Subjectivist Theories of Normative Language

H.W.A. Evers

Thesis submitted for the degree of Doctor of Philosophy at the University of Oxford.

Acknowledgements

Several people have been important to me during the writing of this thesis. I would like to thank my supervisor Ralph Wedgwood for guidance and encouragement. His supreme knowledge and insight made this work far better than it would otherwise have been. I am also grateful to Stephen Finlay for extremely generous comments throughout the years (even in very difficult times). I would also like to thank him for allowing me to read and discuss his book manuscript *Confusion of Tongues*. I would like to thank Natalja Deng for stimulating discussion and very generous financial and spiritual support. I would like to thank my parents too, for their wonderful support throughout my life. I would like to thank Leo Gough and Louise Hanson for giving me a place to stay in Oxford and London in the last few weeks before submission. Finally, I would like to thank the following institutions for scholarships and bursaries: the Royal Institute of Philosophy, the Arts & Humanities Research Council, the Prins Bernhard Cultuurfonds, the Stichting Fundatie van de Vrijvrouwe van Renswoude te 's-Gravenhage, the Deutscher Akademischer Austausch Dienst (for the Michael Foster Memorial Scholarship) and Jesus College, Oxford.

Abstract: Subjectivist theories of normative language

On the assumption that there are no objective normative facts, what is the best theory of normative language? I try to answer this question.

Chapter 1 argues for a presumption against noncognitivism and explains why error-theories are of limited interest: they concern adverbs and adjectives like 'moral', but not words like 'ought', 'good' and 'reason'. This narrows down the options: the best subjectivist theory of normative language is a truth conditional, non-error-theoretic account.

Chapter 2 argues for contextualism about normative statements. Contextualists hold that their truth conditions (can) vary with the context of utterance.

Chapter 3 starts the assessment of contextualist theories. It looks into Humean accounts. Problems are revealed with both Harman's and Schroeder's versions.

Chapter 4 develops a form of indexical relativism according to which the truth of normative statements depends on contextually salient rules. I present imperative-based analyses of 'ought' and 'reason' and show how they can explain why '*A* ought to *X*' entails that the balance of reasons favours that *A X*-es.

Chapter 5 further develops the theory of chapter 4 and applies it to the words 'good' and 'must'. It turns out to be hard to analyse 'good'. It also emerges that 'must' and 'ought' cannot be given different truth conditions.

Chapter 6 explains Stephen Finlay's end-relational theory. On this account, normative statements concern the relation in which acts or objects stand to contextually salient ends. In the case of 'ought' and 'good', this relation is one of probability raising.

Chapter 7 discusses and answers some familiar objections to Finlay's view.

Chapter 8 raises some new problems, related to the fact that normative judgments are often made in the light of *several* ends.

Chapter 9 explains why the end-relational theory is nonetheless the best subjectivist theory of normative language.

Table of contents

Chapter 1. Normative language, noncognitivism and error-theories

1.1 Introduction

Normative language is used to express evaluations of actions, states of affairs and objects. I mean 'action' to be understood broadly so as to include "mental" action like changing one's beliefs. 'To express' is intended to cover all possible forms of linguistic conveying. The phrase, then, is neutral with respect to whether normative language is fact-stating or not. 'Evaluation' is intended to cover different phenomena, including assessments of requirements, reasons, and judgments which place things on an evaluative scale. The following sentences are examples of normative language:

(1) You ought to keep your promises.

(2) Torture is morally wrong.

(3) That is a good car.

(4) We have reason to believe that life evolved.

(5) If you want to win, then you have to pay attention.

It is hotly disputed how to understand such sentences. Some say they (purport to) describe objective normative facts. Others say they don't describe any kind of fact at all. Yet others say they do describe facts, but not objective ones.

This thesis is about normative language. Its leading question is this: *On the assumption that there are no objective normative facts, what is the best theory of normative*

language? It argues for a view which falls within the third category: i.e. normative language is fact-stating, but not objective-fact-stating. The assumption (no objective normative facts) does not by itself preclude any of the views described: normative language may purport to describe facts which do not exist.

1.2 Objective normative facts

What is meant by 'objective normative fact'? It is surprisingly hard to define this notion. Rather less ambitiously, I will characterize it.

Some things are the case no matter how we (human agents) feel, what we believe or are disposed to do: that life evolved, the Earth is round, the house on fire, the cat asleep (etc.) are all independent of our mental states. They are objective facts in precisely this sense: that whether or not they obtain is determined not by our beliefs, desires, experiences, dispositions, etc. Furthermore, whether they obtain is also not determined by human conventions like traffic rules or monetary systems.[1] By contrast, whether the soup tastes great, the sand is hot, the girl pretty or the driver committing an offence is not independent of human mental states or conventions. I will say that facts like the latter are subjective, whereas facts like the former are objective.

[1] Of course whether 'cat' *means* cat is not independent of our mental states or conventions, but that is beside the point. The question is whether it depends on our mental states or conventions that a sentence with a given meaning is true, or whether what is stated by a sentence with a given meaning actually obtains.

An objective *normative* fact would thus be a fact about what is good, wrong, ought to be done (etc.) which obtains independently of human mental states or conventions. They would be independent in the same way in which facts about the shape of the Earth and the wakefulness of cats are independent of such things (although the *nature* of the facts involved may differ).[2]

Here is one qualification: an act might be morally wrong because it causes pain, and pain is a mental state. So there is a sense in which some normative facts do depend on mental states (even according to objectivists). But this sense is not the one I have in mind. Normative (or, more narrowly, moral) objectivists believe the following: *that* an act is wrong because it causes pain (if it is wrong) is independent of human mental states or conventions. Its wrongness does not constitutively depend on us having various stances towards or prohibitions on pain-causing (and similarly for other normative statuses).[3]

But we need another qualification: according to Michael Smith, 'It is (morally) desirable that *A X*-es' is true just in case any observer who is (a) sufficiently informed

---

[2] It should be clear from the foregoing that when I characterize views as 'subjectivist' I don't mean to say they are examples of a particular metaethical view sometimes called *subjectivism*. According to this view, the moral statement '*X* is wrong' means 'I (the speaker) disapprove of *X*'. As I will use the term 'subjectivist', this particular view is a species of subjectivist views about morality.

[3] The notion of constitutive dependence is from Wedgwood (1997). According to him, the wrongness of an act constitutively depends on certain mental responses just in case it is *essential* to the act's being morally wrong that it (the act) stands in some relation to mental responses. In other words: the essence of moral wrongness involves some kind of relation to mental responses; that's just what moral wrongness *is*.

and (b) fully rational would desire that *A X*-es ((1994), chapter 5). So Smith makes the normative (or at least moral) status of an act depend on the mental states of ideal observers. Nevertheless, his theory could be thought of as objectivist. The reason is that all ideal observers would have the same desires (at least with respect to *A*'s *X*-ing). There would be no justification for this conjecture unless the coincidence of desires is somehow inevitable.[4] Assuming this to be the case, there is only one uniquely correct answer to any given question about moral desirability.

My characterization of objectivism excludes Smith's theory. We could remedy this by demanding that the normative facts are independent of *actual* as opposed to *hypothetical* mental states. But that is not a good idea. Philosophers like Bernard Williams (1981a) and Gilbert Harman (2000a) make the truth of normative claims depend on what a subject would be motivated to do were s/he not subject to ignorance or failures of reasoning (see chapter 3). Although these views make the normative status of an act depend on hypothetical mental states, they are not intuitively objectivist. For, unlike Smith, Williams and Harman allow that different people could be motivated to do different things even when conditions are ideal. So there is a sense in which they deny that there is only one uniquely correct answer to normative questions.

This suggests another way to characterize objectivism. We might say that objectivists about (certain) normative facts believe there is one uniquely correct answer to (the relevant) normative questions. But this will not do either. Suppose that one is

[4] In my view, Smith fails to provide good reason that it is.

a speaker-relativist, according to whom '*A* ought to *X*' means that I, the speaker, approve of *A*'s *X*-ing. Surely a form of subjectivism if anything is. But if '*A* ought to *X*' means that I approve of *A*'s *X*-ing, then the question 'Ought *A* to *X*?' is (presumably) a question about whether I approve of *A*'s *X*-ing. So long as it is determinate whether I approve of *A*'s *X*-ing, there is a uniquely correct answer to that question.

Of course, our first characterization of objectivism does rule out speaker-relativism. According to that characterization, normative facts do not constitutively depend on human mental states or conventions. But this made Smith into a subjectivist. Problems like these illustrate why it is hard to define objectivism in such a way as to rule in and out everything one intuitively wants to rule in and out. A satisfying definition will likely be complex and disjunctive, especially if it aspires to eliminating vagueness. My best stab at a simple statement is the following:

> **Objectivism:** an objectivist about normative facts believes that normative facts do not obtain in virtue of any variable ends, rules or attitudes of human beings and/or institutions.

The insertion of 'variable' is intended to put Smith into the objectivist camp. My characterization is vague in certain respects, but there is no need for greater precision. The fact that we are good at finding fault with proposed definitions shows that we have an intuitive grip on what should count as objective and subjective. This is why a characterization, based on some examples, will suffice.

Instead of *objectivist* and *subjectivist* views, I could have used the terms *realist* and *antirealist*. These are sometimes used to mark the same distinction. However, some influential definitions of realism bypass the question of objectivity. They demand no more of realism about some area of discourse than that some statements in the area are actually true under a nonrevisionary interpretation (Sayre-McCord (1988) and Railton (1996)). Such definitions make anyone into a realist who is neither an expressivist nor an error-theorist about the actual use of the relevant fragment of language. But, in my view, much philosophical interest in the "realism-debate" concerns the *nature* of the relevant phenomena. For example, philosophers want to know in what sense, if any, morality depends on human beings. This is why I describe the assumption underlying this thesis as the denial of objective normative facts, rather than the denial of realism.

1.3 Truth

In this chapter, I intend to motivate the focus of this thesis. It concentrates on non-error-theoretic, truth conditional theories of normative discourse. I don't know exactly what truth is (although I suspect it is correspondence), but I will take 'truth' to be a word that stands for a real property of propositions (or perhaps for a relation between propositions and the world). So by 'truth' I will not mean what deflationists mean by it, such as a syntactic device to express agreement or to indirectly express propositions without repeating them (Ayer (1936), Strawson (1950), Horwich (1990)). I will

assume, then, that '*p*' and "*p*' is true' do not mean the same, nor that the meaning of 'is true' is exhausted by the schema:

[*p*] is true if and only if *p*,

where '[*p*]' is some mentioned proposition and '*p*' is '[*p*]"s object-language equivalent (as defended by Horwich (1990)).

The assumption that truth is a property is relatively common, even if it is controversial exactly what truth is. It has significant theoretical benefits. For example, if truth is a property, we may be able to explain (an important aspect of) sentence meaning in terms of it (as in Davidson (1967) or Lycan (2010)). If it is merely a device for expressing agreement with an utterance, then we cannot explain its meaning in terms of the device. Similarly, if truth is a property, we can characterize the distinction between cognitive and noncognitive states by means of it: the first aim at truth (or accurate representation), the second have some other role. But there is probably no useful sense in which beliefs aim at agreement. We can further characterize the nature of assertion, logical validity, and a variety of other notions. For these reasons, I will assume that truth is a real property.

The assumption that truth is a property may not rule out so-called relativism about truth (Kölbel (2002), MacFarlane (2007)). Truth relativists make the truth of a proposition depend not just on a world (or a world and a time), but also on additional paramaters, like a standard of taste or a moral framework. It is not clear that this affects the *concept* of truth. The view that 'Socrates is sitting' expresses a proposition

which can be true and false at different times does not obviously require a novel conception of truth (which might still be correspondence). What it seems to affect is the question whether propositions exemplify the property of truth *eternally* (relative to a world). For example, if the proposition expressed by 'Socrates is sitting' really is *Socrates is sitting at 3 o' clock in the afternoon in the year 410 B.C.*, then it either corresponds or fails to correspond to reality. It makes no sense to say that it corresponds at some times but not at others. If, on the other hand, 'Socrates is sitting' expresses a proposition which does not specify the time at which he is sitting, then it sometimes corresponds and sometimes fails to correspond to reality. But it seems that either view can accept the correspondence theory of truth. However, I don't think that truth relativism (as applied to normative language) is a viable alternative to more traditional subjectivist theories. I will briefly explain why.

According to MacFarlane, the principal advantage of relativism is that it allows genuine disagreement between speakers with different values. More traditional *contextualists* would not, because:

> 'The contextualist takes the subjectivity of a discourse to consist in the fact that it is covertly about the speaker (or perhaps a larger group picked out by the speaker's context and intentions). Thus, in saying that apples are "delicious," the speaker says, in effect, that apples taste good to her (or to those in her group). In saying that a joke is "funny," she says that it appropriately engages her sense of humor (or that of her group).' (MacFarlane (2007), p. 18)

So when John says 'Apples are delicious' and Mary says they're not, then they don't disagree if they refer to different people (e.g. themselves). If relativism is true, however, then they *would* disagree because the relativist about 'delicious' does not build the reference to the speaker (or some other group) into the proposition expressed by 'Apples are delicious'. That way, we can both have 'subjectivity' (the idea that properties like being funny and delicious depend on subjective standards) *and* genuine disagreement:

> 'When I say that apples are delicious and you deny this, you are denying the very same proposition that I am asserting. We genuinely disagree. Yet this proposition may be true for you and false for me.' (Ibid., p. 21)

How is this supposed to work?

In standard (formal) semantics, the truth value of a proposition varies with the "circumstances of evaluation", which are usually taken to include at least (and at most) a possible world (MacFarlane (2005), p. 307). But the fact that something holds at the actual rather than some other possible world is not normally thought of as part of what is asserted or believed. For example, the proposition expressed by 'Grass is green' is not *Grass is green in the actual world*, because else it would be true as evaluated from all possible worlds. If this makes sense, MacFarlane thinks it should also make sense to put other elements into the circumstances of evaluation:

> 'Taking this line of thought a little farther, the relativist might envision contents that are "sense-of-humor neutral" or "standard-of-taste neutral" [...] and circumstances of evaluation that include parameters for a sense of humor, [or] a standard of taste [...]. This move would open up room for the truth value of a proposition to vary with these "subjective" factors in much the same way that it varies with the world of evaluation. The very same proposition – say, that apples are delicious – could be true with respect to one standard of taste, false with respect to another.' (Ibid., pp. 21-22)

But I question the existence of the requisite contents. What the truth relativist needs is for the property expressed by 'delicious' to be nonrelational. It has to be a property which is nothing to do with the relation in which the speaker stands to apples. Why? Because else relativism would collapse into contextualism. But although what is claimed of apples is nothing to do with the relation in which anyone stands to apples, whether or not they *exemplify* this property *is* determined by the standard of taste of an assessor of the proposition. I have no idea what such a property could be (nor does Jesse Prinz who makes a similar complaint in (2007), p. 182).[5]

[5] It is telling that relativists always argue for parameters on which the truth of a normative proposition might depend, taking it for granted that there *is* a proposition whose truth might so depend. None has ever offered anything analogous to conceptual analyses of words like 'good', 'reason', 'ought', etc. I fear that their restriction to high-level semantics blinds them to the fact that there might be nothing for words like 'good' and 'ought' to mean compatible with their views about the parameters of truth.

Here's one way to put the problem: if the property of being delicious is unrelated to our responses, then how come its exemplification depends on responses nonetheless? Compare this with the property of being triangular. The property of being triangular is nothing to do with subjective responses to the world. As a result, whether or not things exemplify it does not depend on subjective responses to the world either. But relativists need these things to come apart for 'funny' and 'delicious'.

However, there may be something relativists can do to guarantee the existence of the requisite properties. This is to identify properties with functions (Egan, Hawthorne & Weatherson (2005), Brogaard (2008)). I will briefly discuss this suggestion.

Some philosophers identify properties, like being triangular, with a function from a possible world to a set of individuals (the extension of the property at that world). To exemplify a property at a world is then a matter of falling within the extension at that world. Of course, the relativist maintains that being delicious is not merely relative to a world. It is also relative to a standard of taste. So the property of being delicious can't be a function from a world to a set of individuals, but has to be a function from a world + standard of taste to a set of individuals.

The relativist could say that the property of being delicious *is* a function from a <world, standard of taste> pair to a set of individuals. So long as this function exists, the property also exists. Identifying properties with functions explains why the property of being delicious is itself independent of our responses to the world, while the exemplification of the property does depend on our responses. Functions are abstract objects and therefore independent of human responses like enjoyment. But

whether an object exemplifies the property of being delicious *does* depend on variable responses, since the function that 'is delicious' denotes takes standards of taste as arguments.

Notice that relativism now stands or falls with a controversial view about properties as functions. That is undesirable by itself, but I also think this view of properties is false. I'll give three arguments. Some of them are directed at an identification of nonrelative properties with functions. But I take it that if the view that properties are functions is not plausible in the case of nonrelative properties, it is not plausible in the case of relative properties either.

First, identifying properties with functions appears to be conflating a tool for modelling properties with those properties themselves. It's a bit like conflating a graph that represents the temperature of an object over time with that temperature itself. George Bealer expresses this thought as follows:

> 'How implausible that familiar sensible properties are functions - the color of this ink, the aroma of coffee, the shape of your hand, the special painfulness of a burn or itchiness of a mosquito bite. No function is a color, a smell, a shape, or a feeling.' (Bealer (1989), p. 1)

Second, it seems that you can be familiar with a function from possible worlds to a set of individuals, without being familiar with the property that it is supposed to be. A colour-blind person may (in principle) know what objects a function associates

with a possible world (any possible world) without knowing the property they have in common (like redness).

Third, if one identifies properties with functions from possible worlds to sets of individuals, then necessarily coextensive properties would be the same properties. But having three sides is not intuitively the same property as having three angles.

So I don't think it is correct to think of properties as functions. But unless relative properties like being delicious can be identified with functions from <world, standard-of-*x*> pairs to sets of individuals, it is doubtful that there is anything which meets the requirements on relative properties.

For these reasons, I will assume that truth relativism is not an option in the semantics of normative language. I will proceed on the assumption that truth is a real property and that normative truth is absolute (i.e. nonrelative). A normative proposition is *absolutely true* if and only if standards held by an assessor of the proposition are irrelevant to its truth, except insofar as the proposition is about these standards.

1.4 Truth and meaning

In this thesis, I discuss analyses of words like 'ought', 'right', 'good', 'permissible' and 'reason'. Many proceed by specifying (informative) truth conditions for sentences containing such terms. According to a widespread view about sentence meaning, the meaning of a sentence in a context (or the proposition it expresses) is (at least) its truth condition (see, e.g., Lycan (2010)). I will not question this tradition here and

assume that the analyses discussed do specify, on the right-hand side of the bi-conditional, a proposition expressed by the sentence mentioned on the left-hand side. This does not rule out that there might be additional aspects to the meaning of a sentence, like modes of presentation or illocutionary force.

In what follows, I motivate further desiderata for an answer to our leading question (what is the best subjectivist theory of normative language?). I first explain why I prefer truth conditional over non-truth conditional approaches (sections 1.5 to 1.6). I then argue against error-theories (1.7 to 1.10).

1.5 Noncognitivism

Non-truth conditional theories of normative discourse have been defended throughout the twentieth century. Their focus tends to be on moral language, but they could be extended (after all, moral language is only a subset of normative language).

Non-truth conditional theories deny that the (essential) function of normative language is to state any kind of fact. Different philosophers flesh out the details of their views in different ways (e.g. Ayer (1936), Hare (1952), Gibbard (1990), (2003), Blackburn (1993), (1998)). What binds them together is the idea that the primary function of normative language is to express something other than a belief. By this phrase I don't mean that what is expressed by normative sentences is not propositional (although some, like Ayer, may have thought so). What matters is that the speaker does not stand in the belief-relation to the (primary) content of a normative utterance.

The qualification 'primary' in 'primary function' is necessary to include Hare's prescriptivism in the class of non-truth conditional theories. Hare believes that normative utterances have two types of content: prescriptive and descriptive ((1952), chapter 7).[6] The prescriptive content is what makes a statement normative. Hare thinks of it in terms of imperatives. But, typically, normative statements also have descriptive content. According to Hare, if I say 'That is a good car', I do two things: (1) I commend the car (recommend that you buy it) and (2) I indicate that it meets certain standards (has properties which explains why the car meets the standards). The speaker does stand in the belief-relation to the latter content.[7]

Hare's *prescriptivism* is slightly different from Blackburn's and Gibbard's *expressivism*. Blackburn and Gibbard believe that normative utterances express various noncognitive attitudes (desires, preferences, intentions, commitments, emotions). They (appear to) deny that what is expressed is best thought of in terms of imperatives, and also (appear to) deny that normative utterances have a descriptive element (although such elements may be inferred from context).

Non-truth conditional theories entail that there is no such thing as normative belief, since the role of normative utterances is not to describe reality. Since knowledge (cognition) entails belief, these views are often classified as *noncognitivist*. I

[6] Hare refers to prescriptive content as *evaluative* content.

[7] If it is correct to think of prescriptive content in terms of propositions, then it may well happen that the prescriptive and descriptive content of an utterance are the same proposition. In that case, what gives the utterance a prescriptive element is the fact that the speaker stands in some other relation than the belief-relation to that content.

cannot discuss these theories exhaustively, but I will argue that there is a strong presumption against them.

1.6 Problems of noncognitivism

My aim in this section is to explain why noncognitivism is a difficult view to defend. In the next, I'll discuss its advantages. We will see that they don't seem great enough to take on all its burdens. However, I don't pretend that my assessment is definitive.[8] Noncognitivism comes in many varieties and I won't discuss them all. I merely aim to create a presumption against the idea that the semantics of normative language involves non-descriptive elements.[9]

The best-known problem for noncognitivism is the Frege-Geach problem. It arises because noncognivists claim that the (primary) meaning of a normative utterance is given by something other than a belief or its content. For example, Hare characterizes the primary meaning of normative sentences in terms of a speech-act performed by them (i.e. recommending) ((1952), chapter 5). Blackburn characterizes their meaning in terms of the expression of noncognitive attitudes like disapproval (e.g. 'murder is wrong' expresses disapproval of murder) ((1998), chapter 3).

---

[8] My discussion bypasses contemporary kinds of *hybrid* noncognitivism, according to which normative sentences express both beliefs and desires (or other types of attitudes). See especially Ridge (2006), (2007). For a critical assessment, see Schroeder (2009).

[9] There is an extensive literature on all forms of noncognitivism. Some of the best recent discussion is found in Schroeder (2008) and (2010).

Peter Geach (1960), (1965) pointed out that sentences like 'murder is wrong' can be embedded in larger constructions like 'Either murder is wrong, or it is not wrong'. Someone who asserts this disjunction does not express disapproval of murder, nor does s/he recommend abstaining from murder. So if the meaning of 'murder is wrong' in assertoric contexts is given by a speech-act performed / noncognitive state expressed, it seems the meaning of the sentence must differ in contexts where the speech-act is not performed / the noncognitive state not expressed. But 'murder is wrong' clearly has the same meaning whether asserted or not. After all, 'If *p*, then *q*; *p*; Therefore *q*' results in a valid argument even when '*p*' and/or '*q*' are replaced by normative sentences.

The Frege-Geach problem, then, is that the noncognitivist's explanation of the meaning of normative sentences seems to entail that they have different meaning in asserted and unasserted contexts. In his (2008), Mark Schroeder summarizes what noncognitivists have said (and should say) in response (referring to Hare (1970)):

> 'Their answer is that normative sentences have the same meaning when embedded as when unembedded because the meaning of the complex sentence is a *function* of the meaning of its parts. Compare the ordinary truth-conditional semanticist's explanation of why 'grass is green' means the same thing when embedded in 'grass is not green' as when unembedded. *Prima facie*, expressivists suggest, there ought to be a problem. After all, 'grass is not green' does not have the same truth-conditions as 'grass is green'. So if meaning is truth-conditions, then how could 'grass is green' mean the same thing when

> appearing with 'not' as when appearing without it? The answer is that what we say is that the truth-conditions of 'grass is not green' are a function of the truth-conditions of 'grass is green' – a function that is given by the meaning of 'not'. Similarly, expressivists claim, 'murder is not wrong' expresses some mental state *distinct* from that expressed by 'murder is wrong'. But which mental state this is, is a function of the mental state expressed by 'murder is wrong', and the function is given by the meaning of 'not'.' (Schroeder (2008), p. 20)

As Schroeder argues, however, if the meaning of 'not' as applied to normative sentences is a function that takes you from a noncognitive state to a noncognitive state, then unless the same is true for 'not' as applied to nonnormative sentences (or a combination of normative and nonnormative sentences), noncognitivists are forced to say that 'not' is ambiguous ((2008), chapter 2 and chapter 7). Similar observations apply to other sentential connectives like 'and', 'or', 'if ... then', and to operators like 'knows that' and 'it is possible that' (the same point is made by Copp (1995), p. 17) This seems rather unappealing.

Schroeder's way around the problem is to suggest that all descriptive sentences (sentences with truth conditions) express the same kind of noncognitive attitude as normative ones (the attitude of *being for*). But as he notes himself, this may seem like a *reductio* of noncognitivism, and seems to run counter to what motivated it in the first place: namely a distinction between normative and nonnormative language in terms of a distinction between what is expressed by each (belief and some other kind of

attitude). Furthermore, Schroeder argues that the required hypothesis makes it hard to assign the right meaning to many constructions in natural language (among which past-tense operators, necessity operators and generics) (see (2008), chapter 12).

Even if noncognitivists could avoid claiming that all sentences express noncognitive attitudes (as argued by Wedgwood (2010)), it is hard to see how they could avoid ambiguity for at least some natural-language expressions. Take 'to know'. Many philosophers think that '*A* knows that *p*' entails that *A* believes that *p* and that *p* is true. But if noncognitivism is correct, '*A* knows that murder is wrong' cannot entail either that *A* believes that murder is wrong or that it is true that murder is wrong (remember that I'm assuming a nondeflationary theory of truth). So it seems that 'to know that *p*' must amount to something else for normative and nonnormative claims. And if Schroeder is right that his proposal does not give plausible interpretations of past-tense operatores, modal vocabulary and generics, these may turn out to be ambiguous as well.

So noncognitivists might end up assigning different meanings to quite a few terms that seem univocal. It is certainly fair to say that if they wish to adopt a truth conditional semantics for fact-stating language, then their overall theories of language will be more hybrid than the theories of those who include normative language in the fact-stating division. Noncognitivists' theories will be more hybrid not only in the sense that they require different machinery for sentences in different domains (factual and normative), but also *within* the normative domain, since all normative sentences contain at least some descriptive terms.

Even if truth conditional semantics turns out to be incorrect for fact-stating language, it is not plausible that noncognitivism will be correct for it (not even noncognitivists propose to apply their theories across the board; Schroeder is the first to suggest this). But any cognitivist analysis will be compatible with whatever theory of fact-stating language turns out to be correct (so long as it retains the ultra-plausible idea that fact-stating sentences *have* conditions under which they are true). So cognitivist theories of normative discourse have the advantage of being relatively isolated: they are proposals about the meaning of normative terms in particular, without affecting the interpretation of a variety of other terms. As a result, they require a simpler view of language than noncognitivists are likely to be saddled with.

A second well-known difficulty for noncognitivism is to explain the logical relations between normative utterances (see e.g. Schroeder (2008), chapter 2). If noncognitivism is correct, then such utterances cannot be truth-apt (except in a deflationary sense). So why do '(1) If murder is wrong, then *p*' and (2) 'Murder is wrong' entail that *p*? Normally, this is explained in terms of truth-preservation: the inference is valid because, necessarily, if (1) and (2) are true, then it is true that *p*.

Noncognitivists usually appeal to a "logic of attitudes" (e.g. Gibbard (1990), chapter 5 and Blackburn (1988); Gibbard does not refer to his view this way, but I believe their proposals are in crucial respects similar). On this sort of view, normative utterances inherit their logical relations from the (in)compatibility of associated ideal worlds (or rather the descriptions of such worlds). For example, 'Murder is wrong' and 'Murder is not wrong' are logically inconsistent because they respectively correspond to (say) a world in which murder is avoided and one in which murder is not avoided.

One can then explain why *p* follows from (1) and (2) in this way: assuming that 'If *p* then *q*' is equivalent to 'either not-*p* or *q*', premise (1) corresponds to an ideal world of which it is true that either murder is not avoided or *p*'. Premise (2) corresponds to an ideal world of which it is true that murder is avoided. Taken together, these rule out that not-*p* (in the standard way). It therefore follows from (1) and (2) that *p*.

But there are at least two difficulties with this kind of answer to the problem of logical relations. The first is that ideal worlds plausibly correspond to desire-like states, which take propositions as their objects. But Blackburn is also an expressivist about the word 'beautiful' and related terms of aesthetic appraisal ((1998), p. 110). Whatever attitude is expressed by 'Baroque music is beautiful', it is not plausibly desire-like. Or even if it is, it not easy to describe exactly what ideal world corresponds to the sentence. The problem arises from the fact that judging that baroque music is beautiful is compatible with an indefinite number of practical attitudes towards it (it may be that, even though it is beautiful, we should not listen to any). But from 'If baroque music is beautiful, then *p*' and 'baroque music is beautiful' it follows that *p* just as much as in the moral case. Similar problems may arise for other types of evaluative judgment (like 'That soup is delicious!').

If you didn't find the previous case convincing, try the following: what ideal world corresponds to the judgment '*A* has a reason to *X*'? Surely not a world in which *A X*-es, since the reasons for *X*-ing may be outweighed by the reasons for not-*X*-ing. Perhaps the most plausible option here is to say that the judgment that *A* has a reason to *X* corresponds to a world in which *A* takes the reason (or the fact which provides it) into account. In other words, it is a world in which *A* lets the reason play a certain role

in his or her practical deliberation. But it may be hard to specify what role this is without using normative vocabulary (is it the role of taking it to count *in favour of* something?). Furthermore, there may be counterexamples to any analysis of '*A* has a reason to *X*' in terms of a positive attitude towards having the reason play a certain role in one's deliberation. For example, Ronnie may have a reason to go home, since Bradley has organized a surprise-party for him. But if Ronnie took this reason into account, the surprise would be ruined (the example is from Schroeder (2007)).

