontinuous phenomena at the macroscopic level, so it seems plausible that, among the degree-of-confidence profiles for which one's belief that one feels cold always counts as being suitably reliable, some of those degree-of-confidence profiles are continuous. One likely candidate for such a profile is that given in figure 3: in it, one's degree of confidence over time is nearly as in figure 2, but the corners of the profile are "rounded off" so as to make the degree-of-confidence function $c(t)$ continuous for all t such that $t_0 < t < t_n$. Note that at this point no assumption is being made that a *normal human* could have such a degree-of-confidence profile; all that is being assumed, at least for now, is that *some* possible creature could.

However, now we have trouble, for the degree-of-confidence profile given in figure 3 is a counterexample to (F-SAFETY). As before, let α_j be the last of our cases in which one counts as believing that one feels cold, and let α_{j+1} be the first case in which one does *not* count as believing that one feels cold.

Assuming as before a penumbral connection between "feels cold" and "believes," α_j will also be the last case in which one feels cold, and α_{j+1} the first case in which one does not feel cold. Moreover, because the degree-of-confidence profile is continuous, we can always choose our α_j and α_{j+1} such that one's degree of confidence that one feels cold in α_{j+1} is slightly lower than one's degree of confidence that one feels cold in α_j; if this isn't the case when the interval between cases is one millisecond, we can make it so by choosing a smaller interval, such as one microsecond or one nanosecond. Then in α_j one believes that one feels cold with a given degree of confidence, and in sufficiently similar case α_{j+1} one has a slightly lower degree of confidence that one feels cold, despite its no longer being true that one feels cold. So according to (F-SAFETY), one's belief in α_j that one feels cold is not reliable enough to constitute knowledge. However, by assumption the degree-of-confidence profile in figure 3 is one such that one's belief that one feels cold always meets, for as long as it lasts, the minimum degree of reliability required for knowledge. So, lest we hold that only discontinuous, perfectly reliable degree-of-confidence profiles such as that found in figure 2 allow one's belief that one feels cold to be reliably based in the way required for knowledge for the entire time it lasts, we must give up (F-SAFETY).

Could one resist this conclusion by insisting that, in order for one's belief that one feels cold to be suitably reliably based for the entire morning, one's degree of confidence that one feels cold must indeed take a discontinuous drop once one stops feeling cold? Such a response is available, but it would divorce reliability talk in the case of knowledge from reliability talk in other domains. A common way of motivating reliability constraints on knowledge is first to note our practice of rating devices (such as thermometers) as reliable when they generally serve their purpose, and then to extend that idea to humans by thinking of our belief-forming mechanisms as nothing more than complex, biological devices for attaining true beliefs and avoiding false ones.[25] However, by the criterion of reliability being proposed, almost no physical device would count as always being reliable enough for our purposes. For example, if a certain light is built to turn red whenever the temperature in some room is above 32°F, then in order for

[25] Cf. Williamson (2000), p. 101: "The use of the concept *is reliable* here is a way of drawing attention to an aspect of the case relevant to the application of the concept *knows*, just as one might use the concept *is reliable* in arguing that a machine ill serves its purpose."

that light to qualify as sufficiently reliable for the entire time that it is in use, the light would have to discontinuously change in color at the precise instant at which the temperature in the room rises above 32°F; as no physically implemented light could ever do that, no real-world version of such a light would ever count as reliable in this way. But that seems far too demanding a standard of reliability: we can easily imagine a light built for this purpose that we would rely on, and be right to do so, without its color ever making such a discontinuous jump; indeed, we could even build such a light if we wanted.[26] And if a requirement of discontinuity is too demanding when ascribing the label "reliable" to physical devices, why should it fail to be too demanding when ascribing that label to human cognitive systems and their outputs?

Thus I take the various versions of the freezons example to show that (F-SAFETY) does not specify a genuine necessary condition on knowledge. One reply on Williamson's behalf would be to concede the point, but then try to find some alternative to (F-SAFETY) that could still do the work necessary in justifying (MAR). The basic idea would be to argue that, even if figures 1 or 3 depict the freezon and degree-of-confidence profiles of a creature for whom the condition *that one feels cold* is luminous, humans can never have such profiles, even when they do everything they can to decide whether they feel cold, so the condition in question is not luminous for creatures like us. However, I have grave doubts about the possibility of such a strategy ever succeeding. First, notice that it depends upon our being able to find a new criterion for sorting the freezon/degree-of-confidence profiles that result in one's belief that one feels cold being reliable enough for knowledge during its entire duration from those that do not. How to decide upon this new criterion is a delicate affair: it's just not clear what sort of constraints are in play that would allow us to sort the profiles in a suitably precise, and yet non-arbitrary, manner. But more importantly, once we have this new criterion in hand, we would then need an additional argument establishing that humans always fall on the unreliability side of the divide—that their degree-of-confidence curves over time are *of necessity* too far from the sorts of curves given in figures 1 or 3 for them to have a belief that they

[26] Williamson might insist that, at the precise instant at which the room's temperature is (say) 31.99°F, our light does not qualify as reliable, though it might well have qualified as reliable several seconds earlier when the room's temperature was (say) 31.5°F. However, why say that? Why not say instead that the light was reliable enough for our purposes the entire time, though of course it could have been a bit more reliable when the temperature was 31.99°F? To repeat a point made earlier: Williamson seems to be conflating *sufficient reliability* at a given time with *perfect reliability* at that time.

feel cold that is reliable for the entire time it lasts.