The second reason why the logic of attitudes may not solve the noncognitivist's problems is that it presupposes that the attitudes expressed by normative utterances are the right sorts of thing to ground logical (in)consistency. Some philosophers doubt that simultaneously desiring that *p* and desiring that not-*p* is problematic in the same way as it is to believe that *p* and to believe that not-*p* (e.g. Schueler (1988) and Schroeder (2008)). In other words, some believe it is not *logically* inconsistent. The least that noncognitivists should do is identify mental states expressed by normative utterances which could plausibly explain why accepting that murder is wrong and that it is not wrong amounts to a logical mistake.

Mark Schroeder recommends that they use a primitive state of *being for* which could be expressed by all normative (and ultimately *all* kinds of) statements. In the version of noncognitivism that he favours, 'Murder is wrong' and 'Murder is not wrong' respectively express *BeingFor*(blaming for murder) and *BeingFor*(not blaming for murder). He does say that noncognitivists will have to argue for the claim that being for *p* and being for not-*p* are in the relevant way inconsistent, but he finds this reasonable and assumes it for the sake of argument ((2008), pp. 59-60).

I think this is surprising. It doesn't seem plausible to me that being for is primitive. What is it to be *for* something? Part of what it is to be *for* the death penalty is to desire its instigation, or to prefer it or at least approve of it. But Schroeder does not believe that any of these states could ground logical relations. Furthermore, *being for* appears to amount to different things in different contexts. If I am for Manchester United, I don't approve of Manchester United, but I desire that they win. Again, Schroeder does not believe that desire can underpin logical relations.

It could of course be that the state of being for is composite, consisting of a desire *and* something else left out by my analyses. And whatever is left out might then account for the sense that 'Murder is wrong' and its negation are logically inconsistent. But I have no idea what this something else might be and, in any case, the burden of proof rests on the noncognitivist.

So at least those philosophers who believe that mental states like desire and approval cannot ground logical relations should object to noncognitivism. I am not entirely convinced of this myself, but at least truth conditional views of normative language sidestep the problem altogether.

This brings me to my third objection to noncognitivism. I've already noted that it is not clear what ideal world is supposed to correspond to '*A* has a reason to *X*'. This is related to the fact that it is not clear what attitude is expressed by the sentence, or (alternatively) what object it takes. It is not the attitude of being for *X*, or the attitude of desiring that *A X*-es. Nor does it appear to be desiring that *A* takes the reason into account in practical deliberation, for the reason mentioned: it does not seem to work across the board. In addition, it may be hard to specify what *taking into*

account is supposed to be without using normative vocabulary. So I believe that noncognitivism does not look promising as a theory about reason-vocabulary.

It is now clear why there is a presumption against noncognitivism: (1) it may well assign different meanings to a variety of terms that seem univocal, (2) it has difficulty explaining the logical relations between normative utterances and (3) it is hard to see what attitude is expressed by judgments like '*A* has a reason to *X*'. It would be worth tackling these problems, or biting some of these bullets, if noncognitivism had powerful advantages unique to it. But this does not appear to be the case.

1.7 Advantages of noncognitivism

Noncognitivism has at least three advantages. First, it does not postulate objective normative facts, which some think are incredible (the *locus classicus* is Mackie (1977), chapter 1). But the denial of objective normative facts is not unique to noncognitivism. All subjectivist metaethical views deny their existence. This advantage, then, hardly gives us reason to take on the burdens of noncognitivism.

Second, if there is an intimate connection between sincere normative judgment and motivation, then noncognitivism can explain it. The view that this connection is in some sense necessary is called *normative judgment internalism* (NJI). According to it, a person cannot sincerely make a normative judgment without being defeasibly motivated to act in accordance with it. The reason why the motivation is defeasible is that it is implausible that one would never act against one's better judgment.

I think NJI is probably true for some normative judgments, although perhaps not all. A first qualification is that many normative judgments do not concern oneself. I may judge that Jim should finish his degree, but since Jim is the one who should do something, *I* cannot be motivated to act in accordance with it. So perhaps NJI is primarily true for first-personal normative judgments (judgments to the effect that I should do something, or that it would be better if I did something, etc.). However, even for second- or third-personal judgments, there might still be an indirect link with motivation. Sincerely judging that John should finish his degree may be impossible without being defeasibly motivated to act on the sorts of considerations that prompt my judgment about Jim.

A second qualification is that the judgment that *X* is artistically valuable is not obviously related to motivation in the way that some moral judgments seem to be. It might be possible to judge that a film is artistically good without being inclined to watch it (perhaps because it is too gruesome). If the link with motivation is more tenuous here, it is related to the fact that not all normative judgments are evaluations of *action*. Judgments of artistic value may fall into this category. For such judgments, NJI may not be true.

However, even if NJI were not true for all normative judgments, there would still be something to explain, namely why it is true for some judgments. Noncognitivism can explain this as follows: a normative judgment for which NJI holds is an expression of a noncognitive state. The particular states associated with those judgments are intrinsically motivating. So, anyone who makes a sincere

judgment expressing that state will be defeasibly motivated to act in accordance with it.

However, noncognitivism is not the only view that can explain the truth of NJI. One view I discuss in this thesis says that '*X* is right' is a statement about the requirements of rules to which the speaker is him/herself committed (at least in those cases where NJI is plausible in the first place). Since the speaker's commitment to rules is (in part) determined by his or her noncognitive states, this view explains why someone who sincerely judges that *X* is right is defeasibly motivated to act in accordance with the judgment. So in this respect, noncognitivism does not have an advantage over the cognitivist view just sketched.

The same is true for Stephen Finlay's view, according to which normative judgments are end-relational: they are (roughly) statements about what acts are conducive to contextually salient ends (Finlay (2009)). Whenever NJI is plausible, the identification of the end will proceed via the noncognitive states of the speaker. This allows us to predict that anyone who makes a sincere normative judgment of the relevant type will be defeasibly motivated to act in accordance with it. So although noncognitivism can explain NJI, several cognitivist alternatives can do so too.

The final advantage of noncognitivism is that it explains why normative judgments seem to resist reductive analysis. Many people think that analyses of normative vocabulary in terms of nonnormative vocabulary leave something out: namely the peculiarly normative element of normative language (e.g. Hare (1952), chapter 7 and Boghossian (2006)).

G.E. Moore famously argued for this conclusion as follows: for any proposed definition of 'good' in terms of some naturalistic predicate *N*, the question whether something which is *N* is good makes sense (and so, *mutatis mutandis*, for other terms like 'right' and 'ought'). The fact that this question appears to be open shows that no reduction of normative vocabulary to naturalistic vocabulary is possible ((1903), chapter 1, §12). Since the same holds for supernatural vocabulary, we might as well say that no *reductive* analyses (of any kind) are plausible.

The noncognitivist could marshall Moore's argument in favour of his or her position (as Blackburn indeed does in (1998), p. 86). S/he could urge that noncognitivism provides a good explanation of why reductions seem unsatisfying. They seem unsatisfying because they ignore the fact that the function of normative language is not to state any kind of fact at all (not even a special kind of normative fact, as Moore would have it). Rather, the normative element in normative language consists in the fact that something is *commended*, or that some noncognitive state is *expressed* rather than *described*.

Noncognitivism may provide an explanation of why reductive analyses seem to leave something out. But their explanation overgeneralizes, since nonassertoric force is defeasible. It is not clear that anything is commended in the following sentences, for example:

(6) In Japan, people ought to burp during dinner.

(7) Adultery is good for destroying marriages.[10]

(8) This tree has got good roots.[11]

Nor are favourable attitudes expressed in disjunctions or attitude reports. So it seems too drastic to suppose that the semantics of normative terms involves noncognitive elements. The defeasibility of recommendation (etc.) suggests that there might be an alternative explanation of why reductions seem inadequate. In chapter 7, section 7.2, we will see that pragmatics allows us to explain why normative sentences (in assertoric contexts) tend to carry informational significance in addition to the communcation of facts. The robustness of this connection might explain why reductive analyses seem to leave something out.

More could be said. However, I think enough has been said to cast doubt on the claim that the advantages of noncognitivism are great enough to take on all its burdens. I therefore think we should concentrate on truth conditional accounts of normative language. I consider an account truth conditional just in case it denies that the semantics of normative terms involves non-truth conditional elements.

In the remainder of this chapter, I explain why I will also set aside error-theories.

[10] This example is from Finlay, ms, chapter 6.

[11] This example is from Foot (1961).

1.8 Error-theories

Error-theories are truth conditional accounts of *moral* language in particular. According to them, moral statements purport to describe moral facts. But they don't succeed in doing so, because such facts do not exist. The point is often put by saying that moral language involves a false presupposition. The presupposition is that moral properties and relations are objective in some special, problematic sense.[12]

So error-theorists think that the presupposition affects the application conditions of moral concepts (and thereby the truth value of moral statements). What exactly is this presupposition? This is not always clear. John Mackie suggests (amongst other things) that moral properties and relations are supposed to provide agents with reasons to act irrespective of their desires or commitments:

> 'The ordinary user of moral language means to say something about whatever it is that he characterizes morally [...] [which] involves a call for action [...] that is absolute, not contingent upon any desire or preference or policy or choice.' (Mackie (1977), p. 33)

[12] Notice that my definition of objectivism (section 1.2) makes the error-theorist into a moral objectivist. This is correct in the sense that their *conception* of what a moral fact would be is objective. What makes error-theories relevant to this thesis is that they share the subjectivist's belief that no objective moral facts exist. So although the error-theorist shares the objectivist's semantics, s/he shares the subjectivist's ontology.

Stephen Finlay describes the presupposition as '*absolutism* about the normative authority of moral value' ((2008), p. 348). This notion may be preferable to that of reason-giving. Finlay's subjectivist theory does not tie the truth conditions of statements about reasons to the desires of *agents* (see chapter 6, section 6.5). But that does not commit him to anything that Mackie would consider "queer" (such statements reflect commitments of the *speaker* instead). But 'normative authority' needs some explication too, and so does absoluteness.

Richard Joyce describes the presupposition as follows:

> 'The Mackian error theoretic argument claims (a*) that moral discourse presupposes non-institutional desire-transcendent reasons and non-institutional categorical imperatives, while mainting (b*) that all genuine desire-transcendent reasons are institutional and all genuine categorical imperatives are institutional.' (Joyce (forthcoming))

By an *institutional* categorical imperative Joyce means the following: a demand for action, such that a social institution allows it to be applied to people irrespective of their particular desires and interests. An example of such demands would be requirements of etiquette (a social institution). An institutional desire-transcendent reason is (presumably) a reason such that a social institution allows its ascription to people irrespective of their desires. An example is the reason gladiators have not to throw sand in their opponent's eyes (from Joyce (2001)). This in turn is determined by the rules of gladiatorial combat (another social institution).

Joyce agrees with Mackie that moral language presupposes the existence of non-institutional desire-transcendent reasons and imperatives. He also agrees that no such reasons and imperatives exist (or could be "valid"), so that all interesting moral statements are false.[13]

It's a little odd that Joyce singles out *institutions* as the things that moral requirements and reasons are supposed to be independent of. I take it that an institution like etiquette is a more or less generally recognized set of rules that are in some sense arbitrary. The same is true of gladiatorial combat. But many ("genuine") normative judgments are not in this sense related to a social institution, like 'You ought to see that film'. True, this judgment is (possibly) sensitive to the agent's desires (based on the fact that s/he might enjoy it), but 'Get out of my house!' hardly needs to be. Nor is the latter obviously related to a social institution. The point is this: more might be required for a reason or imperative to be problematic (in the sense required for an error-theory) than that it is non-institutional and desire-transcencent.

But we can bypass these problems. My characterization of objectivism in section 1.2 gives us the material we need. So let us say that moral language presupposes that moral facts do not obtain in virtue of any variable ends, rules or attitudes of human beings and/or institutions.

An error theory of moral language will then consist of two claims:

[13] Except for analytic ones, like 'Everything that is good, is good' and some negations like 'Murder is not wrong'.

(i) *Semantic:* the truth of moral sentences requires the existence of objective moral properties and relations (i.e. properties and relations that do not obtain in virtue of variable ends, rules or attitudes of human beings and/or institutions).

(ii) *Metaphysical:* there are no objective moral properties and relations.

It follows that many ordinary moral statements like: 'You morally ought to keep your promises' and 'Torture is morally wrong' are false.

Like noncognitivism, error-theories have been defended primarily for moral language. But unlike noncognitivism, error-theories cannot simply be extended to other normative domains. They are usually defended on the grounds that there is something special about moral language in particular, something which generates the error underlying moral but not other forms of normative discourse. For example, John Mackie says:

> '[T]here are certain kinds of value statements which undoubtedly can be true [...], even if, in the sense I intend, there are no objective values. Evaluations of many sorts are commonly made in relation to agreed and assumed standards. The classing of wool, the grading of apples, the awarding of prizes at sheep dog trials, flower shows, skating and diving championships, and even the marking of examination papers are carried out in relation to standards of quality or merit, which […] are fairly well understood and agreed by those

who are regarded as experts in each particular field.' (Mackie (1977), pp. 25-26)

So Mackie is not an error-theorist about statements like (3) 'That is a good car'. Nor does he want to be an error-theorist about (4) 'We have reason to believe that life evolved' or (5) 'If you want to succeed, you have to pay attention'. Nor need he be an error-theorist about (1) 'You ought to keep your promises', unless the 'ought' has moral connotations. Mackie's error-theory, then, is limited. It is a theory about statements like (2), i.e. about moral language in particular.

This shows that if there is an error in moral language, it is not located in words like 'ought', 'good' and 'reason'. After all, they do not radically change meaning in nonmoral contexts,[14] and there is no error there. So if moral language does involve an error, it must be located in the adverbs and adjectives that (implicitly or explicitly) qualify words like 'ought', 'good' and 'reason'. For example, the word 'morally' in 'What John did was morally wrong' is what makes this judgment false. Presumably, the relevant adverbs and adjectives get to do this because our concept of morality is the concept of an objective entity.

[14] At least the hypothesis that they do is costly: it would lead to radical proliferation of the number of concepts expressed by 'ought', 'good', 'reason', etc. And this is not the only reason to think that the hypothesis is false: if 'ought' had radically different meaning in nonmoral contexts, it would be mysterious what the function was of adverbs like 'morally' and 'prudentially'. They would not be required to generate the meaning of 'ought' in moral contexts compositionally by combining the sense of 'moral' with the (domain-neutral) sense of 'ought'.

This makes the error-theory of limited interest to my project. My project is primarily about normative words like 'ought', 'good' and 'reason', but not just in moral contexts. It also shows that an argument against the error-theory should target the idea that our concept of morality is the concept of an objective entity. It is to the latter task that I now turn.

1.9 The presumption against error-theories

There is at least a presumption against error-theories of moral language. The presumption exists because of a principle of charity governing the interpretation of language. Other things being equal, one's assignment of semantic values to terms should guarantee the correctness of fundamental aspects of their use. So if there is a reasonable assignment which does not condemn all interesting moral claims to falsehood, it is to be preferred. This means that if the data are at least compatible with a "success-theoretic" interpretation, it has a significant advantage.

In what follows, I will summarize Stephen Finlay's discussion of the data in (2008). Finlay argues that they are in fact compatible with a success-theory. I have little to add to his discussion.

As indicated, Finlay describes the purported presupposition as '*absolutism* about the normative authority of moral value'. We can take this to mean that moral values are independent of variable ends, rules or attitudes of human beings and/or institutions. What evidence might there be for this presupposition?

Reflective evidence: Richard Joyce ((2001), p. 97) claims that absolutist interpretations of moral language are explicitly avowed by users of that language. But (a) this is far from universal, since many seem to accept some kind of relativism. (b) How to interpret people's claims is not straightforward of itself. If we ask 'Is it a fact that torture is wrong?' it is not clear whether positive responses reflect metaethical or normative ethical beliefs (one could interpret the question as asking whether one is *certain* that torture is wrong). And so we cannot put much weight on such claims in the first place. (c) Even a universally accepted theory about the reference of a term can be false. It was universally accepted that water was an element (instead of a compound) for centuries. But we don't say that people did not refer to water at the time. So beliefs about the semantics of a term needn't affect the truth conditions of statements containing it.

Linguistic evidence: the surface appearance of (many) moral statements is absolute or categorical. E.g. we say: 'You ought to keep your promises', not 'If you want to ..., then you ought to keep your promises'. But many normative statements that are clearly relative are superficially absolute as well. The captain of a rugby team does not prefix his statements with 'If you want us to score' or 'In order to win this game'. In a context where the end is clearly shared, it is not required to state the obvious. So if moral standards are typically shared by members of a society, or assumed to be shared, or easily identifiable (as, arguably, they are), then this can explain the surface absoluteness of moral statements.

Appraisal evidence: in assessing the moral quality of an act, we (typically) don't take into account whether it did or would obstruct the agent's desires or ends or standards. We say that Hitler's actions were wrong even if we are convinced that he did not share any relevant concerns. All this shows, however, is that moral judgments are (probably) not sensitive to the desires or ends of *agents*. It does not show that they are not sensitive to the desires or ends of *appraisers*.

The next three forms of evidence all derive from situations in which we know or believe that the addressee does not share our standards, ends or desires. First, *address evidence:* people address surface-categorical moral judgments even to people whom they know do not share our concerns. Second, *expectation evidence:* in addressing people in this manner, people seem to think they might succeed in influencing the hearer by providing her with categorical reasons. Third, *disputation evidence:* even in fundamental moral disputes, people seem to proceed on the assumption that they are discussing the same subject matter (i.e. of what is morally right). But if their judgments are relative to different standards or ends, then they are really talking past each other.

In response, Finlay makes several useful points: (a) these phenomena are not as widespread as one might think. It is very rare to meet people who do not share any of our concerns (this is true even in pluralistic societies). But it is even rarer to *believe* that they do not. The phenomena adduced only count as evidence if the appraisers acknowledge that they are involved in a dispute with someone who is morally alien. But it is not clear how much evidence of this kind there is. Furthermore, it is not clear

that we *would* engage in disputes with people we believe to share none of our concerns (we may be wary of this enterprise). (b) Even if the relevant phenomena occur, there might still be explanations of them which don't require an absolute semantics. For example, there might be a rhetorical point to insisting that an act is wrong. One might be communicating one's insistence that the agent abstains from this behaviour (even if there is no hope that s/he will do so). This brings us to the final piece of evidence that Finlay discusses.

Reactive attitude evidence: we blame those who (we judge) acted wrongly regardless of their own ends or standards. This betrays a belief that those blamed failed to act on reasons that were authoritative for them (again, regardless of their *own* ends or standards).

In response, Finlay first questions the relevance of this alleged fact to the question at hand: the question is primarily whether judgments about wrongness, rightness, goodness, badness (etc.) involve the ascription of absolute moral properties and relations. Even if blaming someone for doing something bad/wrong (etc.) presupposes a commitment to the existence of reasons with absolute authority, this needn't show that judgments about badness and wrongness (etc.) do so too.

I don't think this response is particularly strong, however. If the reasons which make the act bad or wrong are not the same as those which the blamees are supposed to have ignored, then what reasons *are* they supposed to have ignored?

But Finlay's second response is more promising. This response involves an alternative theory of what it takes to be blameworthy. It is partly motivated by the

observation that we don't always blame people for not responding to reasons that have authority for them. For instance, we don't blame people for not responding to prudential or instrumental reasons. So '[i]t is plausible to suppose that we blame people only when they act against ends or standards that are important to *us*' (Finlay (2008), p. 359).

This does not explain why we don't blame animals or people who act out of ignorance or incapacity. But this can be accounted for by a requirement of knowledge or ability to know (which animals lack under all conditions and humans under some). So it seems that reactive attitudes are not very good evidence for a false presupposition either.

These results are important in light of the presumption against error-theories described at the start of this section. Considerations of charity tell us that if the data are at least compatible with a success-theoretic interpretation, it is to be preferred. I think Finlay's discussion shows that the evidence is compatible with a contextualist theory according to which moral statements implicitly refer to (a particular kind of) standards or ends that can in principle vary with the speaker.[15]

[15] Some philosophers who defend error-theories about the literal content of moral utterances nonetheless believe that moral language is (in an important sense) kosher. They believe that moral propositions are not really asserted and not intended to depict reality. Rather, they are useful fictions for influencing and coordinating action (Kalderon (2005), Nolan, Restall & West (2005)). Although fictionalism may in certain respects be preferable to simple error-theories, it is no more plausible in the

We can say a little more (although not very precisely). It is not just that we should be charitable in our assignment of semantic values to terms, we should also assign them on the basis of what application of the term is *sensitive* to.[16] For example, we assign cows as the referent of 'cow' because we apply the term 'cow' to cows. The difficulty in making this thought precise is that we can be wrong in applying a concept. But presumably there is a link between a reasonable assignment of referents and the dispositions of subjects to affirm or retract certain judgments. For example, it is plausible that 'water' refers to H_2O even on Twin Earth only if some subset of the community *would* retract judgments that some stuff is water upon discovering that it is composed of XYZ. If no one had this disposition before the rise of chemistry, there is no justification for the claim that the reference of 'water' was already confined to H_2O. So 'water' is sensitive to H_2O only if the linguistic community has a disposition to apply the term to H_2O and not to other things.

Now suppose there are no mind-independent normative facts. In that case, people will make normative judgments on the basis of their own attitudes, standards or ends. It then seems that the application conditions for normative terms will consist of facts about the relation in which acts and objects stand to these attitudes, standards or ends (whether forbidden or required, etc.). For example, people will have a disposition to say '*X* is wrong' just in case *X* is forbidden by the standards that they are invoking. So if subjectivism is true, then the use of normative terms will be sensitive

---

light of charity. After all, fictionalism still entails the falsity of the literal contents of all interesting moral statements.

[16] Finlay develops a similar thought in (2008), pp. 365-369.

to the instantiation of relational properties. This makes it plausible to assign relational truth conditions to normative sentences.

1.10 Conceptions of morality

Let me add one more observation. I've said that the semantic culprit of the falsehood of moral statements is likely to be expressions that qualify words like 'ought', 'good' and 'reason'. These are (implicit or explicit) adverbs and adjectives connected to the concept of morality. So the error-theorist thinks that our concept of morality is the concept of an objective set of principles or reasons.

The attraction of this view is somewhat diminished when we ask what an alternative conception might look like. One might think the alternative would have to involve the idea that morality is subjective instead. But that is not so. An alternative conception could be *neutral* about the metaphysical status of the relevant principles or reasons. Such a conception would be functional and/or determined by the content of morality.

In (1984) David Wong characterizes morality by means of its function. According to him, the point of moral rules is to resolve 'internal conflicts of requirements [...] that affect others' and 'interpersonal conflicts of interest in general' (p. 38). This is a functional characterization. A content-characterization would (also) mention particular things that a recognizably moral system tends to promote or protect. These may include (human) life, welfare, the sustaining of interpersonal cooperation, etc. I don't think anyone can provide a characterization of the function

and/or content of morality that would not harbour some vagueness or has some unwanted implications. But this is true of almost anything and hardly a reason to reject functional or content-conceptions of morality.

The advantage of such conceptions is that they leave it open what the metaphysical status of the relevant system is. To see that this is an advantage, consider prudential 'ought's. Is it plausible that the prudential judgment that John ought to *X* entails information about the metaphysical status of prudential principles (their objectivity)? I don't think it is. Unless we have strong reason to think moral principles are different, we should prefer an anti-metaphysical conception of morality.

1.11 Conclusion

I have tried to justify (some of) the focus of this thesis. Its leading question is: *On the assumption that there are no objective normative facts, what is the best theory of normative language?* I have argued that truth conditional theories are (*prima facie*) preferable to noncognitivism, since its advantages don't seem to outweigh its problems. I have also argued that the data do not support error-theories of moral language, against which there is a strong presumption. So we are looking for a non-error-theoretic, truth conditional view of normative language.

In the next chapter, I will present an argument for contextualism, the view that the proposition expressed by normative utterances varies with the context of utterance. This further limits the theories to be discussed.

Chapter 2. An argument for normative contextualism

2.1 Introduction

In the previous chapter, I've argued that noncognitivism is not our best bet when it comes to normative language. I've also suggested that an error-theory would not tell us enough about the meaning of words like 'ought' and 'reason', since they do not merely occur in moral contexts. Error-theories are primarily theories about the contribution of adverbs and adjectives like 'morally' and 'moral'. But even here, an error is not obviously favoured by the evidence.

In this chapter, I want to narrow down the options even further. I will present an argument for contextualism about normative words. Contextualism is the view that the truth conditions of normative judgments vary with the context of utterance.

Quite generally, a contextualist about a word *w* believes that the contribution it makes to the truth conditions of sentences containing it can change with the context of utterance.[17] This kind of view is extremely plausible for gradable adjectives like 'is tall'. Someone can be tall compared to some people, but not compared to others. So whether '*X* is tall' is true depends on a comparison class.

Contextualism is also plausible for knowledge-ascriptions (see e.g. DeRose (1992)). Whether '*A* knows that *p*' is true seems to depend on the standards of

[17] This formulation may need supplementation. It makes one a contextualist about any word that is polysemous. But this is not obviously wrong. Contextualists about knowledge do believe that the meaning of 'know' varies with the standards for knowledge salient in the conversation.

knowledge applied. If nothing important is at stake, I know that John was in the office because an otherwise reliable source told me. But in a context where the police investigate a murder, I may be reluctant to say 'I *know* that he was in the office'.[18] Evidence that would otherwise suffice for a true knowledge ascription may be insufficient if the standards are raised. And so whether one knows that *p* seems to depend on the standards of knowledge applied.

Note that the details of contextualism about gradable adjectives are different from contextualism about knowledge. It is not that *A* knows that *p* because *A* is *more* of something compared to certain other people (as in the case of tallness). Rather, *A* knows that *p* because *A* satisfies certain standards (it doesn't matter whether others do so more or less). So the kind of relationality involved in sentences with gradable adjectives is different from the kind of relationality involved in sentences with the verb 'to know'. The former involves a relation to a comparison class, the latter a relation to standards. But at least there is nothing unfamiliar about context-sensitivity. In this chapter, I will provide some arguments for contextualism about normative words. I will also attempt to rebut important arguments against it.

In this thesis, I am primarily interested in propositions as the meanings of sentences used in contexts. Contextualism is first of all a view about the truth conditions of sentences uttered in a context. It can also, but needn't, be a claim about the proposition expressed by a sentence in a context. A contextualist about *w* who believes that meanings *are* truth conditions is also committed to the claim that the

[18] This example is taken from DeRose (2010).

contribution made by *w* to the proposition expressed by sentences containing it can vary with context. And so is anyone who believes that two sentences-tokens cannot express the same proposition if their truth conditions differ. I certainly accept the latter. So in my view, contextualism about 'ought' is the view that the contribution made by the word 'ought' to the proposition expressed by a sentence involving it varies with context.

2.2 The context-sensitivity of 'ought'

It should be uncontroversial that a word like 'ought' is in certain respects context-sensitive. Take the following sentences:

(1) John ought to be faithful.

(2) This place ought to be a club.

(3) It ought to rain tomorrow.

(3) involves what is sometimes called the 'predictive' ought. Presumably, what makes it true that it ought to rain tomorrow is in certain respects different from what makes it true that John ought to be faithful. Even in normative uses of 'ought', there seem to be differences. For example, 'John ought to be faithful' is specifically addressed to *John* (he is the "owner" of the requirement). But 'This place ought to be a club' does not appear to be owned by anyone in particular.

And there is another kind of context-sensitivity (called *information-relativity* by Björnsson & Finlay (forthcoming)). Suppose we are with the police and our best information strongly suggests that the hostage is held in the Hilton. We might then, apparently truly, say: 'We ought to concentrate our efforts on the Hilton'. But an onlooker who knows that the hostage is held in the Ritz might, again apparently truly, say: 'They ought not to concentrate their efforts on the Hilton, they ought to concentrate their efforts on the Ritz'. A reasonable way to secure the truth of both claims is to say that the truth of the first depends on the state of information of the police, whereas the second depends on the state of information of the onlooker.

In the next section, I will argue that there is context-sensitivity even within normative uses of 'ought' that are addressed to the same agent and indexed to the same state of information. This type of context-sensitivity of 'ought' plausibly carries over to other words, like 'good' and 'reason'. Since it is usually what people have in mind when they talk about normative contextualism, this is what I'll take the term to mean.

2.3 An argument for contextualism

It is widely accepted that 'ought's can be relative to types of consideration. Moral considerations may favour *X*, but prudential considerations may favour the opposite. If this is right, then at least certain uses of 'ought' appear to be relative to (something like) standards, rules or ends. So some 'ought' propositions may involve a reference to a variable standard (or principle, end, etc.; henceforth I will omit these alternatives).

It may seem that this consideration does not really support the view that 'ought' propositions contain a reference to standards. One can hold that 'ought' propositions are not relational, even if supported by different kinds of consideration. However, I think this option has a drawback. If 'ought' is not relational, it cannot be simultaneously true that an agent ought morally to *X* and prudentially to not-*X*. If 'ought' is not relational, it is either the case that *A* ought to *X* or it isn't. Likewise, an object *O* cannot both have and lack a property *F* at the same time. Even though we can make conditional judgments about what the agent *would* ought to do provided certain considerations were the only ones applying, it wouldn't *actually* true that s/he ought, say, morally to *X* but prudentially to not-*X*.

The situation would be analogous to one where there are reasons of type A to *believe* that an object has some property, and reasons of type B to believe it does not. But those do not *make* the object have the property relative to A-type reasons and not have it relative to B-type reasons. It either has it or it doesn't. Similarly, although one may suppose (for moral reasons) that *A* ought to *X*, that does not make it the case that *A* ought to *X* relative to those reasons. Not, that is, if 'ought' is not intrinsically relational.

This latter consideration loses its force if the judgment '*A* ought morally to *X*' itself expresses a subjunctive conditional (something like 'Had morally considerations been the only ones applying, then it would have been true that *A* ought to *X*'). But, firstly, '*A* ought morally to *X*', does not appear to express a conditional (and no one I'm aware of actually defends this view). Secondly, if I say that *A* ought morally to *X*, part of what I indicate is that the moral considerations that support this judgment

actually apply. But that cannot be inferred from the subjunctive: 'If moral considerations were the only ones applying, then *A* ought to *X*'. The subjunctive may be true even if no moral considerations actually apply.) Third (and for the same reason), if '*A* ought morally to *X*' expresses a subjunctive conditional, we cannot infer from it that there are moral reasons for *A* to *X*. But we can. Therefore '*A* ought morally to *X*' does not express a subjunctive conditional.[19]

My argument for contextualism about 'ought' is that it explains why '*A* ought morally to *X*' can be true at the same time as '*A* ought prudentially to not-*X*'. Contextualism explains this by hypothesizing that 'ought' propositions have an argument place for (something like) a system of standards.

Once we allow this, however, our background assumption of subjectivism virtually forces us to accept that the standards can vary even *within* a single normative domain (like morals or prudence). For the assumption would make the standards (at least in some interesting cases) depend on the variable commitments of people. And why would someone who makes, say, sincere moral judgments be referring to the commitments of others? It is more likely that they would be referring to their own

[19] It may be possible to solve the last two problems by tweaking the content of the conditional. But let us not forget that I am engaged in a particular project. It is the project of finding a success-theoretic, truth conditional theory of normative language on the assumption that there are no objective normative facts. Although there is logical space for a position which satisfies these requirements and holds that 'ought' is not relational, it is hard to see what it would be. So even if relativization of 'ought' to a certain type of consideration does not favour a relational theory over an objectivist, nonrelational semantics, it does favour a relational *subjectivist* semantics.

standards. But if these vary among people, then the truth conditions of moral judgments will vary too.

Is there some independent reason to think that the standards implicitly referred to in normative statements can vary *within* a single normative domain? I think there is.

Imagine a far-away island. This island has never been visited by people from a different culture. Anthropological research soon reveals that the natives have many taboos, and also insist on many actions. The relation between their utterances and practices of punishment and reward strongly suggest that they have normative vocabulary. Some of their statements are best interpreted as claims involving 'ought'. Some are best interpreted as *moral* 'ought's. Assuming there is an argument place for a system of standards in 'ought' propositions, what system should we assign to those expressed by native moral statements?

If there is a large discrepancy with what Europeans judge is morally required, it seems uncharitable to interpret either Europeans or the natives as largely mistaken about the requirements of a single system (provided, of course, that both groups get the nonmoral facts right, at least for the most part). If these conditions are satisfied, it makes a lot more sense to assign different systems to their respective utterances. Therefore, different standards (can) plausibly matter for the truth of 'ought' judgments *within* a normative domain.[20]

[20] For my purposes, it does not matter whether this kind of scenario is actual (although I think it is). Its possibility suffices to establish that 'ought' propositions within a single normative domain may in principle refer to different systems.

There is, of course, nothing special about the islanders. Islands may exist within our own society (as indeed they do). If there is sufficient overlap between standards, it needn't be apparent to speakers that they are referring to different systems. But they may do so nonetheless (we will return to these issues in section 2.4).[21]

We need to ask why this is so. It is not because the moral system a speaker is referring to supervenes on the practices in force in (his/her portion of) society. That would not allow sufficient room for dissidence. Rather, what system a speaker is referring to is determined by his or her commitments. This explains why societal norms fill the argument place for at least many native 'ought's: we can expect a substantial portion of the population to be committed to the standards and practices generally endorsed on the island.

However, there are many contexts in which the argument place in 'ought' propositions is filled by a system to which the speaker is not him/herself committed. For example, we may be discussing what one ought *legally* to do. Or the speaker may be engaging with the *hearer's* moral perspective. It is hard to specify exactly when the speaker's own commitments determine the system referred to in an 'ought' claim. But it is plausible that they matter for, e.g., sincere moral judgments that are not 'engaged'

[21] Some may wish to claim that there isn't much fundamental disagreement in the world. It may be that most superficial disagreements are explained by false beliefs and subsequent mistakes in the application of standards. That would be helpful, since it makes it easier to defend a contextualist view like mine. If people tend to refer to the same system within a normative domain, the famous problem of disagreement is not as urgent as it seems.

in the way described. Of course, even here speakers needn't be referring to different systems. That depends on how much fundamental moral disagreement exists (see also footnote 21). All I meant to argue was that the standards implicitly referred to by 'ought' statements can in principle vary even within a single normative domain.

In summary, my argument for contextualism about 'ought' is this:

First stage

Premise 1: '*A* ought morally to *X*' can be true even if *A* ought not to *X* from some other point of view.

Premise 2: If 'ought' were nonrelational (i.e. didn't have an argument place for something like a system of standards) then premise 1 would be false.

Therefore: 'Ought' contains an argument place for (something like) a system of standards.

Therefore: Contextualism is true for 'ought' claims made within different normative domains (morals, prudence, etc.).

Second stage

Premise: The argument place of (many) moral 'ought' propositions expressed by natives on the island is filled by (something like) a nonwestern system of standards.

Therefore: Contextualism is true for 'ought' claims within a single normative domain.

More informally, the argument runs as follows: first I've noted that 'ought's can be relativized to types of consideration (moral, prudential, etc.). This makes it plausible that there is an argument place for (something like) standards in 'ought' propositions. Next, I've argued that the way the system is determined (namely via the commitments of the speaker) makes it possible that different speakers can refer to different systems even within a single normative domain. The extent to which this happens is of course an empirical matter, and the less it happens the easier it is to defend contextualism (as we shall see).

I kept inserting remarks qualifying the assertion that 'ought' propositions contain an argument place for a system of standards (I said that they contain a place for *something like* a system of standards). The reason is this: the observation that 'ought's can be relative to normative domains does not by itself support a standard-relational view as opposed to, say, an end-relational one, or even a view according to which 'ought's are relative to motivations. In subsequent chapters, I will discuss these options.

2.4 Arguments against contextualism: problems of disagreement

Although there is a good argument in favour of contextualism, many have been brought against it. The most serious of these are all, directly or indirectly, related to the problem of disagreement.

Contextualism about 'ought' does not (always) allow contradictory propositions to be expressed by speakers who respectively say '*A* ought to *X*' and '*A* ought not to *X*'. For example, if the first statement refers to a system of standards accepted by one speaker, and the second to a system accepted by the other (and the two are relevantly different), then they do not contradict each other. They merely state facts about the requirements of different systems.

The related problem is sometimes called the problem of disagreement: on the face of it, '*A* ought to *X*' and '*A* ought not to *X*' appear to be incompatible. But contextualism does not always make it so. Some authors consider this problem insurmountable. For example, Mark Timmons (1999) dismisses contextualism on the basis of it alone. And recently, Lars Binderup wrote:

> 'A well known and in my view insurmountable problem with [moral contextualism] is that it fails to account for the very *possibility* of moral disagreement between speakers who do not share moral standards.' (Binderup (2008), p. 410)

Notice that contextualism does allow some disagreement of the kind described (let's call it 'propositional disagreement' or 'p-disagreement', since it concerns disagreement about the truth value of the same proposition). Any view that does not

involve first-personal possessive determiners (as in '*My* system requires *X*') allows that two speakers referring to the same system express incompatible propositions if the one says that *A* ought to *X* and the other denies it.

The problem of p-disagreement is most likely to arise in cases where the speaker's own commitments determine the systems to which s/he is referring. If this is not the case (as in discussions about legal requirements, or what one ought to do *given* someone's standards) the participants in the conversation are likely to refer to the same system anyway. And if they don't, the intuition that they are expressing contradictory propositions is not very strong to begin with (imagine two people inadvertently referring to different legal codes). So from now on I will discuss the problem of p-disagreement as a problem arising in contexts where the speakers' commitments determine the systems to which they are referring.

It's possible that the problem of p-disagreement is most pressing for moral discourse. After all, if I say 'The soup is great!' and someone else denies it, we needn't feel that we are contradicting each other. This is because the greatness of soup is related to one's taste in food, which we know is variable. We also accept that some things are beautiful to some people, but not to others. Similarly, when aunt Betty says 'You ought not to speak with your mouth full', we may "disagree" without thinking that she spoke falsely.

Notice that even in cases of moral discourse, p-disagreement is possible when there are differences in the standards accepted by the speakers. Only some of them may be relevant at all. So even if the systems to which they are committed contain different standards, the standards which are not shared may not bear on the situation.

In such cases, it seems plausible enough to construe their respective statements as referring only to subsets of the standards to which they are committed.

But even if there is a lot of p-disagreement, it does not solve the problem completely. Contextualists cannot rule out that there might be cases where there is no single proposition whose truth is disputed by participants in the conversation. Although true, this observation is not enough to reject contextualism, for at least two reasons: (1) there may be good explanations of the appearance of p-disagreement and (2) there may be other kinds of disagreement available to account for the sense of disagreement.

2.5 Explaining the appearance of p-disagreement

I've said that the problem of disagreement is the problem that '*A* ought to *X*' and '*A* ought not to *X*' appear to express incompatible propositions, but that contextualism does not guarantee that this is always so. One way to defend contextualism would then be to explain why the appearance might arise in cases where there is no contradiction.

Before I do this, however, it is worth asking who is supposed to be the subject of the appearance and what it is supposed to be. Is it supposed to be a more or less explicit *belief* that we are literally contradicting each other in cases where there is no contradiction? It does not strike me as plausible that ordinary speakers make finegrained distinctions between different types of disagreement. It may be unclear exactly what the content is of the thought 'I disagree with *S*'. Perhaps this is not p-

disagreement in the first place. Ordinary speakers may feel that having different standards (with respect to the same act) is a way of disagreeing.

However, even if we assume that ordinary speakers have the sense of expressing contradictory propositions in cases where they don't, there may be ways to explain this appearance. Stephen Finlay notes that normative debate ordinarily proceeds on the assumption that at least some relevant standards are shared ((2008), p. 353, 356). This assumption is usually justified because, as a matter of fact, most people do share values. He rightly notes how rare it is to meet someone who is morally alien. Furthermore, moral disagreements are (often) based on disagreements about the nonmoral facts. People may harbour different (and to some extent inchoate) beliefs about the effects of certain actions. This makes it harder to be sure that the difference is in standards, and could explain the sense of contradiction.

In order for this response to solve the problem, it had better be plausible that people *always* (or nearly always) make this assumption. Robert Streiffer argues that it is not because of the extent and people's awareness of moral disagreement (Streiffer (2003), pp. 14-15).

Björnsson and Finlay retort that an assumption of common standards

> 'could be reasonable even in the face of extensive and intractable disagreement; moral standards might reasonably be thought to be highly abstract and difficult to apply -perhaps even indeterminate- as would be the case if Kantians or Utilitarians were correct about the principles of morality.' (Björnsson & Finlay (forthcoming))

They further add that:

> 'if our sense of disagreement depends on the assumption of a common standard, that could explain why it is less clear when the standards are *strikingly* different. If we consider the moral beliefs of (e.g.) a New Guinean headhunter prior to 'civilized' contact instead of the moral beliefs of a 19th century American [...], it is arguably much less obvious that we have an intuition that those moral beliefs contradict our own.' (Ibid.)

So Björnsson and Finlay think the appearance of contradiction can be explained by the fact that people tend to assume some relevant commonality in standards, to which I would add the difficulty of making sure that we agree on all the facts.

But I'm not sure that this explanation is sufficiently plausible. What is plausible is that, in discussion, moral claims are justified by means of nonmoral features of the world. Participants do presuppose that these features carry weight with others. (It is also plausible that if someone is not at all responsive to any such feature, we might cease to engage.) But it may become apparent that features don't carry the *same* weight with others, as seems to be the case in many political disputes (e.g. the protection of individual liberties versus the protection of the people from harm). It seems implausible that we are never aware of such differences.

So Björnsson and Finlay's idea might derive its plausibility from a narrow understanding of a standard as something which determines what nonmoral features have moral *relevance*. It's true that most people share standards in this sense. But it's less plausible that people share standards if by 'standard' we mean a principle which not only determines the relevance of features, but also their weight. It seems that Björnsson and Finlay's explanation requires people to assume that everything carries the same weight with everyone. I'm not convinced this is the case. However, I'm also not convinced that the problem of disagreement is quite as it was formulated.

2.6 Explaining the appearance of non-p-disagreement

The previous discussion was based on the assumption that it appears to ordinary speakers as if relevant moral statements always (or mostly) express contradictory propositions. I've indicated that I am somewhat sceptical of this. The sense of "contradiction" may also arise from an awareness that people express commitment to incompatible standards. If so, the problem of disagreement needs to be reformulated: the problem then becomes to explain why speakers would have a sense of disagreement at all in cases where no contradictory propositions are expressed. But this needn't be a sense of *p*-disagreement. This makes our task significantly easier.

Even though speakers with different standards may not be disagreeing in the sense of asserting contradictory propositions, they may be disagreeing in some other sense. For example, they may have a disagreement in *attitude*. Having a disagreement in attitude amounts to taking opposing attitudes towards the same thing (e.g.

respectively approving and disapproving of it). Given that to subscribe to a moral standard is (*inter alia* and *ceteris paribus*) to approve of conformity to it, this kind of disagreement is available to the contextualist.

I once heard the worry expressed whether disagreement in attitude is disagreement at all. It was backed up by the following example: If I prefer a biscuit and you prefer a chocolate, in what sense are we disagreeing? We can of course define technical notions of disagreement, but they don't give us disagreement any more than defining 'unicorns' as horses gives us unicorns.

However, this objection does not seem very strong, since what we respectively prefer in this example is that *I* do something and that *you* do something. We don't have a preference with respect to the same act.[22] So if people feel that they disagree with someone about *X* even if they know that the person holds very different standards, then they are right: they disagree in what they respectively prefer with respect to *X*.

[22] If required, we may be able to answer the objection further by circumscribing disagreement in attitude more closely. It does not sound implausible to say that we disagree if I approve of *X* and you do not. So a plausible notion of disagreement in attitude may involve only certain *types* of attitude. There might be a difference between the attitude of approval and that of preferring something. The notion of preferring can be analysed in terms of desiring something more than something else. One might argue that, although approval also involves something like this, it is not exhausted by it. I don't approve of whatever I desire most.

2.7 The point of moral discourse and further problems of disagreement

Of course, even if the sense of disagreement can be explained by means of a disagreement in attitudes or a clash of standards, we still have to explain some linguistic phenomena. For example, it seems that if your judgment '*A* ought morally to *X*' is true relative to your standards, I should not regard it as false if *my* standards disallow it. If we do that nonetheless, it is evidence against a contextualist treatment of moral discourse.

Notice that if people tend to presuppose that moral standards are shared, it's not surprising that I regard your judgment as false. For if the standards are shared, one of us must be wrong about their application.

If people don't make an assumption of shared standards, then the contextualist has yet to answer this and some related objections. But I think they can all be answered in the light of the following hypothesis: that there is a reason why people tend to assess the correctness of moral judgments from their own normative perspective (their own standards). The reason is that moral thought has a function not confined to learning facts about the world. It is essentially concerned with protecting and promoting certain values. This hypothesis is advanced (in slightly different forms) by Gibbard (1990), Francén (2008) and Björnsson & Finlay (forthcoming).[23] It

[23] Francén seems to think that people assess the correctness of other people's moral judgments relative to their own standards because it 'is important to reach a (common) conclusion about what to do, not to discuss what to do on this or that morality' ((2008), p. 123). However, if all we were interested in was to reach a *common* conclusion, we might as well agree to do whatever the other person thinks. So I

explains why we focus on our own standards when we assess other people's moral judgments for correctness:

> 'Part of what it is to engage in a *moral* (or categorically normative) practice is to subscribe to a particular standard, to the exclusion of any rival, as determinative of what to do, both for one's own conduct and others'. Characteristically, what fundamentally matters to us when we make moral judgments is agents' conformity with the moral standards to which we ourselves subscribe. Our interest in the truth of (standard-relative) propositions is therefore derivative upon this concern, and the truth of propositions relativized to *other* standards is irrelevant to us. To address the question of whether Huck was right in thinking that not telling on Jim violates Huck's standards, *when the issue is how to act in circumstances like Huck's*, is therefore a perverse fixation on truth for truth's sake, in neglect of what is relevant to the purpose of moral thinking and moral discourse.
>
> What would be relevant to our concern's, however, is to assess the moral judgments of others like Huck *as if* they had been made in relation to *our* standards and to express agreement or disagreement with the resulting ought-propositions, which *are* relevant to us.' (Björnsson & Finlay (forthcoming))

---

think that Björnsson and Finlay's way of putting it is better: we want to promote and/or protect certain values. Which values? The ones that *we* hold dear.

It is plausible that, in moral discourse, our interest in truth is secondary to our interest in the protection and promotion of (our own) values. This explains why the proposition I assess for truth or falsity is not (necessarily) the one expressed by my interlocutor. It is (sometimes) one closely related to it. When John says '*A* ought to *X*' I don't check whether *X*-ing is required by John's standards, but I check whether it is required by mine.

This hypothesis (the Point Hypothesis) provide the resources to answer all the standard objections to contextualist analyses of normative language. We've already seen how they make sense of cases like the following (where John and Mary hold different standards):

A John: Female infibulation is wrong.

Mary: That's false. It is not wrong at all.[24]

Berit Brogaard claims that the contextualist has to say that Mary is linguistically confused in A, like she is in:

B: John: I am hungry.

[24] This example is based on Brogaard (2008), p. 387. Notice how rare it is for people to say something like 'This moral view is *false*' (I challenge anyone to find such a statement by non-philosophers). But this is probably irrelevant. It suffices that Mary is willing to assert what appears to be the negation of John's statement.

Mary: I disagree [that's false]. I am not hungry at all.

(Brogaard (2008), p. 387)

However, if our hypothesis is correct about the point of moral discourse, then Mary wouldn't be confused. She would be confused precisely if she said: 'I agree. Female infibulation is not wrong (from John's perspective)'. She would have misunderstood the point of moral discourse.

Brogaard also claims that the contextualist has a problem with the fact that people retract earlier 'ought' claims when they change their minds:

> 'Suppose John has recently visited a country where female infibulation is routinely done. Upon his return he has completely changed his mind about it. The following exchange then takes place between him and Mary:
>
> C. John: Female infibulation is not wrong at all.
>
> Mary: But a few months ago you said that it was wrong.
>
> John: I know. I was mistaken.
>
> If John means different things by 'wrong' on the two occasions, as contextualism says, one might expect John to respond differently, for instance, with 'Yes, but I was right back then as well; it's just that my moral standards were different back then', or 'Yes, but I was right back then as well; I just

didn't mean the same thing by "is wrong"'. But this is not how people would normally respond.' (Ibid., p. 388)

However, if people evaluate 'ought' claims from their own normative perspective, it is not hard to see why John would say that he was mistaken. He was mistaken in the sense that female infibulation is not forbidden by his *current* standards.[25]

It may seem, however, that the Point Hypothesis makes trouble for propositional attitude reports (ibid., pp. 388-389). Isn't the contextualist required to say that if Mary reports John's beliefs about infibulation, she does so incorrectly? Suppose that Mary says: 'John believes that female infibulation is not wrong'. If Mary's use of 'wrong' is relative to *her* standards, then the proposition which she claims that John believes is different from the one that he actually believes. For 'wrong' as used by John is relative to *his* standards.

Björnsson and Finlay (forthcoming) plausibly respond that 'in attributing beliefs we are interested in understanding the subject's attitudinal state'. In other

[25] Notice, though, how underdescribed Brogaard's example is. John visited some country where female infibulation is routinely done. Upon his return he has completely changed his mind about its wrongness. But *why* is that? It seems quite possible that witnessing the process, or talking to people who underwent it, or learning about the social stigma attached to not having it done changed his beliefs about its benefits and harms. If so, he needn't have changed his standards. But the example is effective only if John plausibly changed his standards, rather than his nonmoral beliefs.

words: 'to believe' is an operator that shifts the circumstance of evaluation of 'is wrong', namely to the standards held by the subject.

2.8 Final thoughts

I think these responses are satisfactory. But this thesis is about normative language in general, not just moral language. Although it may be true that the point of *moral* discourse is to protect and promote values, is this plausible in other cases too?

I don't see why not, although there might be differences in how much people *care* to protect and promote the values in the relevant domains. The less they care, the more likely they are to refrain from saying things like: 'You're wrong!'. Art critics and food critics might care enough to generate all the patterns that we see in moral discourse.

To support this idea, consider the BBC's motoring progam *Top Gear*. On this show (or at least in some series), there is a thing called *The Cool Wall*. It is divided into three sections: subzero, cool and uncool. The presenters' task is to pin up photographs of cars onto this wall in accordance with their coolness. They often disagree about which cars are cool and argue for their views.

Few people believe that these are disputes over objective matters of fact. Most presume that cars are cool only relative to standards. But the disputes show all the features of moral arguments: reasons are provided and rival views declared to be wrong (even if it's clear that co-presenters hold completely different standards).

Top Gear is admittedly humorous. But the source of humour is not, in this case, the linguistic patterns of assent and dissent. They are perfectly familiar, despite the subject matter's subjectivity. This supports both the idea that the degree to which we care will determine what we are prepared to say, and contextualism about normative words in general.

It is probably also true that people tend to "go relativistic" much quicker in nonmoral cases than they would in moral ones. That is to say: once it is clear that others cannot be persuaded of the coolness of a car, people are more likely to give up, making explicit the relativization: 'Oh well, *I* like it anyway'. But it is a mistake to take the fact (if it is a fact) that it happens less often in morals as strong evidence for a difference in semantics. For, firstly: there is presumably some kind of relation between normative terms like 'good', 'right' and 'ought' as used in moral and nonmoral contexts. If standard-relativity is likely for many of these uses, there is reason to think that it is the same with moral ones. Secondly: as long as explicit relativization occurs at least *sometimes* in moral discourse, both contextualists and invariantists have some explaining to do. The contextualist has to explain why moral language often appears nonrelative, but the invariantist has to explain why it doesn't always. Thirdly: my argument for contextualism in section 2.3 is not refuted by the considerations discussed in section 2.7.

Does explicit relativization occur in moral discourse? It might well occur in contexts where more peripheral moral problems are discussed. By 'peripheral moral problems' I mean problems not related to (human) life and death, basic goods, justice, etc. (areas which we tend to feel most strongly about). Explicit relativization might

occur in judgments about whether to vote at elections, whether one is allowed to eat meat, whether one has to inform the neighbours before one has a party (etc.). Explicit relativization may occur in more central cases too, but I don't claim to know this. It is an empirical question that philosophers are too keen to pronounce on from their armchairs.

2.9 Conclusion

In this chapter, I have presented an argument for contextualism about 'ought': it explains why '*A* ought morally to *X*' and '*A* ought prudentially to not-*X*' can be simultaneously true. I also considered some objections against contextualism. The upshot of my discussion is that the purpose of normative language allows us to explain why we don't always evaluate the same proposition in normative discussion. However, the extent to which this happens should not be exaggerated: people often refer to the same system of standards (or the same subset of a system). Furthermore, contextualism about 'ought' allows great semantic unification: it provides a uniform treatment of 'ought' as used in moral and nonmoral contexts.

In the rest of this thesis, I will assess the merits of various forms of contextualism. Prominent among these are Humean approaches to normative language. They are the subject of the next chapter.

Chapter 3. Humean theories of normative language

3.1 Introduction

In the previous chapter, I argued that contextualism about normative language is plausible. Contextualism about 'ought' is the view that the truth conditions of 'ought' claims vary with the context of utterance. I argued that the premise that it can be simultaneously true that *A* ought morally to *X* but prudentially to not-*X* places a restriction on the semantics of 'ought'.

One way to allow this is to say that '*A* ought to *X*' can be analysed as 'The balance of reasons favours that *A X*-es'. In that case, the simultaneous truth of (1) '*A* ought morally to *X*' and (2) '*A* ought prudentially to not-*X*' can be secured by the following hypothesis: (1) means that the balance of moral reasons favours that *A X*-es and (2) means that the balance of prudential reasons favours that *A X*-es.

A view that takes this line will have to say that the concept of a reason is conceptually prior to the concept of ought. Unless the view is objectivist (and thus not relevant to this thesis), there have to be mind-dependent or institutional facts which make it the case that something is a reason. By far the most prominent subjectivist view about reasons is Humeanism. I will consider any view Humean according to which a thing acquires the reason-status in virtue of the *agent's* motivational states.

I will discuss two versions of Humeanism: Gilbert Harman's and Mark Schroeder's. My justification for this is twofold: first, they purport to satisfy the requirements we have placed on theories of normative language: they are intended as

subjectivist, truth conditional views that are not error-theoretic. Second, Harman (at least in early work) and Schroeder are fairly explicit about the semantics of normative discourse. But I will argue that both views are problematic.

3.2 Harman's view

Harman is usually seen as a particular type of indexical relativist about moral language: someone who believes that moral statements implicitly refer to a moral framework. This makes sense because he often talks like this:

> 'Just as it is indeterminate whether a dog is large, period, apart from any relation to a comparison class, so too […] it is indeterminate whether an action is wrong, period, apart from any relation to [a moral framework].' ((2000a), p. 4)

In (1996), he describes his position by saying that '*X* is wrong' does not have a complete truth condition independent of a moral framework:

> 'For the purposes of assigning truth conditions, a judgment of the form, *it would be morally wrong of P to D*, has to be understood as elliptical for a judgment of the form, *in relation to moral framework M, it would be morally wrong of P to D*. Similarly for other moral judgments.' ((1996), p. 4)

This too sounds like a kind of relativism. The next chapter will be devoted to such views. In this chapter, I concentrate on Harman's Humean ideas.

In his early articles, Harman stresses the relation between 'ought' statements and statements about reasons. 'Ought' statements are presented as equivalent to statements about reasons:

> '[I]n each of its uses, the word 'ought' is used to speak of things for which someone has reasons. The epistemic 'ought', for example, is used to speak of things that there are reasons to expect, as when we say that the train ought to be here soon. The evaluative 'ought' is used to speak of what there are reasons to hope or wish for or take some other positive attitude towards, as when we say that there ought to be more love in the world or that a knife ought to be sharp. The simple 'ought' of rationality is used to speak of something for which an agent has reasons of any sort to do, as when we say that a burglar ought to wear gloves. And the moral 'ought' is used to speak of things an agent has moral reasons to do, as when we say that a burglar ought to reform and go straight.' ((2000c), p. 42)

Harman distinguishes between the *evaluative* 'ought' and what he calls the *moral* 'ought' in the passage quoted. Evaluative 'ought's are equivalent to statements containing words like 'good' and 'bad'. Moral 'ought' statements (or 'directives' in Wiggins's terminology of (1998a), pp. 95-96) are equivalent to statements containing words like 'right' and 'wrong'. The judgment that *A* ought morally to *X* is equivalent

to the judgment that it is morally right of *A* to *X*, and the judgment that *A* morally ought not to *X* is equivalent to the judgment that it is morally wrong of *A* to *X*. The difference between evaluative and directive (moral) 'ought' statements is that only the latter are equivalent to statements to the effect that the agent of whom it is said that s/he ought to *X* has (moral) reasons to *X*. This is not true for evaluative 'ought' statements, even though they are equivalent to statements about reasons in a different sense (to which I will return).

What is it for an agent to have a reason to do something? In (2000d), we find the following characterization of 'sufficient reason':

> 'Now, presumably, someone has a sufficient reason to do something if and only if there is warranted reasoning that person could do that would lead him or her to decide to do that thing.' ((2000d), p. 86)

In (2000c), Harman claims that someone has no sufficient reason to accept some moral principle *P* if and only if that person can fail to accept *P* without being in any relevant way irrational, stupid, or uninformed ((2000c), p. 43).[26] From these characterizations, it is clear that Harman subscribes to something like Bernard Williams's conception of reasons as considerations that could potentially motivate an

[26] In (1996), p. 46, Harman makes the qualification that one can have sufficient reason for several incompatible acts. In that case, one needn't be irrational, stupid or uninformed in doing one rather than the other. So henceforth this type of formulation is to be understood against the background assumption that the subject has no sufficient reasons for any incompatible action.

agent (Williams (1988a)). Harman allows for a certain idealization in that he allows for cases where an agent has reasons that are not recognized by the agent, but only insofar as this lack of recognition is explained by 'ignorance of relevant (nonmoral) facts, a failure to reason something through, or some sort of (nonmoral) mental defect like irrationality, stupidity, confusion or mental illness' ((2000b), p. 30). Harman's stress on nonmoral failings reflects the fact that he does not allow for objective moral facts, the failure to recognize which could, for the objectivist, constitute a failure of rationality all by itself.

Harman's conception of reasons puts certain restrictions on the correctness of directive 'ought' statements. If in making such a statement a speaker implies (among other things) that the agent judged has a sufficient reason to *X*, then it has to be true of him or her that absent akrasia, nonmoral failures of reasoning, mental health (etc.) the agent will decide to *X*. This is the case only if the agent has certain motivational attitudes *M* which (can) lead him/her to recognize some consideration *C* as a reason to *X*. Because the speaker also recognizes *C* as a reason to *X*, directives involve a presupposition of shared attitudes (this is why Harman calls directive 'ought' statements 'inner moral judgments'; they are made from within a certain evaluative stance ((2000a), p. 4)). In the same work, Harman puts it as follows:

> 'As used by the relevant sort of speaker, 'Ought (*A*, *X*, *C*, *M*)' means roughly this: given that *A* has motivational attitudes *M* and given *C*, *X* is the course of action for *A* that is supported by the best reasons. In judgements using this sense of 'ought', *C* and *M* are often not explicitly mentioned but are indicated

by the context of utterance. Normally, when that happens, *C* will be 'all things considered' and *M* will be attitudes that are shared by the speaker and audience.' ((2000a), p. 10)[27]

What about evaluative judgments, judgments to the effect that some state of affairs is good or bad, or that it is good or bad that someone acts a certain way? Harman is clear that these are not equivalent to claims about reasons in the way directives are. With regard to the evaluation that it was evil of *A* to *X*, Harman even claims the opposite:

> 'If someone *S* says that *A* (morally) ought to do *X*, *S* implies that *A* has reasons to do *X* and *S* endorses those reasons. If *S* says that *B* was evil in what *B* did, *S* does not imply that the reasons *S* would endorse for not doing what *B* did were reasons for *B* not to do that thing. In fact, *S* implies that such reasons were not reasons for *B*.' ((2000a), p. 8)

Yet Harman does say (at least in other articles) that evaluative judgments are equivalent to statement about reasons in some sense. Remember that he says that

[27] I have replaced the action-variable '*D*' with '*X*' in this and all subsequent quotations from Harman's early work.