It is here that we run into a familiar problem. We saw in §4 that no version of Williamson's anti-luminosity argument that attempts to derive (MAR) from a coarse-grained version of the safety requirement would succeed if there exists a certain kind of constitutive connection between the coarse-grained states of feeling cold and believing that one feels cold. However, there might also exist an analogous constitutive connection *at the fine-grained level* between one's feeling of cold as measured in freezons and one's degree of confidence that one feels cold. And moreover, this constitutive connection might be such that, whenever one has done everything one is in a position to do to decide whether one feels cold, the graphs over time of one's feeling of cold measured in freezons and one's degree of confidence that one feels cold would be as in figures 1 or 3, or at least close enough to those graphs for one's belief that one feels cold to count as reliable enough for knowledge during its entire duration. Of course, one might doubt that such a constitutive connection exists. But to simply *assume* that it does not, without offering any arguments in support of that assumption, would once again beg the main question at issue, since defenders of luminosity are typically motivated by the thought that there is a tight connection between the obtaining of certain conditions and our beliefs, at least upon reflection, about the obtaining of those conditions. So even if a suitable replacement for (F-SAFETY) could be found—which itself is highly doubtful—then the brunt of the argumentative work in establishing that conditions such as *that one feels cold* are not luminous would still be left to be done.

6. Do the Relevant Constitutive Connections Obtain?

I have presented two ways of attempting to derive the crucial premise in Williamson's anti-luminosity argument from a safety requirement on knowledge: the first attempted to derive that premise from a coarse-grained safety requirement formulated in terms of "all or nothing" belief, the second attempted to derive it from a fine-grained safety requirement formulated in terms of degrees of confidence. In each case, the original version of the argument ran into problems—in the coarse-grained case, it proved impossible to derive the needed premise (MAR) from the coarse-grained safety requirement without

appealing to a dubious sorites-premise-like bridge principle between successive cases, and in the fine-grained case, the proffered fine-grained safety requirement failed to specify a genuine necessary condition for knowledge. Moreover, in each case the best attempt at resuscitating a version of the argument—whether in the coarse-grained case by only running the argument until the first case in which it is false that one feels cold, or in the fine-grained case by proposing an alternative to the original (putative) fine-grained safety requirement—was blocked by the possibility that there might exist a certain sort of constitutive connection between feeling cold and believing that one feels cold (in the coarse-grained case), or between the degree to which one feels cold and the degree to which one believes that one feels cold (in the fine-grained case).

Settling to what extent, if any, there is a constitutive connection—whether at the coarse-grained or fined-grained level—between feeling cold and believing that one feels cold is beyond the scope of the current chapter. But this much seems evident to me: there must be *some* sort of modal connection, constitutive or otherwise, between one's phenomenal experiences (such as one's feeling cold) and one's cognitive states (such as one's believing that one feels cold)—the phenomenal and the cognitive cannot swing apart from each other any which way one likes. This is brought out by considering a thought experiment once proposed by Ernest Sosa in a very different context. Sosa has us imagine a subject who "has a beautifully coherent and comprehensive set of beliefs," yet through the interference of a group of Cartesian evil demons, the subject's sensory experiences are "wildly at odds with his beliefs." So, for example, the subject "believes he has a splitting headache, but he has no headache at all; he believes he has a cubical piece of black coal before him, while his visual experience is as if he had a white and round snowball before him," and so on.[27] Sosa takes this case to be a counterexample to any theory of justification that makes reference only to one's beliefs and the relations between them, since presumably that theory would deem the subject's beliefs to be fully justified, whereas Sosa insists that there still seems to be something epistemically blameworthy about our subject. However, my reaction to this example is quite different: I fail to see why Sosa's case should trouble anyone, since it seems clear to me that the

[27] Sosa (1991), pp. 135-136.

scenario as described is *impossible*—that there *could not* exist a being who counts as having experiences and beliefs, yet those experiences and beliefs are radically disjoint from one another in the way Sosa imagines. Moreover, once one tries to explain why such a being could not exist, one soon finds oneself using phrases like "constitutive connection." Of course, it is one thing to say that a subject cannot have a fully coherent set of beliefs that are wildly at odds with most of her experiences, and quite another to say that, after ideal reflection, a subject cannot believe that she feels cold without really feeling cold. But I take Sosa's example to strongly motivate the idea that there must be some sort of modal connection between the phenomenal and the cognitive realms. Whether that connection is a constitutive one, and whether it is a tight enough connection to block Williamson's argument, is a topic for another day.

Note, however, that even if Williamson were somehow able to prove that the relevant constitutive connection does not hold between the obtaining of the condition *that one feels cold* and one's believing that the condition obtains, in order to extend his anti-luminosity argument to other conditions he would need to argue, on a case by case basis, that an analogous constitutive connection does not exist for each condition to which he applies the argument. For example, in order to use a coarse-grained version of the anti-luminosity argument to show that the condition *it is rational for one to believe that p* is not luminous, Williamson would have to establish that the following is not the case:

> (R-CON) If one has done everything one can to decide whether it is rational for one to believe that *p*, then one believes that it is rational for one to believe that *p* only if it is, in fact, rational for one to believe that *p*.

But for those who have internalist persuasions, the idea that in the limit of inquiry one cannot be wrong about what it is rational for one to believe is extremely plausible. So it will be dialectically difficult, if not impossible, to use a coarse-grained version of the anti-luminosity argument to motivate an externalist conception of rationality according to which it can be rational for one to believe that *p* despite one's not being in a position to know that it is rational for one to so believe. And similar comments apply to attempts to use the anti-luminosity argument (whether in its coarse-grained or fine-grained form) to show that various other conditions, such as *that it appears to one that q* or *that words X and Y have the same meaning for one*, are not luminous: the constitutive connection that Williamson must deny in order to make his

argument work is often precisely what is at stake in claiming that the given condition is luminous.

Few would doubt that a version of Williamson's argument can establish that an external-world condition such as *that the temperature outside is less than 90 °F* is not luminous. We can imagine a morning on which the outside temperature starts at 70°F and then slowly warms up to 110°F, all while a given subject does nothing but carefully attend to whether the temperature outside feels to be less than 90°F. In this case, the analogue of (MAR) obviously holds: if at any time on the given morning one knows that the temperature outside is less than 90°F, then one millisecond later the temperature must still be less than 90°F. Even if somehow one stops believing that the temperature outside is less than 90°F at precisely the point at which the temperature first reaches 90°F, or even if one's degree of confidence in that proposition suddenly takes a sharp drop when the proposition first becomes false, such an occurrence would seem to be nothing more than a fortuitous accident: given the inexactness of our abilities to detect the external temperature around us, there must be some sufficiently similar situations in which one does not stop believing (or in which one's degree of confidence does not significantly drop) at that first point at which the outside temperature stops being less than 90°F. It is only a short step from there to concluding that, regardless of what one's coarse- or fine-grained beliefs are like in the actual world, on the given morning one does not know that the temperature outside is less than 90°F if one millisecond later it is not.

However, for the defender of luminosity, it is not a fortuitous accident that, when one carefully considers the matter, one's beliefs about whether one feels cold perfectly line up with the facts about whether one feels cold. That this could happen would seem miraculous if one holds an *inner perception model of self-knowledge* according to which introspection is fundamentally no different from perception via the five senses: if our faculty for forming beliefs about our own mental states were essentially just an eye turned inward, then given the inevitable inexactness of our external perceptual faculties, it would seem that our inner belief-forming mechanisms must be irremediably inexact as well. But few, if any, of Williamson's targets hold such a conception of self-knowledge.[28] Thus for most defenders of luminosity claims there is an important disanalogy between Williamson's argument as applied to the condition *that the temperature*

[28] For a survey of the various views of self-knowledge that do not involve an inner perception model, see Gertler (2003).

outside is less than 90 °F and his argument as applied to a condition such as *that one feels cold*: in the latter case it is *one's own mind* that one is forming beliefs about, and one does not do this by standing outside of oneself, as it were, and using a quasi-perceptual faculty to detect the goings-on within.[29]

One of the upshots of Williamson's anti-luminosity argument is supposed to be that, because we have no cognitive home of mental states whose obtaining we are always in a position to know of, the type of epistemic access we have to our own mental states is qualitatively no different from the type of epistemic access we have to external-world conditions—the difference is a matter of degree, not of kind. But if my arguments are correct, then there is a sense in which for Williamson's anti-luminosity argument to have any chance of succeeding, he must already assume that conclusion, since he must assume that there is not a special, tight connection between, say, our feelings of cold and our beliefs about those feelings, or between the degree to which we feel cold and the degree to which we believe that we feel cold. If Williamson's anti-luminosity argument succeeds, then it is possible for us to be *epistemically disengaged* (that is, disengaged at the level of knowledge) from our minds, the meanings of our words, and what rationality demands of us; but he can only establish that conclusion by assuming that, even after ideal reflection, we can be *doxastically disengaged* (that is, disengaged at the level of belief) from our minds, the meanings of our words, and the demands of rationality.