> 'The evaluative 'ought' is used to speak of what there are reasons to hope or wish for or take some other positive attitude towards, as when we say that there ought to be more love in the world or that a knife ought to be sharp.' ((2000c), p. 42)

In the same essay, we learn that Harman does not make much of the distinction between reasons that *one has* and reasons that *are there*. This distinction, then, does not help to answer the question in what sense evaluative judgments are statements about reasons.[28] It seems reasonable to suppose that Harman would say that someone who judges that there ought to be more love in the world judges at least that s/he him/herself has reason to hope for this, and perhaps also that the audience has such reason. But when an evaluative judgment is made about some particular agent (for example, that it was bad *of A* to *X*), it is not implied that the agent has such reason too. After all, Harman believes that unlike directives, evaluative 'ought' statements can be properly made even about people with very different motivations. Harman argues that we cannot make the inner moral judgment that Hitler ought not to have ordered the extermination of the Jews, because Hitler had no reason to refrain from doing so (and inner 'ought' statements imply that the agent has such reason) ((2000a), p. 7). The sense in which we 'cannot' make this judgment is presumably that the judgment would be false. However, we can truly say that it was bad of Hitler to

[28] Harman only allows a distinction between the reasons someone has and the reasons that are there based on a difference in the amount of effort necessary to get the agent to realize that s/he has some reason ((2000c), p. 44).

order the extermination of the Jews, because in doing so we don't imply that Hitler had reason to refrain. The conditions under which such a statement is true depend on who is supposed to have reason to hope or wish for a state of affairs in which Hitler did not give this order. Is it merely the speaker or perhaps also the people s/he addresses? The latter is suggested by Harman in (1996), where he says the following:

> 'In saying, "This rain is bad," Tom means (roughly) that it is bad for himself and his audience; not just that it is bad for himself. The remark, "This rain is bad" is not normally equivalent to "This rain is bad for me." When Tom tells someone else that the rain is bad, he means (roughly) that it is bad in relation to certain goals, purposes, aims, or values that he takes himself to share with his audience.' ((1996), p. 15)

So (early) Harman thinks that moral statements are equivalent in meaning to statements about reasons. In the case of directive 'ought's, these reasons are meant to apply to both speaker, audience and agent. In the case of evaluative 'ought's, they are meant to apply to speaker and audience only.

However, Harman also claims that unrelativized moral statements are somehow incomplete in respect of truth conditions. Let me repeat the quote already given:

> 'Just as it is indeterminate whether a dog is large, period, apart from any relation to a comparison class, so too […] it is indeterminate whether an

> action is wrong, period, apart from any relation to [a moral framework].' ((2000a), p. 4)

But the hypothesis that moral statements are incomplete in respect of truth conditions is incompatible with the hypothesis that they are equivalent in meaning to statements about reasons. For suppose that Harman is right that the concept of a sufficient reason is as follows:

> Someone has a sufficient reason to *X* if and only if absent akrasia, nonmoral failures of reasoning, neglecting facts, mental illness (etc.) the agent will decide to *X*.

If 'It is right for *A* to *X*' is equivalent in meaning to '*A* has a sufficient reason to *X*', then it has complete truth conditions on its own. '*A* has a sufficient reason to *X*' is true just in case *A* would decide to *X* absent akrasia, nonmoral failures of reasoning, neglecting facts, mental illness (etc.). There is no need to relate the statement to a moral framework first.[29]

[29] The only sense in which moral statements are true relative to a moral framework on this proposal is that the truth of a moral statement depends on the moral reasons someone has, which in turn depend on the principles, values, etc. s/he accepts.

3.3 Problems with Harman's view[30]

So on Harman's (early) view, 'ought' statements are equivalent to statements about reasons. It has the advantage of providing clear truth conditions for the claim 'The balance of reasons favours that *A X*-es'. Insofar as this is an inner 'ought' judgment, we can take it to mean that *A* has sufficient reason to *X* and not sufficient reason to do an incompatible act. This in turn is true just in case *A* would decide to *X* were s/he not subject to ignorance (lack of awareness of relevant facts), akrasia and failures of reasoning.

It is less clear what the truth conditions are of the judgment '*A* has *a* reason to *X*'. Here we have two options. We can either say that it is true just in case *A* would be motivated to *X to some extent*, or we can say that *A* has a reason to *X* just in case *A* would decide to *X* were this the only relevant consideration (and *A* not subject to ignorance, akrasia and failures of reasoning). I think neither of these options is very attractive, but we'll encounter plenty of problems even if we ignore this issue.

I believe there are three problems with Harman's view. The first relates to the insertion of akrasia (weakness of will) into the conditions under which *A* would decide to *X*. For what is akrasia? In this context, it is acting contrary to one's judgment that one ought to do something. In other words, it is doing something for which one judges that one lacks sufficient reason. So in the definition of akrasia we find the notion of sufficient reason. But in the definition of sufficient reason, we find the notion of akrasia. This circularity is troubling.

---

[30] This section has benefited from discussion with Natalja Deng.

It seems plain that the notion of sufficient reason is conceptually prior to the notion of akrasia and that the latter should be defined in terms of the former. But when we delete akrasia from Harman's definition of sufficient reason, it is no longer adequate. For in that case, the definition reads:

> *A* has sufficient reason to *X* (and not sufficient reason to do anything else) just in case *A* would decide to *X* were s/he not subject to ignorance or failures of reasoning.

But if *A* is a libertarian free agent, *A* might still decide to not-*X* even if s/he is aware of all the facts and does not make any mistakes in her reasoning about the best option. And so it wouldn't be true that *A* would decide to *X* whenever the specified conditions obtain. Surely Harman does not want his definition of sufficient reason to rule out various views about freedom of the will.

It's not clear to me how to fix the definition so as to avoid the problem. I don't think it helps to weaken the requirement by saying that an agent has a reason to *X* just in case s/he *might X* were s/he not subject to ignorance or failures of reasoning. This seems inadequate for the same reason: a libertarian free agent might decide to do anything. But s/he does not thereby have sufficient (of even some) reason to do anything.

A second worry about Harman's view is that it may not allow enough normative judgments to be true. With respect to moral judgment, Harman's view is not error-theoric because it does not condemn *all* moral judgments to falsehood. For

Harman, the truth of an inner moral judgment depends on whether *A*, the agent, shares the relevant motivations with the speaker. It is true that the agent has sufficient moral reason to *X* just in case that agent would decide to *X* were s/he not subject to ignorance, akrasia and failures of reasoning. But that is true only if the agent is relevantly similar to the speaker. That will sometimes be the case (which is why Harman's theory is not error-theoretic), but it will not always be the case. It can be false even if speaker and agent share relevant values, for the question whether someone would *X* depends not just on his or her values, but also on the *weighting* of those values. And disagreements about the latter are quite common.

Harman could say that people can still make true *evaluative* 'ought' judgments (judgment about what we, speaker and audience, have reason to hope). That may be true (although the audience may also be dissimilar). But even if it is, it does not reduce the number of false inner 'ought's. After all, it is not plausible to interpret an 'ought' judgment as evaluative whenever speaker and agent happen to be dissimilar. For the speaker may not *know* this and falsely assume that the agent is relevantly similar.

At least in the case of morals, it seems that our judgments are not sensitive to psychological information about the agent. By this I mean that people's assessment of the truth of such judgments (or their readiness to make them) is not influenced by this type of information. We may happily say there was plenty of reason for Hitler not to order the extermination of the Jews, even if we know that he lacked the relevant

commitments.[31] Derek Parfit considers it a *reductio* of Humeanism that if one lacked the relevant desires now, one would not have reason to prevent a future period of agony (*On What Matters*, chapter 3). This argument betrays a willingness to ascribe moral and prudential reasons to people irrespective of their current motivational profile.

If Harman is right, this lack of sensitivity shows up in a standing assumption that the agent shares our motivations. But, first, why would the assumption be part of the truth conditions of moral judgments, if our willingness to assert them is not affected by its truth? And second, whenever the assumption is false, the moral judgment will be false as well. So Harman's theory may not allow enough true moral judgments.

I think Harman's response to this objection would be to point out that some of our reason ascriptions do seem to depend on the agent's psychology. If Ronnie likes to dance and there is dancing at the party, then Ronnie has a reason to go to the party. If Bradley hates dancing, he does not have this reason (the example is from Schroeder (2007)). If our best explanation of this phenomenon were Humean, there would be pressure to think an error is involved in the ascription of moral reasons irrespective of the agent's motivations. I agree with this line of reasoning, but I don't think our best explanation of the phenomenon is provided by a Humean theory of reasons. As I will

[31] Notice that Harman's scepticism about a distinction between reasons that one has and reasons that are there bars him from dissociating the moral reasons that were there for Hitler from *Hitler*'s motivational profile (see also footnote 28).

show in chapters 4 and 6, it can be accommodated quite readily by other theories of reasons.

My third objection to Harman's view concerns the order of explanation. If an agent would decide to *X* under certain conditions, we can ask why. The fact *that* s/he would do something does not appear to be what its rationality or normative status consists in. Presumably, the agent would choose to do an act because it has certain properties (like being most conducive to the satisfaction of desires). Aren't these properties what explains that it is the thing to do, or what one ought to do? Harman's analysis leaves these out, unlike other forms of Humeanism.

To sum up: Harman's counterfactual definition of 'sufficient reason' appears to be inadequate. Furthermore, his view probably makes many normative judgments false. And finally, his definition of 'sufficient reason' seems to leave the most important question hanging: *why* would one choose the act? It seems that the answer to this question would reveal something more essential about the conditions under which 'ought' judgments are true.

3.4 Schroeder's view

There is a Humean theory on the market which does a better job of explaining why 'ought' statements and statements about reasons do not seem sensitive to psychological information. This is Mark Schroeder's *hypotheticalism* (2007).

Schroeder agrees that people might say that *A* has a moral reason to *X* even if *A* doesn't care about *X*-ing or its consequences. Like many other philosophers, he

believes that a commitment to the universal validity of moral reasons is implicit in moral language. Yet Schroeder claims that all reason statements do implicitly refer to a class of people, as follows:

> 'For *R* to be a reason to do *X* is for *R* to be an agent-relational reason for [us] to do *X*.' ((2007), p. 18)[32]

By an 'agent-relational' reason to *X*, Schroeder means a reason that an agent has in virtue of the fact that s/he has desires which are furthered by *X*-ing. Depending on context, '[us]' comprises more or fewer people.[33] In the case of *moral* reasons, '[us]' is everyone, or everyone possible. This is supposed to account for the sense that moral reasons universally apply.

It may seem that hypotheticalism makes all moral statements false: after all, not everyone seems to have desires which are furthered by going for the moral course of action. But Schroeder's theory is designed to provide almost anyone with such reasons.

Schroeder notes that it is compatible with the view that reasons depend on desires that reasons don't depend on any particular desire. It needn't be the case that if *A* lacks desire D_1, s/he doesn't have a reason to *X*. For the reason to *X* may also be grounded in D_2 as well, which *A* does have. If a reason *R* to *X* can be grounded in

[32] Schroeder uses '*A*' as a variable for actions and '*X*' as a variable for agents. I have reversed this for the sake of uniformity and to avoid confusion.

[33] This could be just one person, in a limiting case.

several different desires, then the following is true: there is no specific desire D_n such that if *A* lacked it, it would follow that *R* isn't a reason for him or her to *X*.

The fact that reasons can be grounded in several desires is clear from an example of Schroeder's:

> 'Susan wants some coffee. So the fact that there is coffee in the lounge is a reason for her to go there. But perhaps this is doubly determined. For example, perhaps philosophers tend to congregate and talk shop when there is coffee. And perhaps Susan wants to talk shop about some idea she's recently had. This should also explain why the fact that there is coffee in the lounge is a reason for Susan to go there. If this is so, then this is a reason for Susan twice over. The fact that there is coffee in the lounge would be a reason for Susan to go there even if she didn't want a cup of coffee, and it would be a reason for her to go there even if she didn't want to talk shop. So there is no single desire on which it depends.' (Ibid., pp. 108-109)

Of course, this is still not enough to establish that there are any reasons that anyone at all must have. For even if there is not one single desire on which Susan's reason to go to the lounge depends, it depends on a small number which she happens to have. People who have neither a desire for coffee, nor for talking philosophy don't have a reason to go to the lounge in Susan's building. Few philosophers hold that the objectivity of moral reasons would be sufficiently guaranteed if it turned out that they depend on [either a desire for *p*, or a desire for *q*, or ...], where this is a relatively short

disjunction. What Schroeder proposes is that some reasons (including moral ones) depend on almost any desire:

> 'Hypotheticalism's favored proposal for how there could be genuinely agent-neutral reasons is therefore that genuinely agent-neutral reasons are massively overdetermined. They are reasons for anyone, no matter what she desires, simply because they can be explained by any (or virtually any) possible desire.' (Ibid., p. 109)

Schroeder, then, believes that if there are reasons that are explained by any possible desire, they will be reasons for anyone, no matter what s/he desires, as long as there is *something* s/he desires. So genuinely agent-neutral reasons depend on desires, but in a very minimal sense. For, as Schroeder claims, it isn't plausible that beings without desires at all are agents in the first place. And it *is* plausible that agenthood is a necessary condition on having reasons.

Next, Schroeder gives the following definition of a reason:

> '**Reason** For *R* to be a reason for *A* to do *X* is for there to be some *p* such that *A* has a desire whose object is *p*, and the truth of *R* is part of what explains why *A*'s doing *X* promotes *p*.' (Ibid., p. 59)

So Schroeder conceives of reasons as true propositions which play a role in explaining why an act promotes the object of desire in virtue of which they are reasons in the first

place. This definition is more satisfying than Harman's characterization (of sufficient reason), since it identifies *why* a suitably ideal agent might choose to *X*. S/he might choose to *X* because *X*-ing is conducive to *p*.

Now, the basic form that Schroeder's explanation takes (of the fact that certain reasons may be had in virtue of any possible desire) is illustrated by the example of a reason to believe an arbitrary proposition only if it is true. Here is how the explanation goes:

> 'Being in error about [some arbitrary proposition] might lead to being in error about other things, such that being in error about them might lead to being in error about other things, and so on until something might lead to Mary having trouble getting new shoes. If this is right, then for any proposition, Mary's desire to get a new pair of shoes will serve to explain why there is a reason for Mary to believe it only if it is true.' (Ibid., p. 114)

So, the basic idea is that for any proposition, there is a chance that if it is true, believing it could come to bear on Mary's ability or success at buying shoes (being in error might lead to having trouble buying new shoes). Therefore Mary has a reason to believe any arbitrary proposition only if it is true in virtue of her desire to buy new shoes.

I take it that the proposition whose truth plays a role in an explanation of why doing *X* promotes *p* has to be true in the actual world. That the fairies exist cannot be a reason to do something if the fairies don't exist. But it would be odd if all that the

proposition has to do is play a role in an explanation of why doing *X* would promote *p* in some possible world (for example, a world where the laws of nature are completely different). Doing *X* has to promote *p* in the actual world in order for you to have reason to do *X* in the actual world. So I will also assume that the proposition that is true in the actual world has to play a role in an explanation of why doing *X* would promote *p* in the actual world.

Partly for independent reasons, Schroeder takes what he thinks is a very permissive stance on what it is for an action to promote an object of desire:

> '*A*'s doing *X* promotes *p* just in case it increases the likelihood of *p* relative to some baseline. And the baseline, I suggest, is fixed by the likelihood of *p* conditional on *A*'s doing nothing - conditional on the status quo.' (Ibid., p. 113)

The idea behind making the likelihood of *p* relative to doing nothing is to ensure that it is easy to generate reasons. As long as it is slightly more likely that *p* will occur if *A X*-es than if s/he does nothing, *A* has a reason to *X*. And the easier it is to generate reasons, the more likely it is that anyone will have moral reasons.

Like Harman, Schroeder analyses 'ought' statements in terms of reasons.[34] But the difference between Schroeder and Harman is that Schroeder's theory makes it

[34] Schroeder says that '*A* ought to *X*' is true just in case the set of all the reasons for *A* to *X* outweighs the set of all the reasons for *A* not to *X* ((2007), p. 130).

easy for people to have reasons in virtue of their desires. So speakers can hold on to 'ought' and 'reason' statements even if the agent lacks desires that are obviously furthered by the action. If Schroeder is correct, it suffices that the agent has some desire or other. This may provide an answer to the worry about sensitivity. The worry was that since 'ought' statements don't seem sensitive to psychological information about the agent, there seemed insufficient reason to build the agent's desires into their truth conditions. But if reasons are grounded in almost any possible desire, then it is quite natural that 'ought' statements aren't sensitive to the particular desires that people have. However, there are three sources of concern.

3.5 Problems with Schroeder's view

The first problem is related to Schroeder's claim that moral reasons can be grounded in almost any possible desire. This can be questioned.

We have seen that Schroeder's strategy in arguing for agent-neutral reasons is to have a very permissive criterion of what it is for an action to promote an object of desire. He says that *A*'s doing *X* promotes *p* just in case it increases the likelihood of *p* relative to doing nothing or the status quo. But this definition may not generate reasons as easily as Schroeder hopes. First, we need to get clear on what 'doing nothing' is supposed to be. It can hardly be sitting absolutely still. For sitting absolutely still is also doing something for which one may have reasons. But if sitting absolutely still is the base-line relative to which all reasons are determined, then one can never have any reasons to sit absolutely still. After all, whatever the character of *p*,

the likelihood of realising it by sitting absolutely still cannot be higher than the likelihood of realising it by sitting absolutely still.

Another candidate for 'doing nothing' is anything that constitutes not doing *X*, which can be many things. For example, if *X* is giving Katie 100 pounds, then, on this interpretation, I'm not doing *X* if I give her 99 or 101. But it may well be that relative to those baselines, the likelihood of my realising *p* is not in fact raised at all. So perhaps this is not what Schroeder had in mind either.

Since Schroeder talks about 'the status quo', it may seem that 'doing nothing' is something specific that constitutes not doing *X*: something like continuing to do what one was already doing. For instance, if what one is currently doing is staring out of the window, then 'doing nothing' is not giving Katie 99 pounds, but continuing to stare out of the window. But this interpretation creates the same problem as the first: surely one can have reasons to continue doing what one was already doing. But whatever the objects of one's desires, the likelihood of realising them by continuing to do *X* cannot be higher compared to the likelihood of realising them by doing the very same.

So it is not exactly clear what 'doing nothing' amounts to. But this question is important. For Schroeder needs it to be easy to raise the likelihood of *p*. But this may be much harder relative to some baselines than others. And that is problematic for his genuinely agent-neutral reasons. In fact, not even his most convincing case seems to work: it is false that any desire whatsoever can ground a reason to believe an arbitrary proposition only if it is true.

Consider whether Mary has a reason to believe a true, but highly abstract proposition about metaphysics in virtue of her desire to buy new shoes. Suppose there is some non-zero probability that believing this proposition affects her success at buying shoes. There is presumably also a non-zero probability that not believing it will not affect her success. Suppose these probabilities are equal (which is a charitable assumption; the probability that not believing the proposition will not affect Mary's success at buying shoes may well be greater than the probability that believing the proposition will affect it). If so, and if we fix the baseline ('doing nothing') at not believing the proposition, then relative to it, the likelihood of p (buying new shoes) is not in fact raised. This means that relative to not believing the proposition, Mary does not have a reason to believe the proposition (in virtue of her desire to buy new shoes). But then relative to what does she have this reason, since not believing the proposition and believing the proposition cover all the options?

So it is much harder to find any genuinely agent-neutral reasons than Schroeder thinks. This also holds for moral reasons. Consider the fact that Katie needs help (an example of Schroeder's). Clearly, logical space contains many scenarios in which helping Katie promotes any arbitrary goal of mine. But that is not enough. The fact that Katie needs help is a reason for me to help her in virtue of some desire of mine only if there is some (actually) true proposition such that it plays a role in an explanation of why helping her raises the probability of realising that desire in the actual world. As before, it all depends on the baseline. Most desires (for ice creams, seeing my parents, helping someone other than Katie, etc.) are so far removed from anything that might be achieved by helping Katie, that the likelihood of attaining

their objects is not going to be higher if I do help her compared to doing something else instead. But whatever 'doing nothing' is supposed to mean exactly, it will presumably at least involve doing something else instead.

If it is not possible to ground moral reasons in almost any possible desire, then many moral reason statements might turn out to be false, especially if '[us]' (the class of people to whom the reasons are ascribed) is *anyone possible*. In that case, Schroeder would not do better than Harman in avoiding large-scale falsehood (in fact, he would do worse).

A second problem with Schroeder's view is that some statements about reasons *do* seem sensitive to the particular desires of people (that is after all the starting point of Humeanism). Ronnie has a reason to go to the party because he wants to dance. Should it turn out that Ronnie lost the desire, we would say he lost the reason too. But hypotheticalism predicts that the fact that there is dancing at the party is almost certainly a reason for him (or anyone) to go, whether or not he wants to dance. For there is almost certainly a baseline relative to which the dancing at the party plays a role in explaining why going to the party raises the likelihood of Ronnie's realizing some object of desire of his.

Schroeder claims that our reluctance to ascribe the reason to anyone is explained by the fact that we ordinarily say that *A* has a reason to *X* only if that reason is relatively *weighty*. So he would say that we'd retract the statement about Ronnie's reason in case he loses the desire to dance because without it, the reason would be very weak. This would be understandable if the weight of a reason were a function of the strength of the desire which grounds it, and the extent to which the act promotes (is

likely to promote) the object of desire. But Schroeder denies this theory, which he calls *proportionalism*.

So whatever Schroeder's alternative is going to look like, it has to explain two things: (a) why the absence or presence of some desires (like Ronnie's desire to dance) make such a great difference to the weight of reasons, and (b) why some reasons (including moral ones) have great weight for all people, even if grounded in desires which are only indirectly related to what the action is likely to secure. This is important because we want to avoid large-scale error in moral language. To do so, we need to preserve the truth of moral 'ought' statements like: 'John ought (morally) not to torture puppies'.[35] According to Schroeder, '*A* ought to *X*' is true just in case the set of all the reasons for *A* to *X* outweighs the set of all the reasons for *A* not to *X* (ibid., p. 130). In other words: 'John ought morally not to torture puppies' is true just in case the balance of moral reasons favours that John does not torture puppies. In order for this statement to be true, the balance of moral reasons that are grounded in John's and everyone else's desires must fall out against John's torturing puppies. I will argue that Schroeder does not succeed in showing this.

3.6 Schroeder's alternative to proportionalism

[35] Although we presumably also want to preserve the truth (at least often) of the judgment that John ought not to torture puppies *all things considered*.

Schroeder believes that *if* it can be shown that anyone has moral reasons in virtue of (almost) any possible desire, then those reasons will be weighty for all people. The basic idea underlying this belief is the following: the weight of a reason depends on reasons to place weight on reasons. Most reasons *not* to place weight on moral reasons are going to be reasons of the wrong kind, so they don't count:

> 'If Ryan can't stand Katie, for example, Ryan may have abundant reasons to place less weight on this reason. But those reasons aren't relevant to its *weight*, because they won't be of the right kind. A reason has a certain weight just in case it is *correct* to place that much weight on it. And correctness is determined by reasons of the right kind.' (Ibid., p. 142)

Earlier in the book, we learn what makes a reason one of the right kind:

> 'The right kind of reasons involved in any activity are the ones that the people involved in that activity have, *because* they are engaged in that activity. So, for example, there are correct and incorrect moves to make in chess. The incorrect moves are ruled out, I think, by reasons to follow the rules of the game. Who has those reasons? Anyone who is playing chess.' (Ibid., p. 135)

So, apparently, in order to find out what reasons are of the right kind to confer or detract weight, one thing we should ask is: in virtue of what activity does Ryan have the reason to help Katie? Now this is a peculiar question. According to

Schroeder himself, Ryan has the reason in virtue of (virtually) any possible desire. But clearly I don't have all of my desires in virtue of engagement in activities. Or even if I do, then not in virtue of engagement in well-defined activities like chess. In virtue of what activity do I have a desire (and thus presumably a reason) to be a professional musician? The activity of making music? The activity of contemplating my future? Both? How do they limit the number of reasons to place weight on my reasons to act so as to become a professional musician? None of this is very clear.

Perhaps these problems can be avoided by making a distinction: activities are relevant only to reasons *(not) to place weight on* reasons. So even though my reasons to play chess or become a professional musician do not stem from being engaged in an activity, the relevance of a reason to place weight on certain reasons does stem from the kind of activity I'm engaged in. But when I have reasons both to decrease and increase the tempo of my song, how does the fact that I am engaged in the activity of making music determine what reasons to place weight on? It seems right that there is something odd about making the tempo of my song depend on the state of my hair. That is, arguably, a reason of the wrong kind to place weight on either reason. So reasons to place weight on my respective reasons to decrease and increase the tempo ought to be reasons that are relevant to *musical* aspects of the song.

This is fine as far as it goes. But how is it supposed to help explain why moral reasons are weighty for everyone? Let me quote a bit from Schroeder:

> 'A reason has a certain weight just in case it is correct to place that much weight on it. And correctness is determined by reasons of the right kind.

> [T]hat means that they must be reasons that everyone who is placing weight on reasons has, in virtue of being someone who is placing weight on reasons. But the activity of placing weight on reasons just is the activity of deciding what to do. So it is simply the activity that every agent is engaged in. So the right kind of reasons with respect to the correctness on placing weight on reasons are precisely the class of *agent-neutral* reasons. It follows that Ryan's idiosyncractic reasons to place less weight on his reason to help Katie are irrelevant, the wrong kind of reason to determine its weight.' (Ibid., p. 142)

Now what is going on here? I think the argument can be represented as follows:

(1) What reasons (to place weight on reasons) are of the right kind is determined by the activity one is engaged in.
(2) The activity one is engaged in in deciding how much weight to assign to reasons is deliberation.
(3) Since deliberation is what *every* agent is engaged in, reasons to place weight on reasons must be agent-neutral (that is, everyone must share them).
(4) Therefore, reasons that not everybody shares cannot be reasons to place weight on reasons.

I think that premise (3) in this argument is false (and doesn't follow from (1) and (2)). Any decision what to do involves deliberation. So if Schroeder is right, no

matter what one deliberates about, only agent-neutral reasons to place weight on reasons are of the right kind. But that is incredible. Suppose I deliberate about whether to have chocolate ice cream or vanilla ice cream. My reason for wanting chocolate ice cream is that I like it. My reason for wanting vanilla ice cream is that I like it. Suppose these are my only first-order reasons. In the context I am in, I have a reason to place more weight on my liking for vanilla ice cream, which consists of the fact that I had a chocolate ice cream yesterday and I like some variation. This is clearly not an agent-neutral reason, but an excellent one to place more weight on my liking for vanilla.

A general problem with Schroeder's argument is that the kind of activity one is engaged in is sensitive to description, and there may be no unique truth about what description is relevant. But different descriptions may have contradictory implications when it comes to the question what reasons are of the right kind (to place weight on reasons). For example: when deliberating about my ice cream, I am engaged in deliberation. But I am also engaged in deliberation about my ice cream. These are two different descriptions of the activity I am engaged in. The first is less specific than the last. Everyone who is deciding about anything at all is engaged in the former, but not everyone who is deciding about anything at all is engaged in the latter. Does that mean that it is *both* the case that only unconstrained agent-neutral reasons are of the right kind to place weight on reasons *and* that only reasons that are shared between people who are deciding *about an ice cream* are of the right kind?

Problems abound. In deliberation, one may ask whether to do one thing (buy an ice cream) or another (helping Katie). But it is hard to see how the activity of

deliberation could by itself furnish one with reasons either to choose the one or the other *without a precise and substantive description of the nature of deliberation* (its objective). For example: if the goal of deliberation is to realise a majority of your desires, then it is clear how being engaged in deliberation gives one a reason to place weight on reasons to do the one, rather than the other. But Schroeder's description of deliberation as 'the activity of placing weight on reasons' is much too general to make it clear how being engaged in *that* limits the number of reasons that ought to be taken into account.

This problem is aggravated by the fact that we cannot make a neat distinction between first-order reasons to *X* or reasons not to *X* and second-order reasons to place weight on reasons to *X* or reasons not to *X*. Suppose I have a reason to go and see some opera. Is the fact that I also have to finish an article a first-order reason *not* to see some opera, or a second-order reason to place less weight on my reason to see some opera? If almost anything can count as a reason to place weight on reasons, then almost nothing is going to be a reason of the wrong kind to place (less) weight on reasons in virtue of the fact that one is engaged in the activity of placing weight on reasons. But then it is unclear why Ryan's dislike of Katie is a reason of the wrong kind.

So it is not clear how Schroeder intends to restrict the number of reasons to place weight on reasons in such a way as to guarantee that the balance of reasons will fall out against torturing puppies for everyone (let alone anyone possible). It is also not clear how Schroeder intends to explain why the absence of Ronnie's desire to dance

would make such a difference to the weight of his reasons to go to the party. And there are further problems.

As we've seen, Schroeder wants to make the weight of a reason depend on the existence of defeating reasons, reasons to place less weight on some other reason. This is illustrated for epistemic reasons in the example with Tom. The weight of the reason to believe that Tom stole a book from the library is reduced by the information that Tom has an identical twin, Tim, from whom you cannot visually distinguish him. But of course, there could be a defeater for this defeating reason too: if we discovered that Tim is in Thailand, that would reduce the weight of the reason to place less weight on my original visual evidence. This could in principle go on indefinitely. If it did, there would not be a fact of the matter about a reason's weight, or at least it would be impossible to establish what it is. So Schroeder hypothesizes that there is a point where defeaters (as a matter of contingent fact) run out. When that happens, 'we can go back through the chain and determine the weight of your original visual evidence that Tom stole a book' (ibid., p. 137).