7. Coda: The Lustrous and the Luminous

I would like to end with a few brief comments about the scope of Williamson's argument.

I take the arguments in §§4-6 to show that Williamson does not successfully demonstrate that every non-trivial condition is non-luminous. However, even if I am wrong and Williamson's argument does succeed in showing, say, that one can be in pain without being in a position to know that one is pain, it would not follow that we could use an analogous argument to show that one can be in pain without being in a position to *justifiably believe* that one is in pain. Continuing with Williamson's light-giving

[29] If Williamson's epistemicist theory of vagueness is false, there will be a second important disanalogy between Williamson's argument as applied to the condition *that the temperature outside is less than 90 °F* and his argument as applied to a condition such as *that one feels cold*, since the former but not the latter condition will have a sharp cut-off point between the cases in which it obtains and those in which it does not.

metaphor, let us say that a condition C is *lustrous* if and only if the following holds:

(**) For every case α, if in α condition C obtains, then in α one is in a position to justifiably believe that C obtains.

A condition that is lustrous shines of its own accord, though not necessarily as brightly as one that is luminous, for the two notions can pull apart: one's position could be such that, if one were to believe that the condition in question obtains, then one's belief would be justified, without the justification that one would thereby have being sufficient to make one's true belief constitute knowledge. And though Williamson's argument might (perhaps) show that the condition *that one feels cold* is not luminous, a parallel argument cannot be used to show that the condition *that one feels cold* is not lustrous, for while it might (perhaps) be true that knowledge requires safety from error, *it is completely implausible to suppose that justified belief requires safety from error.*

Or rather, it is completely implausible to suppose that justified belief requires safety from error, given the common assumption that it is possible to justifiably believe a falsehood. According to the safety requirement on knowledge, one knows that *p* in a given case α only if in some set of sufficiently similar cases, it is true that *p*. What determines that set will vary depending on which version of the safety constraint one is working with: perhaps it is all sufficiently similar cases in which one believes that *p*, or all sufficiently similar cases in which one believes that *p* on a sufficiently similar basis as in α, or all sufficiently similar cases in which one's degree of confidence that *p* is at most slightly less than in α. (See §§4-5 above.) But regardless of how one determines the set of sufficiently similar cases, α will itself be among that set, for α itself is as similar as a case can be to α, and it is trivially true that in α one believes that *p* on the same basis and with the same degree of confidence as one believes that *p* in α. For this reason, the safety requirement for knowledge implies that knowledge is factive: one knows that *p* in some case α only if it is true that *p* in α. And a safety requirement for justified belief would have exactly the same implication: according to such a constraint, one is justified in believing that *p* in a case α only if in some set of sufficiently similar cases, it is true that *p*; as α will be among that set, it will follow that justified belief is factive. However, we can be justified in believing falsehoods: for instance, I might be justified in believing

that my copy of *Knowledge and Its Limits* is sitting on my desk at home even if this belief is false because, unbeknownst to me, someone broke into my apartment and stole the book. So there can be no safety constraint for justified belief.[30]

Thus if we try to show that the condition *that one feels cold* is not lustrous by running an argument parallel to the one that Williamson uses to argue that that condition is not luminous, we will be unable to appeal to a safety requirement for justified belief in order to secure the analogue of premise (MAR), which would read as follows:

(J-MAR) For every integer i ($0 \leq i < n$), if in α_i one is justified in believing that one feels cold, then in α_{i+1} one feels cold.

(J-MAR) simply does not have the same plausibility as (MAR). It seems perfectly plausible that one could *not* feel cold in α_{i+1} and yet have been justified in believing one millisecond earlier that one felt cold; indeed, if α_{i+1} is one of the first few cases on the morning in question in which it is not the case that one feels cold, it seems perfectly plausible that one will *still* be justified in believing that one feels cold in α_{i+1} itself and in any number of cases a few milliseconds later.[31]

Therefore a parallel version of the anti-luminosity argument cannot be used to argue that the condition *that one feels cold* is not lustrous—that is, to argue against the claim that whenever one feels cold, one is in a position to justifiably believe that one feels cold. Williamson, however, is likely to be unbothered by this result: after all, one of the chief slogans of his book is "knowledge first" (p. v). For Williamson it is knowledge that is fundamental in all epistemic matters, and interest in justification is only derivative or secondary. However, it is tempting to see Williamson's anti-luminosity argument as some kind of devastating attack on foundationalism, and it is important to realize that this is simply not the case: most contemporary foundationalists are foundationalists *with respect to justification*, and Williamson's argument leaves justificatory foundationalism completely untouched. For instance, according to James Pryor's version of modest foundationalism, "whenever you have an experience as of *p*'s being the case,

[30] Note that false beliefs can be justified even on Williamson's view according to which (i) evidence is what justifies belief, and (ii) one's evidence is all and only what one knows ("E = K"): even though on this view one's evidence only consists of true propositions, in some circumstances that evidence might be misleading and thus probabilify falsehoods.

[31] For this reason, appealing to an alleged safety requirement for justified *true* belief will not help the anti-lustrousness argument, either.

you thereby have immediate (*prima facie*) justification for believing *p*."[32] And we could easily imagine someone defending the related foundationalist view that whenever one has an experience as of *p*'s being the case, one thereby has immediate *prima facie* justification for believing *that one has an experience as of p's being the case*. Williamson's argument poses no threat to such claims. When Williamson insists that we are cognitively homeless, what he means is that we are cognitively homeless *with respect to our knowledge of what conditions obtain*: even if his argument works (and it doesn't: see §§4-6), we might still have a cognitive home with respect to matters of justification. I leave it open how satisfactory a cognitive home that would be.

[32] Pryor (2000), p. 532.

References

Adler, Jonathan, and Michael Levin (2002). "Is the Generality Problem Too General?" *Philosophy and Phenomenological Research* 65, pp. 87-97.

Alston, William P. (1995). "How to Think about Reliability." *Philosophical Topics* 23, pp. 1-29.

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

Baier, Kurt (1958). *The Moral Point of View: A Rational Basis of Ethics*. Ithaca, NY: Cornell University Press.

Bales, R. Eugene (1971). "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?" *American Philosophical Quarterly* 8, pp. 257-265.

Balog, Katalin (forthcoming). "The 'Quotational Account' of Phenomenal Concepts." Unpublished manuscript.

Bittner, Rüdiger (1974). "Maximen." In Gerhard Funke (ed.), *Akten des 4. Internationalen Kant-Kongress* II.2 (Berlin: Walter de Gruyter), pp. 485-498.

Block, Ned (2006). "Max Black's Objection to Mind-Body Identity." In Dean Zimmerman (ed.), *Oxford Studies in Metaphysics, Vol. 2* (Oxford: Oxford University Press).

Broome, John (2004). "Reasons." In Wallace, et al. (2004), pp. 28-55.

Brueckner, Anthony, and M. Oreste Fiocco (2002). "Williamson's Anti-Luminosity Argument." *Philosophical Studies* 110, pp. 285-293.

Casullo, Albert (2003). *A Priori Justification*. Oxford: Oxford University Press.

Chalmers, David (2003). "The Content and Epistemology of Phenomenal Belief." In Quentin Smith and Aleksandar Jokic (eds.), *Consciousness: New Philosophical Perspectives* (Oxford: Oxford University Press).

Chang, Ruth (2004). "Can Desires Provide Reasons for Action?" In Wallace, et al. (2004), pp. 56-90.

Chisholm, Roderick M. (1956). "'Appear,' 'Take,' and 'Evident.'" *Journal of Philosophy* 53, pp. 722-731.

Chisholm, Roderick M. (1964). "The Ethics of Requirement." *American Philosophical Quarterly* 1, pp. 147-153.

Chisholm, Roderick M. (1977). *Theory of Knowledge*, 2nd edition. Englewood Cliffs, NJ: Prentice-Hall.

Chisholm, Roderick M. (1982). "A Version of Foundationalism." In *The Foundations of Knowing* (Brighton: Harvester Press), pp. 3-32.

Chisholm, Roderick M. (1997). "My Philosophical Development." In Lewis Edwin Hahn (ed.), *The Philosophy of Roderick M. Chisholm* (Chicago: Open Court), pp. 3-41.

Comesaña, Juan (2005). "Unsafe Knowledge." *Synthese* 146, pp. 395-404.

Comesaña, Juan (2006). "A Well-Founded Solution to the Generality Problem." *Philosophical Studies* 129, pp. 27-47.

Conee, Earl (2005). "The Comforts of Home." *Philosophy and Phenomenological Research* 70, pp. 444-451.

Conee, Earl, and Richard Feldman (1998). "The Generality Problem for Reliabilism." *Philosophical Studies* 89, pp. 1-29.

Conee, Earl, and Richard Feldman (2004). *Evidentialism*. Oxford: Oxford University Press.

Crisp, Roger (2000). "Particularizing Particularism." In Hooker and Little (2000), pp. 23-47.

Cullity, Garrett (2002). "Particularism and Moral Theory I: Particularism and Presumptive Reasons." *Proceedings of the Aristotelian Society Supplement* 76, pp. 169-190.

Dancy, Jonathan (1981). "On Moral Properties." *Mind* 90, pp. 367-385.

Dancy, Jonathan (1983). "Ethical Particularism and Morally Relevant Properties." *Mind* 92, pp. 530-547.

Dancy, Jonathan (1993). *Moral Reasons*. Oxford: Blackwell.

Dancy, Jonathan (2000a). "The Particularist's Progress." In Hooker and Little (2000), pp. 130-156.