Notice that it is not very plausible that defeaters and defeaters of defeaters run out, if defeaters are reasons and reasons are come by very easily (which Schroeder needs in order to give everyone reasons not to torture). Perhaps not all reasons will be reasons of the right kind, but what *makes* a reason of the right kind is not clear (as argued in the previous section). Even if we let this pass, however, I think Schroeder's theory of weight is crucially incomplete.

Schroeder gives a recursive account of the weightier-than relation, which is supposed to help us grasp his theory of weight:

'**Weight Base** One way for set of reasons *A* to be weightier than set of reasons *B* is for *B* to be empty, but *A* non-empty.

Weight Recursion The other way for set of reasons *A* to be weightier than set of reasons *B* is for the set of all the (right kind of) reasons to place more weight on *A* to be weightier than the set of all the (right kind of) reasons to place more weight on *B*.' (Ibid., p. 138)

The problem with this account is that *Weight Base* seems fine, but *Weight Recursion* unhelpful. What is it for the set of all the (right kind of) reasons to place more weight on *A* to be weightier than the set of all the (right kind of) reasons to place more weight on *B*? Apparently, Schroeder thinks it is just a question of reapplying Weight Base to the sets of reasons to place weight on reasons:

'Weight Base simply tries to characterize what it takes for an explanation of the weight of some reason ultimately to get started. Weight Recursion tells us how, once it is started, it continues to proceed.' (Ibid., p. 139)

If that's right, then Schroeder's account of the weight of reasons seems to amount to the following: a set of reasons *A* to do *X* is weightier than a set of reasons *B* not to do *X* either if *B* is empty (there are no reasons not to do *X*) or if there are reasons to place more weight on *A* than on *B*. The latter is the case when the set of

reasons to place more weight on *B* is empty (which is not to say that there might not be reasons to place *some* weight on *B*; it's just that there can't be a set of reasons to place *more* weight on *B* than *A*, if we know that there is a set of reasons to place more weight on *A* than *B*).

But this is still unhelpful. What would it *be* for the set of reasons to place more weight on *B* to be empty? What would make it the case? The set of reasons to place more weight on *B* is empty only if the collective weight of the reasons to do *X* is greater than the collective weight of the reasons not to do *X*. So we would have to know something about the relative strength of the reasons for doing *X* and not doing *X*. But it seems that Schroeder has not told us what this is determined by.

The only way to get a complete account here, is to say that whether the set of reasons to place more weight on *B* is empty is determined by some fact about the presence, absence or number of defeaters for the reasons to do *X* and not to do *X*. For example, one could say that the set of reasons to place more weight on *B* is empty if there are more defeaters for the reasons not to do *X* than there are defeaters for the reasons to do *X*. Or one could say that the set is empty if there are more defeaters of the defeaters for the reasons to *X* than there are defeaters of defeaters for the reasons not to do *X*. Or something along those lines.

But no such analysis can work, since defeaters don't necessarily take all of a reason's force away. Schroeder himself accepts this, since he says that Tam makes my original reason for believing that Tom stole the book 'even worse than in the second case'. By a reason being 'worse' than another, he means that it does not have equal weight. But my reason for believing that Tom stole the book still had *some* weight

(despite Tim's undercutting it). This means that the amount by which a defeater reduces the weight of a reason to *X* may not be enough to make the latter weaker than some other reason not to *X*. If so, a reason R_1 to place weight on some reason R_2 may still be stronger than a reason not to place weight on R_2, despite the presence of a defeater for R_1. So there must be something other than the presence, absence or number of defeaters which partly but crucially determines the weight of a reason. Schroeder, however, has not told us what this other thing is. So his account of the weight of reasons is incomplete.

I conclude that Schroeder has not given us enough to understand why his Humean theory will not condemn many commonsensical moral judgments to falsehood.

3.7 'Ought' and 'must'

Before ending this chapter, a word about 'must'. Although Humeanism is versatile in the sense that it may provide theories of 'reason', 'ought' and 'the balance of reasons' (even if problematic), it is not at all clear what Humeans would say about 'must'. 'Must' has a normative use, just like 'ought'. And 'must' appears to be stronger than 'ought', as we can see from the following examples:

(1) Everybody ought to wash their hands; employees must.[36]

[36] This example is from Von Fintel & Iatridou (2008).

(2) You must return the money, but you ought to tell him what happened too.

Humeans say that '*A* ought to *X*' means that the balance of reasons favours that *A X*-es, which Harman analyses either in terms of what the agent would be motivated to do, or in terms of what some group would be motivated to hope for. It is not clear to me what he might say about 'must' and why it appears to be stronger than 'ought'. Since 'ought' is analysed in terms of something that an agent or a group of people *would* do (under certain conditions), it seems natural to think that the difference between 'ought' and 'must' should lie in a difference in the agent's motivation with respect to the act. But it is not clear how to think of this difference.

Things are no better for Schroeder. Since he has no clear theory of what it means to say that something is favoured by the balance of reasons, it is mysterious what 'ought' means. Since I don't know what Schroeder thinks 'ought' means, I am also in the dark about 'must' in his framework.

3.8 Conclusion

I've assessed two Humean theories of normative judgment. Both analyse 'ought' statements in terms of what is required by the balance of reasons. They explain why it can be simultaneously true that *A* ought morally to *X* and that *A* ought prudentially to not-*X*. However, Harman's analysis of 'sufficient reason' is inadequate: it has undesirable implications for debates about freedom of the will. Furthermore, it explains what one ought to do in terms of what one would be motivated to do. But it

does not tell us *why* one would be so motivated. This latter information seemed more essential to the truth conditions of judgments about 'ought' and 'reason'. The third problem with Harman's theory is that it makes too many moral and other normative judgments false. This is the result of building the assumption that others would be similarly motivated as oneself into the truth conditions of normative statements. A fourth reason to doubt Harman's theory is that it leaves it unclear how to think of 'must'.

Although Schroeder avoids Harman's first and second problems, he has given us no reason to think that hypotheticalism guarantees the truth of many intuitively valid moral judgments. This is mainly because he has not given us an intelligible theory which could explain why the balance of reasons falls out against abhorrent acts *no matter what*. But without such a theory, we cannot be sure that '*A* ought morally to *X*' will be true irrespective of the contents or properties of *A*'s desires. Furthermore, Schroeder gave us no more reason than Harman to think that his theory could deal with a distinction between 'must' and 'ought'.

So we'll have to move on to other theories. In the next chapter, I will assess a theory of 'ought' which ties the truth conditions of 'ought' statements to rules to which *the speaker* is committed (at least in certain cases). It is a version of the relativist view to which Harman seems attracted too.

Chapter 4. Standard-relational theories of 'ought' and 'reason'

4.1 Introduction

In the previous chapter, we've looked at Humean approaches to normative language. The main problem we encountered was that Humeans make the truth of (directive) 'ought' judgments depend on the desires (or other motivational states) of the agents to whom they are addressed. But we often seem willing to say '*A* ought to *X*' *irrespective* of *A*'s desires. Mark Schroeder tried to accommodate this by making it easy for an agent to have reasons. For example, anyone with (almost) any desires at all was supposed to have moral reasons. Furthermore, anyone with (almost) any desires at all was supposed to have *strong* moral reasons. But it was unclear how the latter could be justified. This mattered because '*A* ought to *X*' was supposed to be true just in case the reasons for *A* to *X* were stronger than the reasons for *A* to not-*X*.

In this chapter, we will look into standard-relational views of normative language. They are a form of *indexical relativism* in which normative statements implicitly refer to *standards*. They are referred to in the proposition expressed and can vary with the speaker. So, according to this view, normative words like 'ought' and 'good' behave like indexical expressions. Well-known examples of indexicals are 'I' and 'now': their contribution to the proposition expressed depends on who is using them.

Among the philosophers who have (more or less explicitly) associated themselves with this kind of view are David Wong (1984), David Copp (1995) and Gilbert Harman (1996). Although Richard Hare has not, his noncognitivism

presupposes certain elements of standard-relational theories. For example, in (1952), he says:

> 'To become morally adult is [...] to learn to use "ought"-sentences in the realization that they can only be verified by reference to a standard or set of principles which we have by our own decision accepted and made our own.' (Hare (1952), pp. 77-78)

I will return to Hare's standard-relational commitments in chapter 5.

It is important to discuss standard-based theories not just because they have been accepted by some important metaethicists. They also satisfy the requirements imposed on subjectivist theories of normative language: they are success-theoretic, truth conditional accounts that explain why '*A* ought morally to *X*' can be true while *A* ought not to *X* from some other point of view. Furthermore, it seems independently plausible that a statement like '*A* ought morally to *X*' is about the relation in which *A*'s *X*-ing stands to a system of standards or norms (even moral objectivists may be attracted to this kind of view). Lastly, it needn't be part of the proposition expressed by '*A* ought to *X*' that the relevant standards are held *by some particular person* (they can be though of indexically as *the* relevant standards). This increases the theory's appeal by "objectifying" the semantics.

In this chapter, I develop standard-relational theories of 'ought' and 'reason'. I will be drawing on work by Wong (1984), Copp (1995) and myself (forthcoming),

but add or change various features along the way. My aim is to identify the most promising version.

4.2 'Ought' and reasons

The idea that 'ought' statements implicitly refer to standards has been combined with Humean approaches to reasons (e.g. Wong (1984)). But I don't think that is a good idea. We have already seen reason to doubt that statements about reasons (always) involve the *agent's* desires. Furthermore, the combination of a standard-based theory of 'ought' with a Humean theory of reasons is awkward. '*A* ought to *X*' seems to entails that there is reason for *A* to *X*. This entailment is not easy to explain if the standards referred to by '*A* ought to *X*' are determined by commitments of the speaker. After all, Humeanism says that '*A* has a reason to *X*' is true just in case *A* has some desire which is served by *X*-ing. That *X*-ing is required of *A* by a standard of *the speaker's* does not entail that *A* has the requisite desires.

But even if there are no *logical* entailments between 'ought' and the balance of reasons, it is puzzling why 'ought' would be concerned with standards, if reasons are not. It seems that if standards determine what one ought to do, there should at least be a sense in which standards also determine reasons for action.

So I believe a standard-based theory of 'ought' should come with a standard-based theory of reasons. I am not the only one to think so. According to David Copp,

> 'There are reasons of a given kind only if there are standards of that kind with an appropriate standing. For example, [...] the proposition that there is a moral reason to choose *A* is true just in case there is a justified moral standard that calls for *A* to be chosen. [...] And a fact is a *K* reason [a reason of a certain kind] to choose *A* only if, given that fact, a *K* standard calls for the choice of *A*. This is the natural position to take, in light of the standard-based theory of normative propositions.' (Copp (1995), p. 168)

As we will see, the constraint that standards determine reasons matters for the *nature* of the standards that normative statements implicitly refer to.

So what are standards? Since we are interested in analyses of words like 'ought' and 'reason', it is best to keep them out of the standards which determine the truth of normative statements. For if they reappeared in the standards, that would invite the question what *their* meaning is. And this leads to a dilemma: either their meaning has to be explained with reference to other standards (leading to an infinite regress), or it can be explained without reference to standards. But the latter obviates the need to introduce standards into the truth conditions of normative statements (this point is also made by Boghossian (2006)).

No doubt in part for these reasons, standard-relational theorists have suggested that standards are contents which can be expressed by means of imperatives ('Do *X*!', 'Don't do *X*!'). Examples are Wong (1984), p. 40, Copp (1995), p. 9, 20 and myself (forthcoming). Richard Hare says that standards are 'universal imperatives sentences' ((1952), p. 134). Boghossian (2006) agrees that that this is natural as well

(although he disagrees with all standard-relational views). Because I will take standards to be expressible by means of imperatives, I will often refer to them as *rules*.

In (1984), Wong analyses '*A* ought morally to *X*' as follows:

> 'By not doing *X* under actual conditions *C*, *A* will be breaking a rule of an adequate moral system applying to him or her.' (Wong (1984), p. 40)

A drawback of this definition is that it seems to require highly qualified rules. Suppose that Aunt Betty is wearing a hat that I find ugly. 'What do you think of my hat?', she asks. On the one hand, I am not supposed to lie, but, on the other, I am not supposed to hurt her. If the rules which are part of the moral system were simple ('Don't lie!', 'Don't hurt people!'), we would likely break one of them in the imagined situation. But it doesn't follow that we act wrongly (a rule may be less important than another). So in order for Wong's analysis to work, he must have more complicated rules.

I don't think this can be avoided by inserting 'other things being equal' clauses into the rules. For there has to be a determinate fact about whether a rule is broken in order for wrongness to apply to actions. Unless we could in principle spell out the conditions under which other things are equal, no moral rule clearly forbids anything. So Wong has to say that moral rules are complex, highly qualified rules of the kind: 'Don't lie, unless *a*, or *b*, or *c*...'.

Does it matter if the rules are highly qualified? David Copp seems indifferent with respect to the question whether standards are simple or complex:

> 'Let me say that a standard "calls for" a person to do *A* in a given circumstance just in case, if the person failed to do *A* in that circumstance, he would thereby fail to conform to the standard. Let me also point out that although a moral code is a *system* of standards, a code can also be regarded as itself a standard, for it can be construed as specifying a complex criterion for the appraisal or judging of actions and persons.' (Copp (1995), p. 25)

But there is reason to prefer the idea that standards are relatively simple (i.e. can come into conflict with each other). Remember that standards are supposed to give rise to reasons. Copp said that 'a fact is a *K* reason to choose *A* only if, given that fact, a *K* standard calls for the choice of *A*' (ibid., p. 168). We ordinarily think that a fact like '*X* will kill people' is a reason against it. Similarly for '*X* involves lying'. But if there is only one highly qualified standard that grounds reasons, then such facts will not count as reasons. For if lying is sometimes permissible, and there is only one moral standard, then the fact that an act involves lying is not such that, given it, the moral standard calls for its omission. Instead, reasons will be more finegrained facts, such as '*X* involves lying without this yielding benefits *a*, *b*, *c*, etc.' (benefits which lift the prohibition).

So the first reason to prefer relatively simple rules is that a single highly qualified rule falsifies many intuitive judgments about reasons (such as 'The fact that people will die is a reason against the act').

The second reason to prefer simple standards is that reasons are normally thought to have weights. Some reasons are weightier than others. But if a reason is a finegrained fact such that, given it, a single qualified standard calls for an act, then all reasons are compatible. And if reasons cannot come into conflict with each other, there is no need to assign relative weights to them (at least not *within* a normative domain, like morals or prudence; reasons in different domains can still come into conflict).

The third and last reason to prefer simple rules is that they are easier to learn, so that a standard-based theory might gain psychological realism. Relatedly, some psychologists argue that relatively simple rules are involved in actual moral reasoning (Gill & Nichols (2008)).

4.3 A standard-relational theory of 'ought'

As Copp says, rules "call" for certain acts. What is it for an imperative to require an act? Copp takes this notion as primitive. That is one option. We can also say that a rule requires an act of an agent just in case it entails that the agent does it. We can explain this further by hypothesizing that imperatives are propositional contents put forward with imperative as opposed to assertoric force. For example, the rule 'Don't kill!' (as applied to everyone) may have the propositional content *that no one kills*. This content stands in classical relations of entailment with other contents (such as *that the agent does not kill*). Given these assumptions, we can explain what it means for an act to be required of forbidden by a rule as follows:

A rule *requires* an agent *A* to *X* in situation *S* just in case it (i.e. its propositional content) entails that *A X*-es in *S*.

A rule *forbids A* to *X* in *S* just in case it entails that *A* does not *X* in *S*.

It does not matter if rules don't actually have propositional contents. If they don't, we can simply take it as primitive that rules require or forbid certain acts. This will not make a difference so long as there are facts in virtue of which acts are required or forbidden by rules. If such facts exist, we can in principle reformulate the theory developed in this chapter.

Of course, rules do not entail much by themselves. Suppose there is a red button which, if pushed, kills many people. The rule 'Don't kill!' does not by itself entail that the agent does not push the button. It only does so in combination with additional premises, like 'Pushing the red button kills people'. So when I say that an act is forbidden just in case it entails that the agent does not perform the action, I mean that the rule entails this in combination with relevant information about the situation.

I will assume that rules require or forbid *types* of acts. When we talk about *types* of acts, we mean to talk generally about acts which exemplify certain properties (without referring to any particular act token, and without implying that such a token exists). When we talk about act tokens, we refer to specific spatio-temporal events. What type an act token instantiates is determined by the properties it exemplifies.

Now we have enough elements to start semantic analysis. I will state what I take to be promising imperative-based analyses of 'ought' and 'reason'. They are reminiscent of Wong and Copp, but differ in the details.

According to a standard-based theory of 'ought' what one ought to do is determined by rules. Since rules can come into conflict, what one ought to do is not a matter of individual rules, but of a *system* of rules. Roughly, one ought to *X* just in case the relevant system requires *X*-ing.

What is it for an act to be required by a *system* of rules? For ease of exposition, I will stipulate that '*X*' is a variable that takes both acts and omissions as values. So '*X*' can both stand for the act of killing and the omission of refraining from killing. I will sometimes simply talk about *acts*, which is supposed to comprise both doing something and refraining from doing something. Now I can say that *X* is required by a system if and only if *complying* with the system requires that one *X*-es. This can mean either one of two things:

(a) In a simple case, where the rules have no inconsistent implications with respect to *X* in a situation *S*, complying with the system means doing what is required by at least one rule in the system.

(b) In a complex case, where the system contains rules with inconsistent implications with respect to *X* in *S*, complying with the system means doing what is required by the rule with the highest priority. For example: a system may contain both a rule of morality (which requires *X* in *S*), and a rule of prudence (which requires not-*X* in *S*). In this case, *X* is required of you in *S* by the system if and only if (1) the

moral rule requires that you *X* in *S* and (2) the moral rule has higher priority than the rule of prudence.

We can now state a general definition which covers both cases of complying with a system of rules:

Compliance: an agent *A* complies with a system of rules *R* by *X*-ing in *S* if and only if a rule in *R* entails that *A X*-es in *S* and no rule in *R* with the same or higher priority entails that *A* does not *X* in *S*.[37]

Compliance raises the question what having higher priority amounts to. I will explain this in more detail below, but for now I restrict myself to saying that a rule has higher priority than another rule (in *S*) just in case a higher-order rule requires that we act in accordance with the first rule (in *S*).

We can now give truth conditions for 'ought' statements as follows:

Ought: '*A* ought to *X* in *S*' is true if and only if *A* complies with the contextually salient system of rules by *X*-ing in *S*.

Equivalently:

[37] Complying with a system is different from conforming to a system. One conforms to a system just in case one's act is not forbidden by a rule with higher priority than any rule that requires it.

Ought*: '*A* ought to *X* in *S*' is true if and only if the contextually salient system of rules requires that *A X*-es in *S*.[38]

In the next two sections, I will explain some of the elements of this truth condition in more detail.

4.4 Higher-order rules

I said that higher-order rules determine the relative priorities of rules. So what are higher-order rules? A higher-order rule is also an imperative, like first-order rules. The difference is that higher-order rules require us to act in accordance with a first-order rule in a specific situation. This formulation is ambiguous between the following two options:

[38] I said that '*A* ought to *X* in *S*' is true just in case *the contextually salient system of rules* requires that *A X*-es in *S*. This is another departure from Wong's theory. Wong said that '*A* ought morally to *X*' is true just in case by not *X*-ing, A would break a rule of *an* adequate moral system (one or another). So Wong believes that 'ought' statements quantify over systems of rules. They do not refer to any particular system.

There is a presumption in favour of reference over quantification, for the following reason: if several systems with some incompatible impliciations can be adequate, there would be acts which are both morally right and wrong. This is undesirable. But if no two systems with incompatible implications can be adequate, we must be referring to a single system only.

(1) The content of the higher-order rule is general: it tells us to act in accordance with rule R_1 in any situation where it conflicts with certain other rules R_2 ... R_n.

(2) The content of the rule is particular: it tells us to act in accordance with rule R_1 in the situation we are in right now.

I think the first is preferable. Option (2) amounts to a kind of *particularism* about normative judgment: the view that there are no general principles determining the weight of reasons (or, in this case, rules). It could be objected that a higher-order rule with content (2) is not a rule at all, precisely because it lacks generality. But I'm not sure that the lack of generality is the problem. I think the problem is that highly situation-specific rules could not plausibly count as entities that the subject is (implicitly) *aware* of when s/he forms normative judgments. They could not plausibly be what the judgment is *based* on.[39] So if, in assigning referents to terms, we should look at what application of the term is sensitive to (see chapter 1, section 1.9), we could not plausibly say that 'ought' is sensitive to highly situation-specific rules.

[39] Although we should not make the mistake of thinking that subjects should be able to *articulate* the rules on which the truth of their normative judgments depends. David Copp provides the following analogy: 'The standard-based theory does not imply that a person who believes that a given sentence is ungrammatical would be able to articulate a standard that the sentence fails to meet. The theory does not imply that anyone who has a normative belief must be able to articulate a relevant standard.' ((1995), p. 32)

So option (1) is preferable. However, if the relative priorities of rules are not stable across different situations, then it seems that higher-order rules would have to be very complicated (at least in order to be plausible): 'Act in accordance with rule R_1 in case of conflict between R_1 and R_2 ... R_n, except if this, or that, or ...'.

But, first, it's not obvious that the relative priorities of rules vary in different situations. The fact that lying is sometimes permissible and sometimes not does not show that rules have different relative priorities in different situations. If lying is permissible, that is usually because some other rule applies which takes priority. If the other rule does not apply, then it is impermissible. So perhaps higher-order rules needn't be too complicated.

Second, the higher-order rule to which the speaker is referring only has to order the rules which *actually* apply in a situation of conflict. They don't have to order these rules relative to all other rules that might sometimes apply. In a situation of normative conflict, a subject is presumably (implicitly) aware of the conflict between a limited number of first-order rules. S/he may then reach a judgment about what to do all things considered. This judgment will involve reference to a higher-order rule that needn't say anything more general than to prioritize (actually applying) rule R_1 over (actually applying) rules R_2 ... R_n (in any situations where they conflict).

4.5 Rule determination

What is the contextually salient system of rules determined by? As the phrase suggests, it is determined by context. Exactly how this works varies from case to case.

In sincere moral 'ought's, the speaker's own commitments are involved in the identification of the rules. But speakers can also refer to rules they don't (necessarily) endorse. This happens if the relevant rules are the rules of etiquette, legal requirements, or when 'ought' is used in inverted-commas. It also happens when we determine what to do given someone else's principles.

It is important to remember that, although the rules are sometimes determined by the speaker's attitudes, the rules are never identified *as the ones accepted by the speaker*. There are two reasons to avoid this. The first is that not all 'ought' statements are made in the light of rules that the speaker endorses. If so, avoiding positive determiners (like 'my' and 'your') leads to greater semantic uniformity in different normative domains. The second is that it allows more genuine disagreement (see also chapter 2, section 2.4). Suppose that '*A* ought to *X*' meant '*X* is required of *A* by *my* (the speaker's) rules'. Now take two subjects John and Sue. Suppose John and Sue accept the very same rules, but disagree about their application. John says '*A* ought to *X*' and Sue says '*A* ought not to *X*'. Do they contradict each other? No, because John said something about what *his* rules require, whereas Sue made a statement about *hers*. This blocks a contradiction even if John and Sue have the very same rules in mind.

We can avoid this by deleting the positive determiner 'my' from the truth conditions. We should think of John and Sue as pointing to contextually salient rules and claiming things about what they require. In doing so, John and Sue do not identify them as their own (which will be clear from context). This will allow John and Sue to contradict each other if they refer to the same rules.

I will assume that it is clear enough how speakers manage to refer to systems of rules if the rules are generally acknowledged (as is the case with etiquette) or even codified (as is sometimes the case with laws). But how does the speaker manage to refer to a system of rules if they depend on him or her? In this case, there is no rule-book to defer to, nor widely followed practices (at least not necessarily).

I've already indicated that speakers needn't be able to articulate the rules on which the truth of their normative statements depends (see footnote 39). However, truth conditions do not float completely free of anything that speakers do.[40] So what makes it plausible to assign systems of rules to the utterances of speakers in those cases where the rules depend on them?

Here is a suggestion: rules require or forbid *types* of acts, and what type an act instantiates is determined by the properties it exemplifies. A standard-relational theorist could say that what makes it plausible to have rules as part of the truth conditions of normative statements is the role played by features of acts in the practical reasoning of the subject. For example, that an act kills may play the role of disposing one against it: in the subject's psychology, the input '*X* kills' triggers the intention to avoid *X*, other things being equal. Similarly for "positive" features. These

[40] The dependency of truth conditions on the activity of speakers can be quite counterfactual. For example, it is plausible that 'water' referred to H_2O even before the rise of chemistry. This must somehow be related to the fact that at least some suitable subset of the linguistic community *would have* retracted their judgment that something is water if the relevant substance turned out to have a different molecular structure. If no one had this disposition, it is doubtful that the reference of 'water' really was confined to H_2O.

roles can be captured by the theoretical notion of a rule (especially if a rule can be trumped by other rules).

Not all rules will correspond to *pre-established* roles in a subject's psychology (i.e. roles established prior to actual reflection on a case). Imagine a conflict between a moral rule which requires *A* to *X*, and a prudential rule requiring the opposite. The subject may not have encountered such a conflict before. If so, the standard-relational theory predicts that the subject is (at first) unsure whether *A ought* to *X* or not. And this seems right. But if s/he subsequently judges that *A* ought (all-in) to *X*, the theory invokes a higher-order rule which arbitrates between these conflicting requirements. The inclusion of this rule in the truth conditions is justified by three things: (1) the fact that the subject actually reached the conclusion that *A* ought to *X*, (2) the fact that whatever disposed the subject to do so is likely to play a similar role in similar situations and (3) the fact that 'ought' does not radically differ in meaning just because 'all things considered' is inserted in the sentence. The latter consideration does, of course, depend on how plausible it is to think that rules are involved in normative discourse in the first place.

4.6 Advantages of the standard-relational theory of 'ought'

The standard-based theory of 'ought' says that '*A* ought to *X*' is true just in case the contextually salient system of rules requires that *A X*-es. So '*A* ought *morally* to *X*' means that the contextually salient system of *moral* rules requires *X*-ing. '*A* ought

prudentially to *X*' means that the contextually salient system of *prudential* rules requires *X*-ing.

This theory has a number of virtues: first of all, it is unified for all nonepistemic uses of 'ought' (I will return to this issue in chapter 5). Second, it has intuitive appeal, at least for lower-level 'ought's: it is plausible that '*A* ought morally to *X*' is about the requirements of moral rules. But since it is also plausible that 'ought' means something similar in moral, prudential, and other contexts, it is a virtue that the semantics assigns the same value to all such 'ought's (i.e. the relation of being required by some relevant system). Third, it does not build the desires of either speakers or agents into the truth conditions of normative statements. Moral statements are about the requirements of rules, not about anyone's desires (even if the rules are determined by desires). Because normative statements are not relative to the agent's desires, the standard-relational theory secures the truth of many normative statements (including moral ones). Fourth, it allows us to explain why speakers are often motivated to act in accordance with their 'ought' judgments: these judgments often concern rules to which the speaker is him/herself committed. And (in ordinary cases), such commitment will in part be a matter of motivational states. Fifth, as we will see, it is compatible with a theory of reasons which explains why '*A* ought to *X*' entails that there are reasons for *A* to *X*, and also why '*A* ought all-in to *X*' entails that the balance of reasons favours *X*-ing.

4.7 A standard-relational theory of reasons

I've indicated that a standard-based theory of 'ought' should come with a theory of reasons. Such a theory will say that things become reasons in virtue of standards. In order to be a reason, a thing has to stand in a particular relationship to standards (in this case: imperatives).

Following a popular trend, the standard-relational theorist could take reasons for action to be facts (like Broome (2004) and Schroeder (2007) do).[41] But nothing can be a fact which is not the case. So what about situations where we act for reasons, but turn out to be mistaken about the facts? Wouldn't it be better to say that reasons are *propositions*, whether true or not?

I don't think so. The reasons *for which* someone acts may not be facts. But in specifying the reasons for which someone acts, we specify things which this person

[41] In an earlier formulation of the standard-based theory of reasons, I said that reasons are *features* (properties) of acts (forthcoming). I now think this is wrong. It is at least somewhat odd to say that *properties* are reasons. At least as a matter of grammar, it's odd to state properties in answer to a question about the reasons for an act. 'Being such as to cause pain' would be a strange thing to say, whereas it's completely natural to state a (purported) fact: 'It causes pain'.

The second reason to prefer facts to properties is that the reason-making relationship might well be one of explanation: for example, reasons might be things which explain why standards apply to an act (more on this below). And it is quite common to think of *explanans* as propositions, which can be used to state facts. (I owe this point to Robert Schwartzkopff.)

The third reason to prefer facts is that there seem to be such reasons as *there being dancing at the party* (a reason to go). But this is not a property of the act of going to the party, or indeed, of any act (except if we are prepared to countenance artificial properties like 'being such that there is dancing at the party').

takes to be reasons. However, this person would accept that s/he was mistaken about the reasons if the relevant beliefs turn out to be false. This shows that there is a systematic ambiguity in reason-talk: motivating reasons are beliefs about the reasons that there are (whether true or false), but normative reasons are facts (what is stated by true propositions).

If reasons are facts, what makes these facts into reasons? A standard-relational theorist would say: a relationship to rules. But what is this relationship? It seems natural to say that a fact is a reason just in case it explains why a rule requires or forbids an act (Copp's view). I've explained the notions of requiring and forbidding in terms of entailment: an act is required of an agent by a rule just in case the rule (i.e. its propositional content) entails that the agent performs it (in combination with relevant premises). An act is forbidden just in case the rule entails that the agent does not perform the act.

When I say that a reason is a fact which *explains* why a rule entails that the agent does (not do) an act, I don't mean that it is a complete explanation. It is merely part of such an explanation. Suppose that the relevant rule is: 'Seek enjoyment!'. If Ronnie loves to dance and there is dancing at the party, then the fact that there is dancing at the party is a (normative) reason for Ronnie to go. This is because going to the party is a means for Ronnie to enjoy himself. This needs to be added in order for the rule to entail that Ronnie goes to the party. So the fact that there is dancing at the party is merely part of an explanation of why the rule entails that Ronnie goes.

Now we have to be careful not to generate too many reasons. We don't want facts about the laws of logic to be reasons for or against acts. Nor such generic facts as

the agent's being *capable* of *X*-ing. But aren't these equally part of an explanation of why the rule entails that the act is done or not?

The latter problem can, I think, be solved by noting that explanation is relative to assumptions. Some things are better described as part of the presuppositions of explanations (part of the assumptions relative to which certain facts acquire explanatory roles). For example: the laws of logic are ordinarily presupposed when we ask: 'Why is the roof leaking?'. Similarly, the fact that an agent is capable of *X*-ing is not part of an explanation of why the rule entails that the agent does it, because when we ask what reasons there are for the agent to *X*, we already presuppose that the agent is capable of *X*-ing. After all, reasons for action count in favour of things that can be *done*. So facts about the agent's ability are presupposed. I will call facts which are part of the presuppositions of explanations *presuppositional*.

So we have arrived at the following view of reasons:

> **Reason For:** a reason for an agent *A* to *X* in *S* is a nonpresuppositional fact which explains why a relevant rule entails that *A* does *X* in *S*.

> **Reason Against:** a reason for an agent *A* not to *X* in *S* is a nonpresuppositional fact which explains why a relevant rule entails that *A* does not *X* in *S*.

The definition is restricted to *relevant* rules, which are rules that are contextually salient in the conversation or thought process at hand. This is necessary

in order to forestall overgeneration: there is an infinite number of rules in logical space, but that does not mean that there are reasons for anything whatsoever.[42]

I have suggested that reasons are facts which explain why a relevant rule entails that the act is (not) done. I have not said that reasons are facts which explain why *the system* requires the act. This is to allow for the possibility that one has reasons for acts which one ought not, on the whole, to do.

This standard-relational theory of reasons has the advantage of tying reasons to standards (just like 'ought'). In the case of moral judgment, the standards referred to are determined by the speaker's commitments. This allows a speaker to be right that there is plenty of moral reason for Hitler not to murder Jews even if abstaining from such crimes would not further his desires.

But have we not made reasons *too* independent of the agent's desires? Schroeder is right that if Ronnie likes to dance, then he has a reason to go to the party. Bradley, who doesn't like dancing, lacks this reason. Why do some reasons seem to vary with desires of the agent?

The answer is twofold: if the agent's desires matter to our judgment about what s/he ought to do, that will typically be because they are relevant to something which is deemed independently valuable (i.e. independently of the agent's desires).

[42] Or at least, there aren't reasons to do anything whatsoever relative to the system salient in the conversation. There is a sense in which, in a standard-relational framework, there are reasons to do anything whatsoever. Like all reasons, these are relative reasons in the sense that they are reasons relative to certain possible rules.

For example: the desires of the agent matter to what s/he ought to do because their satisfaction affects the agent's happiness, which we consider valuable.

But this does not suffice. Even if *we* are committed to certain values (like the protection of life), we can still make sense of the idea that someone else – say, a psychopath - might lack reasons to abstain from murder. It does not seem plausible that we can make sense of this only by thinking about what is conducive to the psychopath's happiness.

I think the standard-relational theorist should respond to this as follows: when we judge that the psychopath has no reason to abstain from murder, we look at what is valuable from *his* or *her* perspective. If the psychopath does not accept certain standards, then it is in this sense that s/he lacks the corresponding reasons. But from our perspective, we may still judge that there is plenty of reason for him or her to abstain from murder.

These explanations further justify my rejection of Harman's Humeanism in chapter 3. Harman made the truth of directive 'ought' judgments depend on what the agent would be motivated to do. I objected that not all such judgments seem sensitive to the agent's motivations. This made it less plausible that their truth would depend on them. But I also envisaged that Harman might respond as follows: some reason ascriptions do seem sensitive to the agent's mental states (witness Ronnie and the psychopath). If our best explanation of these phenomena were Humean, we'd have reason to think that all statements about reasons involve the agent's motivations. Since the standard-relational theory can make good sense of the phenomena as well, it

has an advantage over Harman's view: it avoids large-scale error in normative judgment.

Before I explain how to think of a reason's weight in a standard-relational framework, let me answer two objections to my truth conditions for reason statements (both suggested by Ralph Wedgwood). They may not generate enough of the reasons that intuitively exist.

The first problem is reminiscent of Buridan's ass.[43] A hungry ass is positioned between two equally desirable stacks of hay. It cannot eat from both, and so it has to choose. Intuitively, the ass has a reason to choose the left but also a reason to choose the right one. What rule is supposed to ground these reasons? Suppose that the rule says: 'Satisfy your basic needs!'. This rule (plus information about the situation) does not entail that the ass chooses the right stack of hay. Nor does it entail that it chooses the left one. Rather, it entails that the ass chooses either the left or the right one (but not both). And so, given the standard-based theory, the ass does not have a reason to choose the right, nor a reason to choose the left stack of hay.

Here, the standard-relational theorist could reasonably deny that the ass has a reason to choose the left and a reason to choose the right stack of hay. Instead, it might be claimed, it has a reason to choose *one or the other* (not a reason to choose the one *and* a reason to choose the other).[44]

[43] The difference being that Buridan's ass had to choose between drinking or feeding.

[44] This response is suggested on behalf of a different theory by Kearns & Star (2009), pp. 238-239.

But if this response is given, we should explain why it seems as if the ass has a reason to choose the left and a reason to choose the right stack. This may be explained by a certain shift in the class of acts the rules are applied to. When we judge that there is a reason to choose the left, we implicitly compare this to choosing neither (staying hungry). In the light of such a choice, the ass has a reason to choose the left. And the same holds for the right stack of hay.

The second problem is as follows: suppose we have a choice between three options. The first is to to lose 10 pounds, the second to win 10 pounds, and the third to win 50. I'll refer to these options as '–10', '10' and '50' respectively.

Our standard-based theory says that a reason to *X* is a fact which explains why a rule entails that the agent does *X*. Suppose we have a rule which says: 'Maximize your profit!'. This rule entails that we choose 50. It does not entail that we choose anything less, and so there is no fact which explains why the rule entails that we do so. But, it might be thought, we do have *some* reason to choose the 10. However, our standard-based theory does not predict this.

I think the standard-relational theorist should question the intuition that we have a reason to choose the 10. Suppose there is no nonprofit-related reason to take less rather than more. So there are no reasons to leave something for others or to avoid seeming greedy. If no such reasons exist, why should we have *any* reason to choose the suboptimal act?

We can then offer an explanation of the intuition that we have some reason to choose the 10: it derives from the true judgment that we have a reason to choose 10 instead of –10 in case of choice between the two. In that case, the 'Maximize your

profit!' rule *would* tell us to go for the 10. This might be put by saying that 10 is *better* than –10 (see also chapter 5). And that explains our intuition.

These responses are not obviously bad. I tentatively note that a theory which allows an *actual* reason to choose the 10 might be preferable (other things being equal). It might also be preferable to allow that the ass has a reason to choose the right and the left stack of hay. But even if it is, these problems are not lethal by themselves. Furthermore, as I will now explain, the standard-relational theory has a clear advantage: it can explain why '*A* ought to *X*' entails that the balance of reasons favours that *A X*-es, and the other way around.

4.8 'Ought' and the balance of reasons

Given that the theory of 'ought' presented above involves many different rules, one might try to correlate the strength of a reason with the position of the rule (to which it owes its status as a reason) in the order of priorities. The higher up the rule, the stronger the reason.[45]

Although I was attracted to this idea in (forthcoming), I no longer believe we can make the strength of a reason correspond to the ranking of a rule. The problem is

[45] Something like this is proposed, in a slightly different context, by John Horty (2007). Horty is interested in formal models of reasoning about belief and action. In Horty's framework, reasons are propositions which are the antecedents of default inference rules (rules of the form '$p \rightarrow q$', although they are not quite material conditionals). The weight of reasons is determined by the priority ordering of the default inference rules.

that reasons can have different weights even if they are reasons in virtue of the same rule. For example, the fact that an act will kill 10 people is stronger reason against it than the fact that it will kill 1. But unless we have separate rules against killing for each natural number, we cannot explain the difference in weight between these reasons in terms of a difference in positioning of rules.

If the relevant rule is 'Minimize killing!', then there is a reason against the act which kills 10 which there is not against the act which kills 1. Or at least, there is such a reason in cases of choice between two acts with these respective consequences. For the rule 'Minimize killing!' entails that you minimize killing, which, in this case, is achieved by choosing 1 over 10. And so there is a reason in favour of the act which kills 1 (since the rule entails that you perform this act), and a reason against the one which kills 10 (since the rule entails that you don't perform this act). Given a minimize rule, then, there are reasons against acts which involve more than the fewest deaths which there are not against acts which minimize the number. Can we explain the difference in weight with reference to this fact?

I doubt it. Two problems emerge: (1) 10 deaths is stronger reason than 1 even in cases where the choice is not between two acts with these respective consequences. (2) The minimize rule entails that the agent does the act which minimizes deaths, and therefore that s/he does none of the acts which involve more than the smallest number. So, given just the minimize rule, there is not more reason against an act which causes 1000 deaths than there is against an act which causes far fewer but not the smallest number.

These problems can be solved by a simple counterfactual theory of weight:

CTW: a reason (or reasons) F_1 is weightier than a reason (or reasons) F_2 iff the act for which F_1 is a reason (are reasons) would be required in case of conflict, and F_1 and F_2 are the only relevant reasons.[46]

This theory makes the weight of reasons depend on/consist in what one ought to do in certain (counterfactual) situations. Since we already have truth conditions for 'ought', we can simply apply those to the imagined situations.

I said that CTW helps to solve the two problems just described. The first was this:

(1) 10 deaths is stronger reason than 1 even in cases where the choice is not between two acts with these respective consequences.

If the relative weight of reasons is a matter of what *would* be required in case of conflict (i.e. were they to be reasons for incompatible acts), then CTW explains why 10 deaths is stronger reason than 1 even in cases where the choice is not between two acts with these respective consequences. After all, even in such cases, it is still true that we *would* be required to choose the smaller number of deaths in case of conflict (or at least this is the case if a minimizing rule applies).

The second problem was as follows:

[46] This is vaguely reminiscent of Joshua Gert's theory of weight in (2005).

(2) We need to guarantee that there is more reason against an act which causes 1000 deaths than there is a against an act which causes fewer deaths, even if neither of them minimizes the number of deaths.

CTW predicts that there *is* stronger reason against an act which kills 1000 than there is against an act which kills 100, since we would be required to choose the 100 instead of 1000. This is true even if neither act is the one which minimizes the number of killings.

But there is a worry. Doesn't CTW get the order of explanation wrong? Shouldn't what one ought to do be determined by the weight of reasons, instead of the other way around? This is not clear to me, since one may argue that to judge that F_1 is weightier than F_2 *is* to judge that, in case of conflict, F_1 is to prevail (i.e. one ought to do the act for which F_1 is a reason). It is not clear that we first need to make some other judgment (a judgment of weight) and then do some reasoning in order to establish that it is indeed the act for which the weightiest reasons exist which one ought to do. Furthermore, a subjectivist about normativity is likely to think that our commitment to rules is determined by non-cognitive states, like preferences. But these same preferences are likely to explain our commitment to differential weights. So, given the subjectivist's moral psychology, one should expect the truth conditions of 'ought' judgments and judgments of weight to be closely related.

The true advantage of CTW is that it straightforwardly explains why '*A* ought to *X*' entails that the balance of reasons favours that *A X*-es, and why 'the balance of

reasons favours that *A X*-es' entails that *A* ought to *X*. Saying that the balance of reasons favours *X* is saying that the collective weight of the reasons in favour of *X* is greater than the collective weight of the reasons in favour of not-*X*. The counterfactual theory of weight proposed above says that reasons are weightier than other reasons just in case the act for which the first are reasons would be required in case of conflict. If we ought to *X*, then *X* is required and not-*X* forbidden. So CTW makes it trivially true that if we ought to *X*, then the reasons for *X* are collectively weightier than the reasons for not-*X*. And the reverse entailment also holds: if it is true of a set of reasons that, *had* they applied, one would have ought to *X*, it follows that one ought to *X* if these reasons do apply. Thus it follows from: 'The balance of reasons favours *X* in *S*' that one ought to *X* in *S*.

It is a significant advantage if a theory of 'ought' explains why '*A* ought to *X*' logically entails that the balance of reasons favours that *A X*-es, and *vice versa*. In chapter 8 we will see that it causes a problem for Stephen Finlay's theory of normative discourse.

4.9 Conclusion

In its current guise, the standard-relational theory seems quite powerful. The theory of 'ought' avoids large-scale error in normative judgment and the theory of reasons and their weight explains why '*A* ought to *X*' entails that the balance of reasons favours that *A X*-es. So far, then, it is a candidate. But 'ought' and 'reason' are not the

only normative words. In the next chapter, we will see how the theory fares when it comes to words like ‘good’ and ‘must’.

Chapter 5. Standard-relational theories of 'good' and 'must'

5.1 Introduction

In the previous chapter, I developed a standard-relational theory of 'ought' and 'reason'. According to it, '*A* ought to *X*' is true just in case the contextually salient system of rules requires *X* of *A*. Rules were conceived of as imperatives, to avoid circularity in the analysis. A reason for *A* to *X* was a fact which explains why a relevant rule entails that *A X*-es (in the situation).

'Ought' and 'reason' are two important normative words that stand in intimate relations with each other. What one ought to do is related to the reasons that there are. If one has most reason to *X*, then one ought to *X*. But there are other words, like 'must' and 'good'. This chapter is almost entirely devoted to the notion of goodness. I will touch on 'must' only near the end. Looking into these words will give us a better sense of the prospects of the standard-relational theory of normative discourse.

5.2 'Good' and standards

It sounds plausible, truistic even, that a thing's goodness depends on the standards for the thing in question. Many philosophers have said this. John Searle says that a prominent meaning of 'good' is 'meets the criteria or standards of assessment or evaluation' ((1962), p. 432). David Wong suggests the same in ((1984)). And David Copp says that to be good is to conform to standards ((1995), p. 26). Although

Richard Hare denies that 'good' can be *analysed* in terms of satisfying standards, he does think that standards determine the *criteria* of goodness. Finally, John Mackie observes that '[e]valuations of many sorts are commonly made in relation to agreed and assumed standards' ((1977), pp. 25-26).

So long as we don't mean anything in particular by 'standard', it is obviously true that goodness is (at least often) a matter of satisfying standards. But it is not so obviously true if we mean *imperative* by 'standard'. In what follows I will consider three different standard-relational analyses of 'good', its comparative ('better') and superlative ('best'). None is free of problems.

5.3 'Good' defined as what we ought to choose

David Wong analyses 'ought' in terms of the requirements of systems of rules, (a bit) like I did in the previous chapter. Like me, he thinks of rules as imperatives. '*X* is a good *Y*' he analyses as follows:

> 'Under actual conditions *C*, *X* satisfies the appropriate standards for *Y*s.' ((1984), p. 70)

Wong says that a 'standard for *Y*s has the form "A *Y* is to be *F*"' (ibid., p. 69). But 'A *Y* is to be *F*' is not an imperative, and I don't think Wong can use this form without further elucidation. For it seems that 'A *Y* is to be *F*' means the same as 'A *Y*

ought to be *F*'. And Wong analysed 'ought' in terms of the requirements of imperatives.

So one option for a standard-relational theorist is to say that 'good' can be analysed in terms of 'ought'. But how would this go in detail? Suppose we say 'This is a good apple'. If 'good' can be analysed in terms of 'ought', and 'ought' is analysed in terms of rules and their requirements, there should be some rules which apply to apples (or at least rules that are in some sense concerned with apples). What might these rules be? One problem here is that imperatives are addressed to agents. But apples aren't agents (one cannot tell them to be juicy).[47]

I used to think this problem could be solved by means of constructions like 'Let *Y*s be *F*', which are (at least grammatically) imperatival. For example, one might think that a relevant imperative for apples could be: 'Let apples be juicy'. But this does not really circumvent the need for an agent to whom it is addressed. Furthermore, even if we could specify an addressee, the imperative is not intelligible as it stands. Since no human agent can make apples juicy by magic, what are the addressees supposed to *do* about the apples? *How* are they to let apples be juicy?

These reflections show that even let-imperatives demand acts, and demand them of someone. So we still need an answer to the question what imperatival contents determine the truth of 'This is a good apple'.

---

[47] Incidentally, this is germane to the question what statements like 'There ought to be no suffering' mean. Such statements seem to express requirements without there being anyone who "owns" them, in the sense that s/he ought to bring it about that the state of affairs obtains. Perhaps they are better thought of as claims about goodness, e.g. 'It would be better if there were no suffering'.

David Wong said that '*X* is a good *Y*' means that, *under actual conditions C*, *X* satisfies the appropriate standards for *Y*s, and he gives the following gloss:

> 'The clause "Under actual conditions *C*" is explicitly or implicitly given in the context of utterance and serves to identify certain parameters for the evaluation of *X*.
>
> For instance, an act may be called good under the condition that its point is to fulfill a certain end. Fasting is a good action under the condition that its point is to symbolize a commitment to end starvation and hunger. It is questionable as a means to losing weight.' (Ibid., p. 69)[48]

Suppose the end is to enjoy eating an apple. Given that aim, one can say that juicy apples are to be *preferred* or *chosen* (by whomever has the aim). If the aim is not part of the imperative which determines the truth of 'This is a good apple', then it would be: 'Prefer/choose juicy apples!'. If it *is* part of the imperative, it becomes hypothetical: 'If one wants to enjoy eating an apple, then choose a juicy one'.

Now we could analyse 'good' (as applied to apples) in terms of 'ought' as follows:

[48] The gloss reveals that it does not really matter that the conditions are actual. Something can be a good poison even if no one in the actual conditions wants to poison someone.

(1) '*X* is a good apple' is true in a context of utterance *C* iff one [i.e. the relevant agents] ought to choose *X* in the relevant situation of choice *S*.

In other words:

(1*) '*X* is a good apple' is true in *C* iff the contextually salient systems of rules requires one to choose *X* in *S*.

The context of utterance includes certain salient ends or purposes which select relevant imperatives. The situation *S* is the situation of choice that we are considering.

Given the analysis of 'ought' developed in the previous chapter, (1*) is true just in case at least one rule in the system entails that the relevant agents choose *X* (in *S*) and that no rule with the same or higher priority entails that they do not choose *X* (in *S*). Also as before, the rules entail this in combination with relevant information about the situation (such as information about the agents' aims).

This analysis may work so long as gradable adjectives are used in the imperatives. After all, one wants to allow that several apples (not just the juiciest and thus best ones) can be good. This is allowed by (1) in virtue of the fact that we are assuming that the relevant system of rules contains a rule which says 'Choose juicy apples!' or 'If one wants to enjoy eating an apple, then choose a juicy one!'. Since several apples can be juicy (can be sufficiently juicier than the members of a comparison class), several apples can be good. But is it plausible that all these apples are such that we ought to choose them? What if we only want (or need) *one* apple?

This problem is the result of a certain reading of 'Choose juicy apples!'. If its meaning is: 'For all apples, choose juicy apples!', then the undesirable entailment holds. (This is true even if we restrict the quantifier to a contextually salient collection of apples.) But there is an alternative reading. According to it the imperative says: 'If one chooses an apple (and wants to enjoy eating it), then choose a juicy one'.[49] This reading does not entail that one chooses all the juicy apples. It entails that we choose a juicy apple each time we make a choice between some apples. If we need five apples, then we make five choices. If we need one, then we make one choice.

Although I wouldn't claim that the following worries are decisive, I don't think (1) is the best analysis of 'good' for the standard-relational theorist. First, there is something odd about defining 'good' in terms of what one ought to choose. That is because in many uses of 'good', 'best' indicates a higher grade of goodness. Shouldn't the *best* or *optimal X* be the one we ought to choose (at least relative to some purpose)? Second, it is not obvious how to distinguish between the meaning of 'good' and 'best' if we analyse 'good' in terms of the requirements of rules (although there might be a way, as we'll see in section 5.3). Third, it makes intuitive sense to say that one ought to choose an apple *because* it is good (where this is not the same as saying that one ought to choose an apple because one ought to choose an apple). Calling an apple 'good' seems to indicate something about its nonevaluative properties. (1) arguably makes the information about the apple's properties too indirectly related to the

[49] I am indebted to Natalja Deng for this distinction.

semantic content of 'good'. For these reasons it does not seem the best option for the standard-relational theorist.

5.4 'Good' defined in terms of 'better'

Another possibility is suggested by Richard Hare ((1952)). Hare defines 'good' in terms of 'better' and 'better' in terms of what one ought to choose (ibid., chapter 12). Strictly speaking, Hare defines an *artificial* notion of 'better than' in terms of 'ought':

> '*A* is a *better X* than *B*' is to mean the same as 'If one is choosing an *X*, then, if one chooses *B*, one ought to choose *A*'. (Ibid., p. 184) (It is clearly Hare's intention that *A* and *B* are the only options.)

But Hare thinks this artificial notion performs many of the functions of the ordinary language one:

> 'Now I think that it will be agreed that '*better* than', as so defined, could do fairly adequately the job that is done in ordinary language by 'better than'.' (Ibid., p. 185)

He then defines an artificial notion of 'good' in terms of 'better', by saying that a good *X* is an *X* better than they usually are (ibid., p. 186).

David Wong objects to this analysis as follows:

> 'Like Bernard Williams, I find no contradiction in the idea that the game of cricket can flourish so much that most cricketeers are pretty good.' ((1984), p. 71)

This is a good objection on a slightly uncharitable reading of Hare's idea. In chapter 12 of ((1952)), Hare notes that 'good' is like 'hot', in that it introduces a comparison class. A good *X* is an *X* which has certain properties which are either lacked or exhibited to lesser degrees by certain other items. So when Hare says that a good *X* is an *X* better than they usually are, we should take him to mean *better than the items in the comparison class*. These needn't include *actual* things or people (thereby allowing that most cricketeers can be good).

Hare believes that the *criteria* of goodness and betterness are determined by standards. He thinks of standards as principles, and of principles as 'universal imperative sentences' (ibid., p. 136). An example is 'Never (or never under certain conditions) say what is false' (ibid., p. 56). When it comes to things like apples, Hare thinks of standards as principles for choosing objects of the relevant kind:

> 'To teach a person – or to decide on for oneself – a standard for judging the merits of objects of a certain class is to teach or decide on principles for choosing between objects of that class. [...] If I say 'That isn't a good motor-car' and am asked what virtue it is, the lack of which makes me say this, and reply 'It isn't stable on the road', then I am appealing to a principle.' (Ibid., p. 134)

What might this principle look like? One might think it is this: 'If one chooses a car, then choose one which is stable on the road'. But, actually, the principle should be of the maximizing form, like 'If one chooses a car, then choose the most (or more) stable one'. The reason is as follows: suppose we have a car which is more stable than another, although neither meets the threshold for stability. In this situation, the rule: 'If one chooses a car, then choose one which is stable on the road' does not explain why Hare's conditional 'ought' is true ('If one is choosing an *X*, then, if one chooses *B*, one ought to choose *A*'). After all, since neither car is stable (full stop), the rule does not require that one chooses either car. This is why the Harean analysis of 'better' requires the imperative to be of the maximizing form.

If we substitute Hare's analysis of 'a good *X*' as 'an *X* better than they usually are' with 'an *X* better than the ones in the comparison class', we obtain the following analyses of 'better', 'good' and 'best' (assuming our standard-relational view about 'ought'):

> **Better:** '*X* is a better *Y* than *Z*' is true just in case, if *X* and *Z* are the only options, one ought to choose *X* over *Z*; i.e. the contextually salient system of rules requires one to choose *X* over *Z*.

Good: '*X* is a good *Y*' is true iff *X* is better than at least some[50] members of the relevant comparison class; i.e. *X* is such that for at least some members of a comparison class, the contextually salient system of rules requires us to choose *X* over that member.

Best: '*X* is the best *Y*' is true iff one ought to choose *X* over all other relevant *Y*s; i.e. the contextually salient system of rules requires us to choose *X* over any other relevant *Y*.

Hare's analysis allows that several apples (not just the best ones) can be good, since many apples can be better than some members in the comparison class. Furthermore, it allows (something of) a distinction between the meaning of 'good' and the meaning of 'best'. Whereas 'good' means 'better than the items in a certain comparison class', 'best' means 'better than *all* relevant items' (including the good ones). If the standards require us to choose the better of two or more options, then this view allows that only the best *X* is the one we ought to choose (namely when we consider all relevant items).

Notice that on this analysis the semantic difference between 'good *X*' and 'best *X*' lies only in the items we ought to choose *X* over. There is no difference in the

[50] This makes it possible for the comparison class to contain some other good items, not just bad ones. It is not advisable to add that *X* is not worse than any item in the comparison class, since *X* may not be best or even optimal.

function of either word. Both serve to indicate that some options are to be chosen over others.

This means that there is much less of a difference between (1) and the current analysis of 'good' than at first appeared. According to the first analysis of 'good', '*X* is a good *Y*' is true iff the contextually salient system of rules requires one to choose *X*. According to the current, '*X* is a good *Y*' is true iff the contextually salient system requires one to choose *X* over certain other items. These are in fact notational variants, since if the standards of goodness are of the maximizing form ('Choose the most *F*!'), then whenever (a) there is more than one option and (b) the system requires us to choose *X*, it will also be true that the system requires us to choose *X* *over* the other options.

Given that there is no significant difference between (1) and Hare's analysis, we should revisit the problems I raised for (1). The first was that, intuitively, only the best *X* is the one we ought to choose. It is now clear that there is a sense in which both (1) and Hare's analysis allow this: they allow it by means of a shift in the items considered. Although it is true that we ought to choose a good *X* over certain items, it is not true that we ought to choose a good *X* when we consider *all* relevant items. If we consider everything, we ought to choose only the best. So the first problem I raised is not at all decisive.

The second problem was that it is not obvious how to distinguish between the meaning of 'good' and 'best' if '*X* is a good *Y*' means that the contextually salient system or rules requires us to choose *X*. But we now know how to proceed: whereas

'good' means 'to be chosen over the contextually salient subset of the options', 'best' means 'to be chosen over *all* relevant options'.

My third objection to analysing 'good' in terms of 'ought' was that it makes sense to say that one ought to choose a certain apple *because* it is good. But on both (1) and Hare's analysis, this really amounts to saying that one ought to choose the apple because one ought to choose it. And that sounds wrong. I suggested that the sense of dissatisfaction arises from the fact that 'a good *X*' seems to indicate something about *X*'s nonevaluative properties.

For Hare himself, the latter problem does not arise. He claims that 'good' has both *evaluative* and *descriptive* meaning. The first consists in the fact that it is used to *commend* (at least in assertoric contexts). But the descriptive meaning of 'good' consists of truth conditional information about the non-evaluative properties of *X* ((1952), chapter 7). The problem does arise for us, for the following reason: the standard-relational theory says that 'ought' means 'is required by the contextually salient system of rules'. *That*'s the descriptive meaning of 'ought', and there is nothing about the nonevaluative properties of *X* in this at all. So if 'good' is analysed in terms of 'better', and 'better' in terms of 'ought', then we don't get the requisite information about *X*'s properties. So it might be worth looking into a third option.

5.5 'Good' defined in terms of satisfying standards

Presumably, Hare would have agreed that 'good' and 'best' carry distinct descriptive meaning: 'best' presumably indicates that certain properties are displayed to the

highest, as opposed to some sufficient degree. But if so, we could analyse 'good' and 'best' directly in such terms (or related ones).

Of course, it wouldn't be plausible to analyse 'good' in terms of the degree to which particular non-evaluative properties are displayed (by some object). For then the meaning of 'good' as applied to apples would have nothing in common with the meaning of 'good' as applied to cars or dancers. There is after all nothing that their respective non-evaluative properties have in common. But it might be possible to analyse 'good' in terms of *satisfying standards*. Although the standards for apples are different than the standards for cars and dancers, good apples have something in common with good cars, namely *that* they satisfy the standards for these things.

Analysing 'good' directly in terms of satisfying standards is probably what David Wong intended in the first place. We saw that Wong analyses 'good' in terms of satisfying the appropriate standards. 'Better' he defines in terms of satisfying standards to a greater degree:

> 'When *X* is a better *Y* than *Z*, under conditions *C*, *X* satisfies to a greater degree the appropriate standard that applies to *X* and *Z*.' (Wong (1984), p. 71)

Wong uses the notion of satisfying *to degrees* in the case of 'better' but not in the case of 'good'. '*X* is a good *Y*' is defined as '*X* satisfies the appropriate standards for *Y*s'. But what does this mean? Does it mean satisfying to the highest degree? That seems wrong, for then no apple that is not the very best can be good. Similarly, it could hardly mean satisfying to any non-zero degree, because then even really bad

apples would be good. So I take it 'good' should mean: 'Satisfies the appropriate standards to some sufficient degree'.

These problems do not arise if the standards involve the positive form of gradable adjectives, like 'stable', 'efficient', etc. ('Choose a stable car!', 'Choose an efficient hammer!'). But if we wish to retain the idea that only the best *X* is the one we ought to choose, then we need the standards to have the maximizing form (for our theory of 'ought' requires a rule which entails that we choose only the *F*-est *X*). It also seems preferable to allow that out of two choices, the better one ought to be chosen. This latter 'ought' is true only if there is some maximizing rule, requiring us to choose the most *F*.

The need for maximizing rules to ground 'ought' claims does, however, lead to problems. If a standard says: 'Choose the most stable *X*', then only the most stable *X* satisfies the standard at all. This is because *being the most F* is not a property that comes in degrees. So if the standards are maximizing ones, we need to explain what it means to satisfy a standard to some non-maximal (and non-minimal) degree. Furthermore, since standards are thought of as imperatives, and imperatives are directed at agents, only humans can conform to or satisfy the standards. But we also want to say that *cars* can satisfy standards.