Dancy, Jonathan (2000b). *Practical Reality*. Oxford: Oxford University Press.

Dancy, Jonathan (2001). "Moral Particularism." *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/moral-particularism/>.

Dancy, Jonathan (2004). *Ethics Without Principles*. Oxford: Clarendon Press.

Feldman, Richard (1985). "Reliability and Justification." *The Monist* 68, pp. 159-174.

Feldman, Richard (2003). *Epistemology*. Upper Saddle River, NJ: Prentice Hall.

Feldman, Richard, and Earl Conee (2002). "Typing Problems." *Philosophy and Phenomenological Research* 65, pp. 98-105.

Fine, Kit (1975). "Vagueness, Truth, and Logic." *Synthese* 30, pp. 265-300. Page references are to the reprint in Rosanna Keefe and Peter Smith (eds.), *Vagueness: A Reader* (Cambridge, MA: MIT Press, 1997), pp. 119-151.

Gertler, Brie (2003). "Self-Knowledge." *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/self-knowledge/>.

Godfrey-Smith, William (1978). "The Generality of Predictions." *American Philosophical Quarterly* 15, pp. 15-25.

Goldman, Alvin (1979). "What Is Justified Belief?" In George Pappas (ed.), *Justification and Knowledge* (Dordrecht: Reidel). Page references are to the reprint in Goldman (1992).

Goldman, Alvin (1980). "The Internalist Conception of Justification." *Midwest Studies in Philosophy* 5, pp. 27-51.

Goldman, Alvin (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldman, Alvin (1992). *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: MIT University Press.

Hare, R. M. (1963). *Freedom and Reason*. Oxford: Clarendon Press.

Harman, Gilbert, and Judith Jarvis Thomson (1996). *Moral Relativism and Moral Objectivity*. Oxford: Blackwell Publishers.

Herman, Barbara (1984). "Mutual Aid and Respect for Persons." *Ethics* 94, pp. 577-602. Page references are to reprint in Herman (1993a).

Herman, Barbara (1985). "The Practice of Moral Judgment." *Journal of Philosophy* 82, pp. 414-436. Page references are to the reprint in Herman (1993a).

Herman, Barbara (1989). "Murder and Mayhem." *The Monist* 72, pp. 411-431. Page references are to the reprint in Herman (1993a).

Herman, Barbara (1990). *Morality as Rationality: A Study in Kant's Ethics*. New York: Garland Publishing. Reprint of 1976 doctoral dissertation.

Herman, Barbara (1992). "What Happens to the Consequences?" In Paul Guyer, Ted Cohen, and Hilary Putnam (eds.), *Pursuits of Reason* (Arlington: Texas Tech University Press). Page references are to the reprint in Herman (1993a).

Herman, Barbara (1993a). *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press.

Herman, Barbara (1993b). "Leaving Deontology Behind." In Herman (1993a).

Herman, Barbara (1993c). "Moral Deliberation and the Derivation of Duties." In Herman (1993a).

Höffe, Otfried (1977). "Kants kategorischer Imperativ als Kriterium des Sittlichen." *Zeitschrift für philosophische Forschung* 31, pp. 354-384.

Holton, Richard (2002). "Particularism and Moral Theory II: Principles and Particularisms." *Proceedings of the Aristotelian Society Supplement* 76, pp. 191-209.

Hooker, Brad, and Margaret Olivia Little, eds. (2000). *Moral Particularism*. Oxford: Clarendon Press.

Horty, John F. (2004). "Principles, Reasons, and Monotony." Unpublished manuscript available at <http://www.umiacs.umd.edu/users/horty/articles/2003-ra.pdf>.

Hurka, Thomas (1990). "Two Kinds of Satisficing." *Philosophical Studies* 59, pp. 107-111.

Jackson, Frank; Philip Pettit; and Michael Smith (2000). "Ethical Particularism and Patterns." In Hooker and Little (2000), pp. 79-99.

Jarvis (Thomson), Judith (1962). "Practical Reasoning." *Philosophical Quarterly* 12, pp. 316-328.

Kagan, Shelly (1988). "The Additive Fallacy." *Ethics* 99, pp. 5-31.

Kagan, Shelly (1989). *The Limits of Morality*. Oxford: Oxford University Press.

Kant, Immanuel (1785/1997). *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor. Cambridge: Cambridge University Press.

Kant, Immanuel (1788/1997). *Critique of Practical Reasons*, trans. Mary Gregor. Cambridge: Cambridge University Press.

Kant, Immanuel (1793/1960). *Religion within the Limits of Reason Alone*, trans. Theodore M. Greene and Hoyt H. Hudson. New York: Harper & Row.

Kant, Immanuel (1797/1996). *The Metaphysics of Morals*, trans. Mary Gregor. Cambridge: Cambridge University Press.

Kant, Immanuel (1798/2006). *Anthropology from a Pragmatic Point of View*, trans. Robert B. Louden. Cambridge: Cambridge University Press.