What to do about these problems? John Searle said that a prominent meaning of 'good' is 'meets the criteria or standards of assessment or evaluation'. Note the use of 'criteria'. If one is asked to state criteria for a good *X*, one responds by listing features. For example, criteria for a good car might be: stable on the road, reliable, comfortable, eco-friendly, cheap in maintenance, etc. This suggests the following:

when we say that a car is good, we are not saying that it conforms to a rule which tells agents to choose cars with certain properties. Rather, we say that it satisfies to sufficient degrees the criteria associated with the relevant rules. If we take criteria to be properties, we can think of satisfying a criterion as exemplifying (to a sufficient degree) the relevant property.

In order to allow grades of satisfaction, we should say that the criteria associated with rules are not lists of properties expressed in superlative form. If the relevant rule is 'Choose the *F*-est *X*!', the associated criterion for *X*es is *F-ness*, not *F-estness*. Since a thing can be more or less *F*, a thing can satisfy the criteria associated with the rule to a greater or lesser degree.

According to our current theory of 'good', '*X* is a good *Y*' means '*X* satisfies to a sufficient degree the criteria for *Y*s associated with the contextually salient system of rules'. This would be circular, if 'satisfying to a sufficient degree' could only be analysed as 'satisfying to a degree which makes it good'. But that isn't so. The sufficient degree might simply be some contextually salient threshold value. With properties expressed by gradable adjectives, the threshold value is determined by a salient comparison class.

Like Hare's, this analysis of 'good' still requires imperatives to ground claims about goodness. In my example an apple is good because it satisfies to a sufficient degree the criteria associated with hypothetical imperatives (like 'If one chooses an apple, then choose the juiciest one'). This means that the apple exemplifies to a sufficient degree whatever properties are mentioned (in superlative form) in the consequent of the imperative. To say that an apple is better than another is to say that

the first exemplifies the properties mentioned in the rules to a greater degree than the latter.

Of course not all properties come in degrees. The property of being a certain age does not, and neither do being rectangular, having siblings and being pregnant. When a rule mentions only one such yes/no property, satisfying the criteria to a sufficient degree can only mean *having the property at all*. In this case, things either satisfy the criteria or they don't. This means that one can say that a thing is better than another (relative to the rule) only in the sense that the one does and the other does not have the property. In this sort of case, it doesn't make sense to judge that some things, all of which are good, are better than others. But that is how it should be with nongradable properties. If the rule requires us to pick people of a certain age, then everyone who has it is (in this respect) equally good. (Of course there might be other features that matter to us, but that does not detract from the point.)

So I propose we take 'satisfying to greater degrees' to cover cases where good-making properties come in degrees as well cases where they don't. In the latter case a thing satisfies the relevant criteria to a greater degree than another iff the first has, and the second lacks, the relevant properties.

Let's recap. I have proposed the following truth conditions for '*X* is a good apple':

> **Good:** '*X* is a good apple' is true iff *X* satisfies to a sufficient degree the criteria for apples associated with the contextually salient system of rules.

'Better' is treated as follows:

> **Better:** '*X* is a better apple than *Z*' is true iff *X* satisfies to a greater degree than *Z* the criteria for apples associated with the contextually salient system of rules.

And 'best' is dealth with thus:

> **Best:** '*X* is the best apple' is true iff, compared with certain items, *X* satisfies to the greatest degree the criteria for apples associated with the contextually salient system of rules.

As we've seen, 'satisfying criteria to some degree' is itself analysed in terms of exemplifying properties. In the case of properties that come in degrees (like juiciness and efficiency), *X* satisfies to a greater degree the criteria for *Y*s than *Z* iff *X* exemplifies to a greater degree than *Z* the properties whose highest grades are mentioned in the rules. In the case of properties that don't come in degrees (like being of a certain age), *X* satisfies to a greater degree the criteria for *Y*s than *Z* iff *X* exemplifies the properties figured in the rules and *Z* does not.

The latter entails that in the case of nongradable properties, the class of good *X*es completely overlaps with the class of best *X*es. But this is as it should be. If all that matters is that *X* is of a certain age, then all *X*es which have it are not only good, but also (in respect of age) the best ones in the set. Perhaps it sounds a little odd to say this, but if so, the impression could be explained by the fact that 'best' connotes

uniqueness.[51] Alternatively, 'best' may connote *gradability*. This could be due to the fact that most of our evaluations are based on gradable properties. However this may be exactly, I don't think it is a threat to this analysis of 'good'.

The current analysis of 'good' has the downside of being hybrid. It requires imperatives and lists of associated properties (criteria). Furthermore, the list of properties can't simply be a list of properties mentioned in the rules. For we often need minimizing and maximizing rules (requiring us to choose the least *F* or the most *F*). But the property of being the least or most of something does not come in degrees. So we need the criteria to be expressed in positive form. None of this is very pretty.

On the upside, the theory explains why '*A* ought to *X because X* is good' seems informative. It seems informative because it means that *A* ought to *X* because *X* satisfies to a sufficient degree the criteria for *X*es. However, there seems to be a problem which afflicts all of the analyses of 'good', 'better' and 'best' discussed.

5.6 Goodness in virtue of various criteria

One might argue as follows: the analysis of 'good' in terms of satisfying criteria seems to work for 'better', 'good' and 'best' *in a single respect*. If all that matters is juiciness, then the juicier the better. But judgments about goodness and betterness can also be made on the basis of several qualities. A hammer may be best because it has a firm

[51] Although we do say things like: 'John and Sue are the best dancers in the class', it would be misleading to say: 'John is the best dancer' if John is no better than Sue.

grip, a hard head and is relatively light. Another, less good hammer may also have a firm grip and a hard head, but be too heavy to handle. It might be suggested that the analysis fails to specify truth conditions for this type of judgment.

One might further reason that it is not obvious how to specify these truth conditions. They are not the following: '*X* is all-in the best hammer' is true iff *X* satisfies to the greatest degree *all the criteria* associated with the contextually salient system of rules. At least this is not correct if by 'satisfying to the greatest degree *all* the criteria' we mean that each individual criterion is satisfied to the greatest degree. A hammer may be best even if it less light than another hammer (its other properties may make it superior).

Nor can we say that '*X* is all-in the best hammer' is true iff *X* satisfies to the greatest degree the criteria *on average*, where this means the following. Suppose we have three relevant criteria. *X* satisfies 1 and 2 completely, but 3 only to a minimal extent. *Y* satisfies 1 and 2 to a minimal extent, but 3 completely. On average, *X* satisfies the criteria to the greatest degree. But *X* need't be better than *Y*, if criterion 3 is far more important than 1 and 2 taken together. So for all-in goodness, we must take into account the weighting of criteria.

This reasoning seems plausible to me, although it might be based on a mistake. Take the example of the hammer which is best in virtue of its combination of properties (having a firm grip, a hard head, and being light). This combination makes a hammer good because the properties combine to make the hammer more conducive to *a single aim*: the driving in of nails. So the hammer is, after all, best in a single

respect: it is the most efficient one. We only need one rule to ground this sort of judgment ('Choose the most efficient tool'!).

It may be that we can't really make sense of goodness in more than a single respect. If we judge that an object is best in virtue of a combination of properties, there may always be a further criterion which that combination fulfills better. But the appearances do count against this. It seems possible to say that an act was morally best in virtue of two distinct criteria. Or that an option was better than others because it was both prudent and morally right. It is not clear what single criterion both of these would help to fulfill. Furthermore, it seems possible that the act is not best in virtue of any one of these, but only in virtue of the combination.

It is not obvious how to make sense of this on a standard-relational theory of the type I am considering. The theory of 'ought' discussed in the previous chapter says that we ought to *X* iff *X* is required by some rule in the contextually salient system, and no rule with the same or higher priority requires not-*X*. If the best *X* is the one we ought to choose, then it seems we should define the *overall best* in terms of the satisfaction of (at least some) criteria associated with rules highest up the order of priorities. But *how* exactly?

Suppose we only have two relevant rules. The first, *R1*, requires us to choose the *F*-est *X*. The second, *R2*, requires us to choose the *G*-est *X*. Further suppose that we have two objects. O_1 exemplifies *F*-ness to a greater degree than O_2, but O_2 exemplifies *G*-ness to a greater degree than O_1. Suppose that O_1 is better than O_2, not because it is more *F*, but because it is more *F* and also a little bit *G*. Had it not been *G* at all, O_1 would not have been better than O_2. What makes this the case?

The following does not: a higher-order rule which requires us to choose in accordance with both rules. If O_2 is more G than O_1, then R_2 entails that we choose O_2. Acting in accordance with both rules is impossible, and does not generate the requirement to choose O_1.

Another possibility is to say that the requirement is generated because, in this situation, R_1 ('Choose the F-est X') is higher up the order of priorities than R_2. But that is not, intuitively, what is going on. O_1 can be better than O_2 not because F-ness is more important than (takes priority over) G-ness. It might be very important that O_1 is G to that limited extent.

It seems to me that the only way to generate the requirement is by means of a rule which tells us to choose an X which is most F, and to some extent G (i.e. to exactly the extent required). Notice that such highly situation-specific rules are required on all previous standard-relational analyses of 'good', 'better' and 'best'. All of them presupposed the truth condition for 'ought' statements developed in chapter 4. So in all situations in which an object is best in virtue of some weighted mixture of properties, it can be true that we ought to choose the object with that mixture only if there is a rule whose content is to choose the object with exactly that weighted mixture.

And the problem is not restricted to goodness. It also affects the theory of 'ought'. An act can be morally best in virtue of contributing to differing degrees to different moral aims. So it seems that some 'ought' judgments will also require highly situation-specific rules. Furthermore, these rules will have to be first-order, since no ordering of simple first-order rules will do the trick.

This looks bad. The theory now depends on many highly situation-specific rules (not just higher-order ones). Some rules (like 'Choose the *X* which is *F* to degree *n* and *G* to degree *n*+1) cannot be constructed out of the materials we already had: a hierarchy of simple rules, like 'Choose the *F*-est *X*' and 'Choose the *G*-est *X*'.

But we need to be clear on *why* this is so bad. Part of our resistance to situation-specific rules might derive from the sense that rules have to do some *psychological* explanatory work. For example, we may think they have to guide people's choices. Highly situation-specific rules cannot do this. But does that matter in the context of a *semantical* inquiry? Why couldn't rules be merely theoretical entities that only play a role in the truth conditions of normative statements?

It is hard to say exactly what is wrong with this suggestion, but I think we should reject it nonetheless. Other things being equal, we should assign truth conditions to utterances on the basis of what they are sensitive to, what they closely vary with, in the way in which cow-judgments are sensitive to cows (see chapter 1, section 1.9). This co-variation is asymmetric in the sense that detection of the relevant entities (cows) is what explains our willingness to affirm certain judgments (like 'There is a cow over there'). But it seems that situation-specific rules are *post hoc* postulates: they are postulated on the basis of the fact that people actually reach a verdict about what is best or what they ought to do. These verdicts are not *sensitive* to the rule (not explained by their implicit or explicit detection).

And *post hocery* is not the only problem. If (many of the) first-order rules are highly-situation specific, then the trouble with reasons that we encountered in the previous chapter returns. A rule like 'Choose the *X* which is *F* to degree *n* and *G* to

degree *n*+1!' does not entail that you choose an *X* which is *only F* to any degree. So the fact that an object is to some degree *F* won't be a reason to choose it. That sounds wrong. The objection can't be answered by shifting the class of objects to which the rule is applied. This manoeuvre was available for a rule like: 'Maximize your profit!' to account for the sense that we have some reason to choose 10 pounds instead of 50 (see chapter 4, section 4.7). This move is not available here because of the highly situation-specific content of the rule. So it seems the standard-relational theory will not generate enough reasons after all.

Perhaps there is a way to avoid this. We could use a move suggested by Finlay for his own theory of normative language (discussed in the next chapter). Finlay proposes that some 'ought' judgments (and presumably judgments about goodness) are made relative to the end of maximizing expected utility.

In (some versions of) expected utility theory (EUT), the expected utility of an act is determined by two factors (1) the value of its associated outcomes and (2) the probability of these outcomes conditional on the act's choice.[52] [53] The expected utility

[52] For an introduction to expected utility theory, see Joyce (1999). See also chapter 8 for a little more detail.

[53] So expected utility values are determined by the values of *outcomes*. How does this relate to the standard-relational idea that the *properties* of acts and objects matter to their goodness? EUT clearly involves the *consequences* of acts, weighted by their probabilities. But this is not a problem. There is no reason why standards could not involve relational properties, like having certain consequences. But EUT is not incompatible with assessing properties (like being to some degree comfortable or cheap) instead of outcomes either. Even if the likelihood that the object exemplifies these properties is

of an act is the product of these two values (the value of associated outcomes times their probability). The higher the product, the higher the expected utility.

The "value" of an outcome is often taken to be determined by the subject's desires (the more you desire an outcome, the more value it has). But it needn't be. It can also be determined by estimates of how much an outcome matters to someone else (which would in turn be determined by this person's desires).

Perhaps the standard-relational theorist can avoid the problem of highly situation-specific rules by hypothesizing that in those cases where the goodness of an option depends on several features, there is only one (very simple) rule: 'Maximize expected utility!'. But can the standard-relational theorist say that betterness in virtue of several features consists in exemplifying the property of having expected utility to a greater degree? That depends.

The property of *having* expected utility is not really a property. Having any (or some) expected utility value is a property, and so is the property of having a particular expected utility value. The first is not suitable to ground judgments of betterness because so long as two options have some expected utility value, they exemplify the property of having *an* expected utility value to the same degree. The second property - having a *particular* expected utility value - *can* be used to ground judgments of betterness. For example, if the property is that of having the *highest* expected utility value (maximizing expected utility), then a suboptimal option exemplifies this

---

sometimes 1, that needn't matter (unless the certainty would unduly raise the expected utility of an option).

property to a lesser degree in the sense that this option does not exemplify the property at all.

But sometimes we compare more than two options. We can say that option 2 is better than 1 and 3 better than 2. However, if 2 does not have the highest expected value, then 2 is not better than 1 in virtue of exemplifying the property of having the highest expected value to a greater degree than option 1. So in order to get the right results, we have to assume that the judgment that 2 is better than 1 is based on a comparison *of 1 and 2 alone*. Compared with just 1, 2 *does* have the highest expected value.

This seems a little awkward, but such moves are hard to avoid for subjectivist theories. First, we've seen that Richard Hare invokes it in his definition of 'better'. Second, Finlay's theory also needs it (or something very much like it), as we'll see in chapter 8. Third, I've argued that this sort of move might reasonably explain the sense that the ass has some reason to choose both the left and the right stack of hay (chapter 4, section 4.7). So I don't think that this problem is decisive.

I doubt that it is incompatible with a standard-relational theory that in some cases, the relevant rule is 'Maximize expected utility!'. Since expected utility is a relation between strength of desire and probabilities, this means that the rule is implicitly about the speaker's mental states (at least in those cases where the utility values are determined by the speaker's desires). This limits the amount of genuine disagreement that the theory allows (see chapter 4, section 4.5). Whenever speaker 1 refers to a rule which requires the maximization of *speaker 1's* expected utility, and speaker 2 refers to a rule which requires the maximization of *speaker 2's* expected

utility, they cannot disagree about the same proposition. It also means that statements about betterness are sometimes implicitly about the relation in which options stand to the speaker's mental states. Although these are costs, the theory had to reckon with the absence of genuine disagreement anyway. It did so by accepting that the point of normative language is not merely to get the facts right, but also to protect and promote certain values. So this cost may be accepted.

A second problem related to the use of expected utility is that it seems to restrict the number of reasons in a bad way. For example, if the only relevant rule is to maximize expected utility, and reasons are facts which explain why a relevant rule entails that the agent acts a certain way, then the only reasons that exist (in this situation) are facts like: *that X has a certain expected utility*. But intuitively, facts like *that a car is cheap* (or cheaper than some other car with which it is compared) are reasons even in contexts where several features matter.

Although this is true, we could reasonably maintain that rules like 'Choose the cheapest car!' are made salient by the fact that expected utility involves people's desires with respect to various properties or consequences. These rules can explain why being cheaper than some other option is a reason to choose it even in a context where the dominant rule is to maximize expected utility.

So far then, the standard-relational theory holds up fairly well. A more serious issue arises once we look at the relation between 'ought' and 'must', which also made trouble for Humeanism.

5.7 'Must' and 'ought'

Linguists point out that 'must' seems stronger than 'ought'. Since 'have to' appears to behave in the same way as 'must' (and serves in its place in past-tense constructions, like 'I had to do it'), examples involving both can be used to substantiate the claim. Von Fintel and Iatridou (2008) give the following examples:

(1) Everybody ought to wash their hands; employees must.

(2) You ought to do the dishes, but you don't have to.

(3) You ought to wash your hands – in fact, you have to.

'Must' seems to indicate some kind of necessity. Perhaps you ought to use a screwdriver, but you *must* give back the money. 'Must' seems stronger than 'ought' at least in the sense that 'must' appears to entail 'ought', but not the other way around. Can the standard-relational theory explain this?

It is hard to see how the theory could account for the difference on the level of truth conditions. After all, '*A* ought to *X*' means that the contextually salient system of rules requires that *A X*-es. What can be stronger than *being required*?

Perhaps one could say that *A must X* in *S* if and only if *A* ought to *X* in *S* and there are no reasons for *A* not to *X* in *S*. But we don't always say that you *must* do whatever is such that there are no reasons against it. Furthermore, it seems perfectly possible that one *must* do something even if there are some reasons not to.

So I think that the standard-relational theorist has to say that the truth conditions of '*A* must *X* in *S*' are identical to the truth conditions of '*A* ought to *X* in *S*'. In order to explain the sense that 'must' is stronger than 'ought', s/he could say that 'must' carries a conventional implicature (e.g. the implicature that the speaker insists on it).

This hypothesis can make reasonable sense of (1) and (2). Although not incompatible with (3), (3) seems to say more (or can be used to say more) than that the speaker *insists* on washing your hands. Some actions seem instrumentally *essential*. E.g. 'In order to evade arrest, Max *has to* mingle with the crowd'. This indicates that it is *necessary* for Max's evading arrest that he mingles with the crowd (there is no other option). But the speaker can insist on acts whether or not they are in this sense essential.

So the idea that 'must' is stronger than 'ought' because it carries a conventional implicature is not quite satisfactory.

5.8 Normative and epistemic 'ought's

There is a second reason to doubt the standard-relational theory of 'ought'. Stephen Finlay (and other before him, like Sloman (1970) and Kratzer (1977)) has drawn attention to the fact that many modal auxiliary verbs like 'must', 'ought' and 'may' are ambiguous between normative and nonnormative readings (2009). For example: 'John may go to his cell' can both be read as a normative statement (he is allowed to go to his cell) or as a statement about what is compatible with the evidence. Similarly, 'John

ought to return tomorrow' can either be a normative or a nonnormative statement. Furthermore, this ambiguity is found in the counterparts of English modals in many other European languages (Kratzer (1977), Von Fintel & Iatridou (2008)). This suggests that it is not accidental like the ambiguity of 'bank', which can be used to mean both 'river edge' and 'financial institution'. It seems plausible that normative and nonnormative modals have a common core which generates both readings.

If we say that the normative 'ought' means 'is required (i.e. entailed) by the contextually salient system of rules', then what are we to say about the epistemic 'ought' in 'It ought to rain tomorrow'? It seems that this also has to be concerned with requirements. We could say that 'It ought to rain tomorrow' means that its raining tomorrow is required (entailed) by the evidence. But this seems too strong. What ought, epistemically, to happen is not always *necessary*. It seems that what ought, epistemically, to happen, is what is *favoured* (perhaps above a certain threshold) by the evidence. It is not clear how to make sense of this on a model where the core meaning of 'ought' consists of a relation of requirement.

5.9 Conclusion

In this and the previous chapter, I've discussed the idea that normative statements implicitly refer to standards, where these are thought of as imperatives or rules. Views of this type have been (and are) defended by David Wong, David Copp, possibly Gilbert Harman and myself. Richard Hare is also committed to certain aspects of standard-relational views. According to the best version of the standard-relational

theory, normative statements refer to systems of relatively simple rules. A multiplicity of rules was needed to generate enough reasons, because reasons were (naturally) thought of as facts which explain why a relevant rule entails that the agent acts a certain way. There were several ways of defining 'good', 'better' and 'best', but the most promising way involved lists of properties which a good *X* displayed to a sufficient degree.

The standard-relational view had the advantage of being able to account for the apparent entailments between 'ought' and the balance of reasons (if in a relatively trivial fashion). The strength of a reason was determined by the requirements of rules under certain (counterfactual) conditions: a reason R_1 is stronger than another reason R_2 just in case the act for which R_1 is a reason would be required in case of conflict.

The theory has certain downsides. For example, it requires us to deny that Buridan's ass has any reason to choose either stack of hay. However, such problems are not clearly strong enough to reject the theory. But there are more serious complications: first, the theory cannot plausibly account for the distinction between 'must' and 'ought'. Second, it seems there is a common core shared between 'ought' in normative and epistemic uses. But the standard-relational theory of the normative 'ought' makes it difficult to see what the normative and epistemic 'ought' might have in common.

These considerations lead me to consider yet another option in normative semantics: Stephen Finlay's end-relational account. We will see that it is superior to all the theories discussed so far.

Chapter 6. The end-relational theory of normative language

6.1 Introduction

In previous chapters I've discussed two versions of Humeanism and standard-relational theories of normative language. We've encountered some serious difficulties. In this chapter, I describe Finlay's end-(as opposed to standard)-relational theory of normative discourse. Since the theory is rich and complicated, I will postpone a critical assessment until later chapters.

6.2 The end-relational theory of instrumentally normative modals

Finlay is aware that 'ought' is but one in a family of related *modal auxiliary verbs*, including 'have to', 'needs to', 'must', 'may', 'might', 'can', 'could' and 'should'. Except for 'ought' and 'should', it is uncontroversial that these are propositional operators. This gives us reason to believe that 'ought' is also one. Just like 'Joe must be 30' can be analysed without loss or addition as 'It must be the case that Joe is 30', so 'Joe ought to stop smoking' can be analysed without loss or addition as 'It ought to be the case that Joe stops smoking'.[54]

[54] Some philosophers believe that 'ought' cannot always be a propositional operator, because that view entails that the logical forms of (1) 'Lexy ought to kiss Lionel' and (2) 'Lionel ought to be kissed by Lexy' are identical, namely: 'O(Lexy kisses Lionel)' (Finlay ms, chapter 4, p. 7). However, it seems that (1) is (roughly) about what Lexy has reason to do, whereas (2) is compatible with Lexy having no

Finlay observes that most modal auxiliaries have both normative and nonnormative readings (with the exception of 'needs to').[55] Take:

(1) Edward has to go to his cell.

(2) Edward may go to his cell.

These can both be read as claims about what is ruled out/in by the evidence (epistemic) or as claims about what Edward is required/permitted to do (deontic). The fact that this ambiguity exists in many languages other than English is strong reason to suspect that it is not accidental, like the ambiguity between 'bank' meaning river edge and 'bank' meaning financial institution. In other words: there is probably a common core involved in both (a point also made by Kratzer (1977)).

In order to find this semantic core, Finlay starts with standard analyses of modal vocabulary in linguistics and philosophy:

> 'Modals are perspicuously analyzed as quantifying over possibilities; necessity modals like 'have to' basically mean *in all possibilities*, while possibility modals like 'could' basically mean *in some possibilities*.' [...] The most obvious and

---

reason whatsoever to kiss Lionel. However, Finlay makes a good case that the personal and impersonal readings *can* be generated on the view that 'ought' is always a propositional operator (ms, chapter 4, pp. 8-9).

[55] Angelika Kratzer also makes this point in (1977).

> simple way of generating different kinds of modality is then by defining (i.e. *restricting*) the domain of quantification or possibility-space *W* in different ways. To be *logically* necessary is to be true in all logically possible world-states, to be *physically* necessary is to be true in all the physically possible world-states, to be *epistemically* necessary for some subject *s* is to be true all the world-states that aren't ruled out by *s*'s information. We would therefore analyze *normative* necessity as truth in all normatively possible (i.e. permissible) world-states. The semantic content of any modal word - its common meaning across all its uses – is then the quantificational operation it performs on its object proposition *p* relative to *W*.' (Finlay ms, chapter 4, p. 10)

So modals can be thought of as propositional operators. They say that some proposition is true in all or some of the relevant possibilities. Since the relevant possibilities can differ, modals have an argument place for a "modal base". Finlay thinks of the modal base as a set of propositions given the truth of which things are possible or necessary.

In the case of normative modals, the modal base includes propositions that describe ends. Ends are possible states of affairs (Finlay (2009), p. 319). Opting for ends as constituents of the modal base is very flexible, because this allows "generic" ends like acting in accordance with systems of rules, but also "specific" ends like evading arrest (the latter is not in any obvious sense a *standard* or *rule*.)

What one normatively must or may do is clearly not relative only to a set of ends. It also depends on specific information about an agent's circumstances. For example, 'whether one may drive home from work depends on what you've been drinking; what Max ought to do if he wants to evade arrest depends upon a host of details about the world around him and his own capabilities' (Finlay ms, chapter 4, p. 11). Furthermore, I've already noted (chapter 2) that normative 'ought's are plausibly relative to the information available to a subject. If the police's evidence suggests that the hostage is held in the Hilton, then they ought to concentrate their efforts on the Hilton. But an onlooker who knows that the hostage is actually in the Ritz can truly say that they ought to focus on the Ritz.

Finlay proposes to put all these relativizations into the modal base, so that we need only one argument place ('*B*' for 'modal base') in our analyses of modals. We can then analyse 'must' and 'may' as follows (ms, chapter 4, p. 11):

$must_B(p)$: In all world-states in W_B, p;

$may_B(p)$: In some world-states in W_B, p.

This idea, applied to hypothetical imperatives, leads to a reduction of instrumentally normative modals. Take the sentence:

(1) In order to evade arrest, Max has to mingle with the crowd.[56]

[56] The example is by Finlay.

If $must_B(p)$ is correct, the meaning of this sentence can be represented by:

(1*) In all world-states in which Max evades arrest, he mingles with the crowd.

Or can it? What if Max has fear of crowds and would *never* mingle with the crowd? It might still be true that he *has* to in order to evade arrest. All this shows, however, is that normative uses of modals like 'must' and 'have to' are not (always) relative to a modal base which includes information about the agent's psychological dispositions. Sometimes, we are interested only in the agent's *external circumstances* ((2009), p. 323; ms, chapter 4, p. 16).

But there is another problem: (1*) seems too weak. It might be true that in all world-states where Max jumps off the roof, he hits the ground hard. But it's not true that *in order to* jump of the roof, Max has to hit the ground hard. Shouldn't there be some indication that mingling with the crowd is a *causal* condition on evading arrest? Finlay argues that it suffices to suppose that (1*) contains an implicit temporal marker, as in:

(1**) In all world-states in which Max evades arrest, he *first* mingles with the crowd.

According to Finlay, (1**) 'suggests metaphysical priority (that Max evades arrest *because* he mingles with the crowd) as well as temporal priority' (ms, chapter 4,

p. 17). The reason is pragmatic: it would be misleading to state the conditional if one didn't want to suggest that Max's escape somehow depends on his first mingling with the crowd. In that case, one could restrict oneself to asserting the consequent. So Finlay believes that Gricean maxims ('Avoid prolixity!') explain why one needn't insert 'because' into the analysis of instrumentally normative conditionals.[57]

Since (1**) appears to state necessary and sufficient conditions for the truth of (1), Finlay believes it amounts to a reductive analysis of the instrumentally normative 'must' (and therefore 'have to', which means the same). The instrumentally normative 'may' can be correspondingly reduced. The meaning of

(2) In order to evade arrest, Max may mingle with the crowd.

can be represented by:

(2*) In some world-states in which Max evades arrest, he first mingles with the crowd.

[57] Some means are not *temporally* prior to the end. For example, in order to murder someone, you have to kill him. Killing is *constitutive* of murder, and thus not temporally prior to it. But it is *metaphysically prior* in that there is no murder without killing, but there is killing without murder. Finlay believes that we also use temporal markers (whether encoded in the grammar or in dedicated words like 'first') to indicate constitutive dependency.

Finlay notes that normative readings of instrumental conditionals depend not just on presenting the consequent as temporally or metaphysically prior to the end's attainment (the *temporal ordering condition*). It also depends on other features of the grammar ((2010a), sections 3 & 4; ms, chapter 4, section 5). Take the following two sentences (from (2010a)):

(3) If Macbeth became king, he *had to have killed* Duncan.

(4) If Macbeth was going to become king, he *had to kill* Duncan.

Whereas (4) is intuitively normative, (3) isn't. But the event in (3)'s consequent is clearly presented as temporally prior to the event in the antecedent.

Finlay believes it is essential to the normativity of (4) that the antecedent contains the prospective 'going to'. Linguists tell us that the function of this phrase is to ascribe to someone at a time t_1 the property of doing something at a later time t_2:

> 'When at *t1* I assert, 'I am going to write a book,' I say that I am (at *t1*) going to write a book (at *t2*). Although the sentence is about a future writing-event, it also addresses what *is now* the case. What is the case about me now is that I possess a particular future, in which I write a book. We can distinguish between the *aspect-time* (here *t1*) and the *event-time* (here *t2*).' (Finlay ms, chapter 4, pp. 19-20)

It is essential to a normative reading of instrumental conditionals that the necessity of the event in the consequent (killing Duncan in (4)) is ascribed to a time prior to the event-time in the antecedent. This requirement Finlay calls the *prior necessity condition*. This condition is not satisfied in (3), since the necessity of having killed Duncan is ascribed to Macbeth at the time at which he is already king (in other words: in (3), the aspect-time coincides with the event-time).

Finlay believes that a 'must' or 'may' conditional has a normative flavour just in case it satisfies the temporal ordering and prior necessity conditions. And that seems true (I discuss two potential counterexamples in the next chapter). So instrumentally normative conditionals may well allow reduction.