Kihlbom, Ulrik (2002). *Ethical Particularism: An Essay on Moral Reasons*. Stockholm: Almqvist & Wiksell International.

Kitcher, Patricia (2003). "What Is a Maxim?" *Philosophical Topics* 31, pp. 215-243.

Korsgaard, Christine (1985). "Kant's Formula of Universal Law." *Pacific Philosophical Quarterly* 66, pp. 24-47. Page references are to the reprint in Korsgaard (1996).

Korsgaard, Christine (1989a). "Kant's Analysis of Obligation: The Argument of *Foundations* I." *The Monist* 72, pp. 311-340. Page references are to the reprint in Korsgaard (1996).

Korsgaard, Christine (1989b). "Morality as Freedom." In Yirmiyahu Yovel (ed.), *Kant's Practical Philosophy Reconsidered* (Dordrecht: Kluwer). Page references are to the reprint in Korsgaard (1996).

Korsgaard, Christine (1996). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Korsgaard, Christine (1997a). "Introduction" to Kant (1785/1997).

Korsgaard, Christine (1997b). "The Normativity of Instrumental Reason." In Garrett Cullity and Berys Gaut (eds.), *Ethics and Practical Reason* (Oxford: Oxford University Press).

Korsgaard, Christine (forthcoming-a). *Self-Constitution: Action, Identity, and Integrity*. Six lectures delivered as the 2002 Locke Lectures at Oxford. Available at <http://www.people.fas.harvard.edu/~korsgaar/>.

Korsgaard, Christine (forthcoming-b). "Acting for a Reason." To appear in V. Bradley Lewis (ed.), *Studies in Practical Reason* (Catholic University Press).

Lance, Mark, and Margaret Olivia Little (2004). "Defeasibility and the Normative Grasp of Context." *Erkenntnis* 61, pp. 435-455.

Lance, Mark, and Margaret Olivia Little (2006a). "Defending Moral Particularism." In James Dreier (ed.), *Contemporary Debates in Moral Theory* (Oxford: Blackwell Publishers), pp. 305-321.

Lance, Mark, and Margaret Olivia Little (2006b). "Particularism and Antitheory." In David Copp (ed.), *The Oxford Handbook of Ethical Theory* (Oxford: Oxford University Press), pp. 567-594.

Little, Margaret Olivia (2000). "Moral Generalities Revisited." In Hooker and Little (2000), pp. 276-304.

Little, Margaret Olivia (2001a). "On Knowing the 'Why': Particularism and Moral Theory." *Hastings Center Report* 31, pp. 32-40.

Little, Margaret Olivia (2001b). "Wittgensteinian Lessons on Moral Particularism." In Carl Elliott (ed.), *Slow Cures and Bad Philosophers: Essays on Wittgenstein, Medicine, and Bioethics* (Durham, NC: Duke University Press), pp. 161-180.

McDowell, John (1998). *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.

McKeever, Sean, and Michael Ridge (2005a). "The Many Moral Particularisms." *Canadian Journal of Philosophy* 35, pp. 83-106.

McKeever, Sean, and Michael Ridge (2005b). "What Does Holism Have to Do with Moral Particularism?" *Ratio* 18, pp. 93-103.

McNaughton, David (1988). *Moral Vision*. Oxford: Blackwell.

McNaughton, David (1996). "An Unconnected Heap of Duties?" *Philosophical Quarterly* 46, pp. 433-447.

McNaughton, David, and Piers Rawling (2000). "Unprincipled Ethics." In Hooker and Little (2000), pp. 256-275.

Mill, J. S. (1863/1987). *Utilitarianism*, ed. A. Ryan. London: Penguin Books.

Millgram, Elijah (2003). "Does the Categorical Imperative Give Rise to a Contradiction in the Will?" *The Philosophical Review* 112, pp. 525-560.

Nagel, Thomas (1970). *The Possibility of Altruism*. Princeton, NJ: Princeton University Press.

Nell (O'Neill), Onora (1975). *Acting on Principle: An Essay on Kantian Ethics*. New York: Columbia University Press.

Neta, Ram, and Guy Rohrbaugh (2004). "Luminosity and the Safety of Knowledge." *Pacific Philosophical Quarterly* 85, pp. 396-406.

O'Neill, Onora (1984). "Kant after Virtue." *Inquiry* 26, pp. 387-405. Page references are to the reprint in O'Neill (1989a).

O'Neill, Onora (1985). "Consistency in Action." In Nelson T. Potter and Mark Timmons (eds.), *Universality and Morality: Essays on Ethical Universalizability* (Dordrecht: Reidel). Page references are to the reprint in O'Neill (1989a).

O'Neill, Onora (1989a). *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press.

O'Neill, Onora (1989b). "Universal Law and Ends-in-Themselves." *The Monist* 72, pp. 341-361. Page references are to the reprint in O'Neill (1989a).

O'Neill, Onora (2004). "Modern Moral Philosophy and the Problem of Relevant Descriptions." In Anthony O'Hear (ed.), *Modern Moral Philosophy* (Cambridge: Cambridge University Press).

Papineau, David (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.

Parfit, Derek (forthcoming). *Climbing the Mountain*. Unpublished manuscript.

Pollock, John L. (1986). *Contemporary Theories of Knowledge*, 1st edition. Savage, MD: Rowman & Littlefield.

Pryor, James (2000). "The Skeptic and the Dogmatist." *Noûs* 34, pp. 517-549.

Rawls, John (2000). *Lectures on the History of Moral Philosophy*, ed. Barbara Herman. Cambridge, MA: Harvard University Press.

Raz, Joseph (1990). *Practical Reason and Norms*, 2nd edition. Oxford: Oxford University Press.

Raz, Joseph (2000). "The Truth in Particularism." In Hooker and Little (2000), pp. 48-78.

Ross, W. D. (1930). *The Right and the Good*. Oxford: Clarendon Press.

Sainsbury, R. M. (1995). "Vagueness, Ignorance, and Margin for Error." *British Journal of Philosophy* 46, pp. 589-601.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Shope, Robert K. (1983). *The Analysis of Knowing: A Decade of Research*. Princeton, NJ: Princeton University Press.

Slote, Michael (1985). *Common-Sense Morality and Consequentialism*. London: Routledge & Kegan Paul.

Smith, Michael (1994). *The Moral Problem*. Oxford: Blackwell Publishers.

Sosa, Ernest (1985). "Knowledge and Intellectual Virtue." *The Monist* 68, pp. 224-245.

Sosa, Ernest (1991). "Reliabilism and Intellectual Virtue." In *Knowledge in Perspective* (Cambridge: Cambridge University Press), pp. 131-145.

Sosa, Ernest (1996). "Postscript to 'Proper Functionalism and Virtue Epistemology.'" In Jonathan L. Kvanvig (ed.), *Warrant in Contemporary Philosophy: Essays in Honor of Plantinga's Theory of Knowledge* (Lanham, MD: Rowman and Littlefield), pp. 271-280.

Sosa, Ernest (1999a). "How Must Knowledge Be Modally Related to What Is Known?" *Philosophical Topics* 26, pp. 373-384.

Sosa, Ernest (1999b). "How to Defeat Opposition to Moore." In James Tomberlin (ed.), *Philosophical Perspectives* 13 (Atascadero, CA: Ridgeview), pp. 141-153.

Sosa, Ernest (2000). "Skepticism and Contextualism." *Philosophical Issues* 10, pp. 1-18.

Sosa, Ernest (2002). "Tracking, Competence, and Knowledge." In Paul Moser (ed.), *The Oxford Handbook of Epistemology* (Oxford: Oxford University Press), pp. 264-286.

Sosa, Ernest. (2004). "Relevant Alternatives, Contextualism Included." *Philosophical Studies* 199, pp. 35-65.

Urmson, J. O. (1975). "A Defense of Intuitionism." *Proceedings of the Aristotelian Society* 75, pp. 111-119.

Väyrynen, Pekka (2004). "Particularism and Default Reasons." *Ethical Theory and Moral Practice* 7, pp. 53-79.

Wallace, R. Jay; Philip Pettit; Samuel Scheffler; and Michael Smith, eds. (2004). *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. Oxford: Clarendon Press.

Weatherson, Brian (2004). "Luminous Margins." *Australian Journal of Philosophy* 82, pp. 373-383.

Wedgwood, Ralph (2002). "The Aim of Belief." *Philosophical Perspectives* 16, pp. 267-297.

Williamson, Timothy (1990). *Identity and Discrimination*. Oxford: Basil Blackwell.

Williamson, Timothy (1992a). "Inexact Knowledge." *Mind* 101, pp. 217-242.

Williamson, Timothy (1992b). "Vagueness and Ignorance." *Proceedings of the Aristotelian Society Supplement* 66, pp. 145-162.

Williamson, Timothy (1994). *Vagueness*. London: Routledge.

Williamson, Timothy (1996). "Cognitive Homelessness." *Journal of Philosophy* 93, pp. 554-573.

Williamson, Timothy (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.

Williamson, Timothy (2004). "Replies to Commentators." *Philosophical Books* 45, pp. 313-323.

Williamson, Timothy (2005). "Replies to Commentators." *Philosophy and Phenomenological Research* 70, pp. 468-491.

Williamson, Timothy (forthcoming). "Why Epistemology Can't Be Operationalized." To appear in Quentin Smith (ed.), *Epistemology: New Philosophical Essays* (Oxford: Oxford University Press).

Wood, Allen (1999). *Kant's Ethical Thought*. Cambridge: Cambridge University Press.