These analyses give us the resources to analyse other normative modals as well. 'Needs to' behaves the same as 'must' (except that it does not have a nonnormative use), and 'might', 'could' and 'can' behave the same as 'may'. For example, in

(5) If you are to make it to Long Beach, you might take the Harbor freeway. (Finlay (2010a), p. 80)

you can replace 'might' with 'could' or 'can' without change of meaning. And (5)'s meaning can be represented by

(5*) In some relevant worlds where you make it to Long Beach, you first took the Harbor freeway.

But what about 'ought'?

6.3 The end-relational theory of the instrumentally normative 'ought'

Finlay's analysis of 'must' and 'may' started with analyses of their nonnormative use. This strategy led to a simple, unified account of the semantics of modals in normative and nonnormative uses. Can the same be done for 'ought'? Finlay thinks it can.

Before I go into this, we should repeat that 'ought' seems weaker than must (see chapter 5, section 5.7). Linguists call it a 'verb of "weak" necessity' (Finlay ms, chapter 4, p. 25; also Kratzer (1977)). 'Must' is a verb of strong necessity:

> 'Whereas 'Max must mingle with the crowd' represents this action as the *only* possibility, 'Max ought to mingle with the crowd' admits other possibilities – that there are other things that Max *could* do instead – and picks this out as the *best* of the possibilities.' (Finlay ms, chapter 4, p. 25)[58]

So it is not plausible to analyse 'ought' in terms of the requirements of rules, as we tried in chapter 4. For this collapses the distinction between 'must' and 'ought', as argued in chapter 5.

[58] Sloman (1970) makes this point as well.

Since 'ought' is also ambiguous between normative and nonnormative uses (like other modals), Finlay seeks its core meaning in the epistemic 'ought', displayed in:

(6) It ought to rain tomorrow.

The 'ought' in (6) is often interpreted in terms of probability, as follows:

(6*) Given *B* (the modal base), it is more likely that it rains than that any relevant alternative transpires.[59]

So the epistemic 'ought' can be analysed as follows (where '*B*' is the modal base and '*R*' a class of relevant alternatives):

$Ought_{B,\,R}(p)$: Given *B*, *p* is more likely than any *r* in *R*. (Finlay ms, chapter 4, p. 26)

If we think of probability as a measure of possibility space (see the appendix for details), we can integrate this analysis with Finlay's theory of modals as operators quantifying over possibilities:

[59] This rendition of the epistemic 'ought' is based on Finlay (2009), p. 323 and ms, chapter 4, p. 26.

Ought$_{B,\ R}$(p): For all r in R, p is true in a greater proportion of the possibilities in W_B than r. (Ibid., p. 26)

Now for instrumentally normative uses. Take:

(7) In order to evade arrest, Max ought to mingle with the crowd.

Following the same strategy as for the analysis of 'must' and 'may', (7) can be analysed as meaning:

(7*) Given *B including that Max evades arrest*, it is more likely that he first mingles with the crowd than any alternative *r*.

If we think of the proposition *that Max evades arrest* as q and the proposition *that he first mingles with the crowd* as p, we can state a general analysis of instrumentally normative 'ought' conditionals as follows (ibid., p. 26):

$\Pr(p|q) > \text{pr}(r|q)$, for all $r \in R$

(where 'q' should really be 'B including q').

This may seem wrong, because there might be a difference between what an agent (with particular psychological features) is most likely to do, and what is *best* or what s/he *ought* to do in order to achieve some end. But we already saw the solution to

this problem in the case of 'have to': what Max *has to* do in order to evade arrest is not sensitive to psychological information about Max. Likewise for 'ought': the modal base relative to which we assess how likely he is to have mingled with the crowd excludes facts about Max's psychological dispositions.[60] This amounts to assigning equal initial probability to each potential means in *R*, which Finlay calls *symmetry of choice* ((2010b), p. 80 and ms, chapter 4, p. 27):

> 'Given the subsequent modal base, i.e. that we have symmetry of choice and that the end actually eventuates, the most likely means to be chosen will always be the means on the choice of which the end has the greatest likelihood of eventuating.
>
> To grasp this equivalence intuitively, suppose that you are aware that Max hopes to evade arrest, and also that only two means are possible: mingling with the crowd, to which you assign a probability of success of 0.9, and bolting for the door, to which you assign a probability of success of 0.1 You have no idea which he will choose, so relative to your information each choice has a probability of 0.5. Now suppose you later learn that Max indeed managed somehow to evade arrest. It would be very natural for you to say, 'He probably mingled with the crowd, then,' or, 'He ought to have mingled with the crowd, then.' (Finlay ms, chapter 4, p. 27).

[60] Strictly speaking, Finlay excludes this information from the 'preliminary modal base B_0 (the modal base prior to restriction to world-states in which the end eventuates)' (ms, chapter 4, p. 27). I don't think this detail matters to our discussion.

So we have arrived at an end-relational theory of the instrumentally normative 'ought' according to which the means one ought to take is the most reliable means, or the means most likely to realize the end:

> 'The end-relational theory offers an explanatory reduction of normative 'ought's about the actual world as being nothing other than propositions about what would be most likely given symmetry of choice and that certain ends eventuate. Judgments about what we ought to do in order to curb global warming, for example, are simply judgments about what we would probably do were we subsequently to succeed in curbing global warming.' (Ibid., p. 27)

Of course, the foregoing pertains directly to instrumentally normative 'ought's, or 'ought' as it appears in instrumentally normative conditionals, like 'In order to evade arrest, Max ought to mingle with the crowd'. But Finlay believes it is the correct analysis for *all* normative 'ought's. This includes uses that are superficially categorical (not explicitly relative to some end). What could justify this claim?

Here is how I see it: 'ought' does not appear to shift meaning in

(8) In order to evade arrest, Max ought to mingle with the crowd

and (statements like)

(9) Max ought to save his friend.

But if that's correct, and if 'ought' in (8) is a propositional operator that quantifies over possibilities in which relevant ends obtain, then this should be true of 'ought' in (9) also.

Additional pressure to accept this derives from the interrelatedness of modal vocabulary. For example, it is plausible that 'must' and 'may' are always relative to a variable modal base. This is why we accept that (i) 'All Maori children must learn the names of their ancestors' and (ii) 'The ancestors of the Maoris must have arrived from Tahiti' can be simulatenously true (the examples are from Kratzer (1977), p. 338). These uses of 'must' are surface-categorical: they do not explicitly state the modal base, or what they are relative to. But they are relative to a modal base nonetheless. Furthermore, the modal base of (i) is plausibly determined by rules of the Maori (W_B is restricted to worlds where Maori children act in accordance with the rules). 'Ought's kinship to 'must' (and other normative modals) therefore raises the probability that it too is always implicitly relational, relative to a set of worlds in which contextually salient ends obtain.

Resistance to the idea that 'ought' is always relational might derive from the influential Kantian idea that moral requirements are *categorical*. This can mean many things, but one relevant sense is that certain acts are to be done *not* because of something else to which these acts are instrumental. They are *in themselves* required. This idea, however, is perfectly compatible with Finlay's analysis. Some acts might be such that they ought to be done *in order that they are done* (and not in order that some

further end obtains). In such cases, the modal base allows only worlds in which the act prescribed is done. (For more objections related to categoricity, see chapter 7, sections 7.3 and 7.4.)

6.4 The end-relational theory of 'good'

Finlay's theory of 'ought' is sensitive to linguistic data concerning modals, and should be taken seriously for that reason alone. It also makes good sense of the difference between 'ought' and 'must'. 'Must' is stronger than 'ought' because 'must' indicates that an act is necessary, whereas 'ought' merely says it is most likely.

What does the end-relational theory say about other words, like 'good' and 'reason'? Because Finlay defines 'reason' in terms of 'good', I shall start with the latter.

In ms, chapter 3, Finlay says that:

> 'Ordinary 'good' sentences of grammatically very diverse kinds [...] appear compatible with the hypothesis that 'good' has the unified, underlying logical form made explicit in sentences of the form ERT[61], 'It is good for *E*, if *p*'.' (Finlay ms, chapter 3, p. 22)[62]

Take the sentence:

[61] This is an abbreviation for 'End-Relational Theory'.

[62] Finlay uses a lower case '*e*' for the end-variable, but I have and will replace these with upper case letters throughout the thesis.

(10) Chocolate is good.

Plausibly, this sentence is elliptical for something more complicated. We can ask: 'Good for what?'. In some contexts, the answer might be: 'It is good to eat'. But that is still not fully explicit. Chocolate might be good *for getting gustatory pleasure* to eat, but it might also be good *for shortening one's life* to eat. This is why we can think of (10)'s logical form as given by ERT:

(10*) It is good for [relevant agent's] getting pleasure (*E*), if [relevant agents] eat chocolate (*p*).

Notice that the agents in *E* and *p* can come apart. This is clear in:

(11) It is good for Finlay, if Schroeder writes about his work.

which itself might be elliptical for: 'It is good for Finlay's book sales, if Schroeder writes about his work'. From this example, we can also see that what takes the position of relevant agents in (10*) needn't be agents of any kind. It might be good for the economy('s picking up) if interest rates are lowered. So the account is very flexible.

Finlay identifies the following grammatical forms in which the word 'good' can occur (ibid., p. 5):

(a) good *to φ*

(b) good *for φ -ing*

(c) good *at φ -ing*

(d) good *for s*

(e) good *with t*

(f) good *as a F*

It also seems that 'good' can (grammatically) take different kinds of objects, as evidenced in:

(g) *t* is good

(h) good that *p*

(i) a good *F*

Finlay believes that ERT is flexible enough to accommodate all these forms and hypothesizes that the prepositional phrases in (a) to (f) make explicit various elements in the propositions *E* and/or *p*. We've already seen that the analysis makes good sense of (a) and (g). But what about (b)?

Take the sentence: 'This stone is good for hammering'. It can be analysed as: 'It is good for driving nails into the wall (*E*), if this stone is used (*p*)'.

(c) is more complicated. Take: 'Ms. Harris is good at teaching.' According to Finlay, this amounts to saying that Ms. Harris is a good teacher (so we are also covering (i)):

'[...] if Ms. Harris is a good teacher, then she is good *at* educating students, which can be explicated as, 'It is good for the education of students (*E*), that Ms. Harris teaches them [(*p*)].' (Ibid., p. 21)

This paraphrase seems ok for teachers, but what about dancers? How to paraphrase 'Sue is good at dancing' (or 'Sue is a good dancer')? Here's a stab: 'It is good for having dances executed correctly (*E*), if Sue performs them (*p*)'. This seems a bit forced. Equally awkward is the analysis of 'Sue is good at chess' as 'It is good for (Sue's) winning chess games, if Sue plays'.[63]

In personal communication, Finlay responded that the awkwardness of these analyses needn't show that they fail to specify the right truth conditions. The paraphrases might seem awkward because we choose not to express ourselves this way. This may have an explanation (convenience, for example).

[63] I think the analysis of 'This is a good painting', where we mean *artistically* good, is also awkward. Perhaps this use of 'good' falls under (f), good *as a painting*. But it's not clear how to explicate this use with ERT. An artistically good painting *can* be good for some end, but what end might it be? Some artistically valuable art might be good *for aesthetic pleasure*, but is it always? And even if it is, should we paraphrase 'This is a good painting' as 'This painting is good for aesthetic pleasure, *if one looks at it*'? This too seems forced to me. It seems more plausible that '*X* is a good painting' means something like '*X* meets the contextually salient standards for paintings'. Unfortunately, I failed to analyse this plausibly in the previous chapter.

We should add that there appears to be something which different ways of being good have in common. It is presumably no accident that the same word 'good' can be used in so many different ways.[64] So we should not settle for diversity in logical form too quickly.

Lastly, even if these analyses did stretch the meaning of 'good' in the examples, there will likely be exceptions to *any* general theory of 'good', as it is used in such a great variety of ways.[65] Finlay may still be right that at least in many uses, 'good' denotes a two-place relation between two propositions *E* and *p*.

So let us look at (d): good *for s*. This form is neatly handled by ERT. 'Running is good *for your health*' can be paraphrased as: 'It is good for your health (*E*), if you run (*p*)'.

What about (e), 'good *with t*'?

> 'For Jerry to be good *with children*, for example, is explicable as its being good for some *E* (e.g. for the children's being happy) that Jerry interacts with children; for Mary to be good *with guns* may be explicable as its being good for

[64] As noted by Von Wright, the same range of application for equivalents of 'good' is found in many different languages (like French, Russian and Greek) ((1963), p. 14).

[65] It is also not clear whether Finlay's ERT makes sense in the case of '*X* feels good'. Does *X* feel good for *E* if *p*? What might *E* and *p* be in this case? I suppose '*X* is good for pleasant sensations, if one touches it' might explicate it. However, it also seems plausible that 'good' is a synonym for 'pleasant' in the sentence '*X* feels good'.

targets' being hit that Mary shoots at them with guns.' (Finlay ms, chapter 3, p. 21).

Lastly, (f). 'This stone is good as a hammer' can be paraphrased as follows: 'It is good for driving nails into the wall (*E*), if this stone is used (*p*)' (this is the same as our explication of (b)).

I haven't explicitly talked about (h), where goodness seems to be a property of states of affairs. But this is clearly covered by ERT: '*p*' in 'good that *p*' is '*p*' in 'good for *E*, if *p*'. In fact, if ERT is correct, then goodness is *always* a property of states of affairs: for goodness is attributed to the potential state of affairs described by *p*.

So Finlay's analysis of 'good' as a relation between two propositions *E* and *p* is flexible enough to accommodate most uses. But it merely explicates the *structure* of 'good'. It does not say what *kind* of relation 'good' establishes between *E* and *p*:

> 'What is the common relation that eating chocolate bears to having sensory pleasure, that the home team's intercepting a pass bears to their winning the game, that the economic mood late in 2008 bears to Obama's winning the presidential election, and that leaves being coloured green bears to pictures being coloured realistically, etc.? This question doesn't seem at all difficult or perplexing, and the philosophers who consider it all give rather similar answers: that *p* "promotes", "serves", "answers to", "satisfies", or "conduces to" *E*. We would question a person's semantic competence with 'good' if they were

unable to recognize some answer of this kind as capturing the meaning with which they use the word.' (Ibid., p. 22)

Finlay elects *raising the probabiliy of* as the semantic value of 'good'. Being good for something, then, is to raise that something's probability. If *p* is good for *E*, then *E* is more likely given *p* than given not-*p*.[66]

The notion of probability is ambiguous between objective and subjective probability. The first is 'probability given the state of the world' and the second 'probability given some set of information of *evidence* about the state of the world' (ibid., p. 23). Although some uses of 'good' concern objective probability (what actually raises the probability of *E*), others involve subjective probability (what raises the probability of *E* relative to our information). Finlay gives the following example:

'Carl goes into hospital with suspected cancer. The doctor runs an initial diagnostic test, which gives no false positives and hence can conclusively confirm the presence of cancer, but gives false negatives in 70% of cases where the cancer is present.

If the doctor reports that the test is negative, it seems correct for Carl to say, 'That's good.' While the test results may slightly raise Carl's subjective

[66] Of course there are many ways in which not-*p* could be the case. You could not go to the cinema by staying at home, but also by going out for dinner. '*E*' may not be equally likely given either description of 'not-*p*'. Complex (but plausible) contextual mechanisms determine what background conditions *B* settle *E*'s conditional probability.

probability that he is cancer-free, they do not of course raise its objective probability.' (Ibid., p. 24)

Context will determine whether subjective or objective probability is involved in statements about goodness.

Finlay's hypothesis leads to the following analyses (where 'to promote *E*' takes the place of 'to raise the probability of *E*'):

'*Good:* '*p* is good for *E*' means that *p* promotes *E* to a positive degree;

Bad: '*p* is bad for *E*' means that *p* promotes *E* to a negative degree (i.e. *demotes* *E* to a positive degree);

Better: '*p* is better for *E* than *q*' means that *p* promotes *E* to a greater degree than *q* promotes *E*;

Worse: '*p* is worse for *E* than *q*' means that *p* promotes *E* to a lesser degree than *q* promotes *E*;

Best: '*p* is best for *E*' means that for all *q* in some comparison class *R*, *p* promotes *E* to a greater degree than *q* does;

Worst: '*p* is worst for *E*', means that for all *q* in some comparison class *R*, *p* promotes *E* to a lesser degree than *q* does.' (Ibid., p. 23)

A promising feature of these analyses is that they explain why the best act (relative to *E*) out of a set *R* is the one you ought (relative to *E*) to do out of *R* (and

vice versa). Remember that '*A* ought to *X* (relative to *E*)' was analysed as: 'Given *B* including *E*, it is most likely that *A* first *X*-ed'. Symmetry of choice guarantees that the act that raises the probability of *E* to the highest value (compared to alternatives in *R*), is also the act most likely done on the assumption that the end obtains. And so it follows from '*X*-ing is the best option (relative to *E*)' that you ought to *X* (relative to *E*) and *vice versa*.

6.5 The end-relational theory of reasons

We need to discuss one more important element of the end-relational theory of normative language. It is the notion of a reason. Finlay starts with the datum that the word 'reason' often occurs in explanatory contexts. For example:

(12) The reason the light isn't turning green is that the car didn't cross the sensor.

(13) The reason this escape plan won't work is that the prison fence is electrified. (Examples are from Finlay ms, chapter 5, pp. 2-3.)

Here, reasons appear to be facts which explain why something is the case. Normative reasons for action (reasons *to X*) initially seem not to fit this picture because, at least grammatically, 'an explanation why to *X*' makes no sense. But Finlay conjectures that this is shorthand for 'an explanation why it would be *good* (in some

way) to *X*' (ibid., p. 6).[67] This, in turn, is analysed in terms of *X* raising the probability of some contextually salient end (as we have seen).

So when we say that there is a reason for you to *X*, we indicate that there is a fact which explains why *X*-ing is in some way good. This existential claim should not be read *de dicto*, but *de re*. In other words: saying that there is a reason to *X* is not saying that there is a fact which explains why *X*-ing is conducive to *some end or other* (there is always *some* end or other). Rather, it is saying that there is a fact which explains why *X*-ing is conducive to some particular end(s). What they are is highly context-dependent (ibid., p. 7).

This account is flexible enough to make good sense of various phenomena. First, we seem to make a distinction between reasons that an agent him or herself has, and reasons that are there to do an act whether or not the agent has them. The psychopath may not have reason to stop torturing puppies, but there may still be moral reason for him or her to stop.

This phenomenon can be explained by the hypothesis that 'reason to *X*' means 'explanation of why it would be good (in some way) to *X*'. For we have seen that it is plausible that 'good to *X*' can be explicated by 'good for *E*, if [relevant agents] *X*'. The identity of *E* is highly context-dependent. When we say 'The psychopath has no reason to stop torturing puppies', we mean that there is no explanation of why stopping promotes ends *that matter to the psychopath* (like inflicting pain). When we

[67] Or at least this is his analysis of a *pro tanto* reason. *Conclusive* reasons can be analysed as facts which explain why one *ought* to *X* (ibid., p. 7).

say 'There is moral reason to stop torturing puppies' we mean that there is an explanation of why stopping promotes *certain moral ends* (like avoiding suffering).

Second, we can make sense of the distinction between normative and motivating reasons. This is not the same as the former distinction, because an agent may *have* normative reasons without acting on them.

Whenever an agent has true beliefs about the facts which explain why acts are conducive to ends, the agent's motivating reasons are simply those facts (which explain why it would be good to *X*) that the agent accepts as reasons and which motivate him or her. Normative reasons (whether the agent *has* them or not) are simply those facts which explain why it would be good (in some way) to *X*.

But there is a complication. In a famous example of Bernard Williams's, a man drinks from a glass that contains petrol, because he mistakenly believes that it contains gin (Williams (1981a)). Here, there is no fact which explains why it would be good to drink from the glass. But it seems he had a motivating reason nonetheless.

In this case, the reason cannot be reported as a fact about what the glass contained. For example, it would be misleading to say: 'The reason the man drank from the glass was that it contained gin'. This suggests, falsely, that the glass did contain gin. Instead, we have to say: 'The reason why the man drank from the glass was that he *believed* that it contained gin'. This may seem to suggest that motivating reasons are attitudes, like beliefs and desires (or belief-desire pairs, as in Smith (1994)). But Finlay provides good reason to reject this:

> 'Normative reasons are things that agents are supposed to take into consideration in their deliberations about what to do. Surely then, when all is going well agents are motivated as a result of their awareness of their normative reasons. And it is most plausible that something is the reason "for which" *s* φ-ed that it is the consideration that *s* accepted as a reason to φ, leading her to φ. By contrast, the rival analysis of motivating reasons – as facts about an agent's psychological attitudes (or just as those attitudes themselves) that causally explain her actions – yields a peculiar mismatch between normative and motivating reasons, because facts about her psychological states are not typically among the normative reasons that an agent attends and responds to in her deliberations.' (Finlay ms, chapter 5, pp. 16-17).

So what happens in the gin-case? There are no facts which explain why it would be good to drink from the glass (at least not relative to the salient ends). But there is a sense in which the agent *did* act for a reason. Finlay proposes that in this case, the motivating reason was 'the supposed but nonexistent *fact that* [*the glass contains gin*]' (ibid., p. 19). We can then interpret the claim that there is a reason for which the man drank from the glass as the claim that there is an explanation *he* accepted for why drinking would be good (i.e. the supposed fact that the glass contains gin). This is compatible with rejecting this explanation ourselves.

Notice that this theory explains the common intuition that if something is good, then there is some reason to seek/do it. For it seems that if something raises the probability of *E* (i.e. is good for *E*), then there are facts which explain this (i.e.

reasons). Since this chapter is already quite long, however, I will postpone a more thorough evaluation until later.

6.6 Conclusion

Finlay's end-relational theory unifies both the semantics of different normative terms and the semantics of normative terms with their nonnormative counterparts. In these respects, it borrows from Aaron Sloman (1970) and Angelika Kratzer (1977), (1981)). It is a powerful theory, according to which 'ought', 'good' and 'reason' are implicitly indexed to contextually salient ends. In the next chapter, I will discuss some problems for this theory. But it can handle all of them.

Appendix: Finlay's model for probability

In this appendix, I will sketch Finlay's account of probability. But the success of his theory does not *depend* on the success of his analysis of probability: 'the end-relational theory should be compatible with a variety of alternative accounts' (ms, chapter 3, p. 28). On one condition, of course: that probability is not itself analysed in normative terms. This would be the case if, for example, by 'the probability that *p*' we meant (something like) 'the degree of credence one ought to have that *p*'. I agree with Finlay, though, that this analysis seems to put the cart before the horse. Intuitively, one ought to have a certain credence in *p*, *because* it has a certain probability (see also Heathwood (2009) for this point).

Finlay believes that the probability of p is determined by the proportion of world-states compatible with certain facts or information in which p. This idea is illustrated by means of the following example:

> 'Stan and Sally are a married couple trying to have a child. They each have one recessive red-haired gene R and one dominant brown-haired gene B. There are thus four possible outcomes: RR, RB, BR, BB. Of these possible outcomes, one out of four is an outcome in which Stan and Sally's child has read hair (RR). Thus, the probability that their child has red hair is .25.' (Finlay, ms, chapter 3, p. 29).

In the example, the facts about genetics are consistent with four possibilities. They form the *possibility space*. But a possibility space needn't be determined by the facts. It can also be determined by the propositions accepted by a subject. Quite generally:

> 'A *possibility space* is a set P of world-types (wt_1, wt_2, wt_3, ...). A world-type is a set of consistent propositions (p_1, p_2, p_3, ...) that partially specify a world-state. For any world-type wt_n in a particular possibility space P, every other world-type wt_i in P is a counterpart of wt_n, such that for every proposition p_m in wt_n, wt_i includes either p_m or its negation not-p_m. A possibility space P consists of world-types whose member propositions can be divided into two groups: those common to every world-type in P (as determined by a closeness relation), and

those that differ from counterpart to counterpart in *P* (as determined by the dimension of comparison). Hence *P* is defined by (i) a set of common propositions, and (ii) a set of disjunctions of contested propositions and their negation;

> *P*=(the set of world-types consisting of (p_1, p_2, ... p_n) and ((q_1 or ~q_1), (q_2 or ~q_2), ... (q_m or ~q_m))).' (Ibid., p. 29)

In the case of subjective probability, the set of world-types is determined by the propositions the subject accepts as true (the common propositions), and the propositions s/he is unsure about (the disjuncts in the set of contested propositions). I take it that the latter are only a subset of all the propositions the subject is unsure about, since there may be infinitely many.[68] The subset would be determined by what is somehow salient for the subject.

In the case of objective probability, the set of world-types is determined by the facts. The probability that *p* at a time t_2 is then determined by the number of world-states compatible with the facts in which *p* at time t_1. So if determinism is true, then the probability that *p* at any time is either 1 or 0.

[68] Finlay also excludes propositions with which the subject is 'unacquainted' (ibid., p. 29).

Chapter 7. Objections to the end-relational theory 1

7.1 Introduction

In this chapter, I discuss some objections to Finlay's end-relational theory, most of which have been discussed in print. I believe he can answers them satisfactorily. In the next chapter, I will discuss more serious difficulties.

7.2 Reducing normativity

The first objection concerns the feasibility of a reductive analysis of normative language. Finlay claims that normative modals can be reductively analysed in terms of what is necessary, possible, or most likely that an agent first did on the assumption that some end has been achieved. Part of his argument proceeds via a reduction of the instrumental 'must' and 'may' in (2010a). He said that instrumental conditionals involving these concepts ('In order that *E*, you must/may *p*') quantify over possibilities compatible with certain ends. They would have a normative flavour whenever two conditions are satisfied (see 6.2 of the previous chapter). The first is the *temporal ordering condition*. This condition is that the event in the consequent is presented as either temporally or metaphysically prior to the end's attainment. The second condition is the *prior necessity condition*. According to it, the necessity/possibility of the event in the consequent must be ascribed to a time prior to the event-time in the antecedent.

Finlay discusses two potential counterexamples to his thesis ((2010a), section 5):

(1) If your opponent is going to win the game, you have to make a lot of mistakes.

(2) If it is going to rain tomorrow, the storm has to veer to the north.

In both (1) and (2) the necessity of the event in the consequent is presented as prior to the event in the antecedent. So they satisfy the temporal ordering condition. Furthermore, the use of 'going to' pushes the event-time of the antecedent to a time after the sentence's aspect-time (i.e. it is *now*, at the aspect-time t_1, necessary to make mistakes / veer to the north *after* which, at the event-time t_2, your opponent wins / it rains tomorrow). So (1) and (2) also satisfy the prior necessity condition. But are they *normative* instrumental conditionals?

Finlay argues that there is no good reason to deny this. His strategy is to consider what is reasonably thought to be necessary for normativity. Two features are sometimes claimed to be essential. The first is illocutionary force. For example, Richard Hare (and noncognitivists more generally) believes that using a normative concept is to perform a kind of act, like recommending. But it seems that in (1) and (2) nothing is recommended (at least not in all contexts where they are used). The second is harder to express clearly. This idea is that normative concepts denote features of the world that are *action-guiding*, *attitude-guiding*, or in some other way

practically relevant (ibid., p. 73). And it seems that (1) and (2) are not (ordinarily) practically relevant.

Finlay points out that the criterion of illocutionary force is highly controversial and appears to have counterexamples (even in assertoric contexts):

> 'Many people find nothing unintelligible about someone saying, 'I know what I ought to do, but I have no interest in doing it', or '*A* is what you *ought* to do, of course. But come and do *B* instead - it's much more fun.' The 'ought' in these sentences still has a clearly normative flavour, even without the illocutionary force that is characteristic of categorical normative claims. The fact that asserting [(1)] and [(2)] typically doesn't involve the illocutionary force of recommendation or prescription is therefore no barrier to their having a minimally normative flavour.' (Ibid., p. 73)

Of course there are contexts in which (1) *can* be used to recommend. For example, if we suppose that you have the aim of letting your oppenent win, then (1) can recommend a course of action (in this context, (1) is also practically relevant). But it doesn't seem *necessary* that something is recommended in order for a sentence to be normative. Even (2) can be used to recommend, if only to a hypothetical agent (like a god).

Finlay further argues that practical relevance or action-guidingness is not a necessary condition for normativity either. He gives the example of

(3) Everybody ought to be loved,

and comments that

> 'Many people find it intelligible that this could be true even if it is not true of every person that there is somebody who ought to love them (or cause them to be loved). If this is right, then there would seem to be normative propositions that aren't guiding for any agent. The fact that [(1)] and [(2)] typically do not express action-guiding propositions is therefore also no barrier to their having a minimally normative flavour.' (Ibid., pp. 73-74)

As before, (2) *can* be practically relevant, if only for a hypothetical agent. So it certainly isn't obvious that Finlay failed to capture the semantic content of instrumentally normative conditionals. Once we allow that normative modals can be reduced in instrumental uses, considerations of simplicity favour this possibility in other cases too (cases where the relation to the end is not syntactically marked). After all, it is more conservative to suppose that 'ought' in 'In order to evade arrest, Max ought to mingle with the crowd' and 'Max ought not to deceive his friends' does not have different meaning.

Finlay's theory will seem even stronger if it can explain why and predict when the use of normative vocabulary tends to carry other types of informational significance (like recommendation or endorsement). This is precisely what he

attempts to do in chapter 6 of his manuscript (an earlier attempt at such an explanation can be found in his (2004); the argument there is substantially the same).

It is sometimes objected to reductive analyses that the use of normative terms is closely linked with the performance of speech acts like recommending or endorsing. But Finlay notes that a reductive analysis is not incompatible with this phenomenon *per se*. Even Simon Blackburn says that in certain contexts, the descriptive predicate 'has south-facing windows' can be used to endorse or recommend, for example in a context where someone is looking for a house with plenty of sunlight (ms, chapter 6, p. 11; Blackburn (1992)).

Furthermore, the idea that (nonassertoric) force is part of the semantics of normative words immediately runs up against the difficulty that such words can be embedded in sentences in which nothing is endorsed or recommended (disjunctions, conditionals, attitude reports). Finlay, on the other hand, can explain why normative sentences (tend to) have such force only in assertoric contexts:

> '[I]f the practicality of a normative sentence derives from the perceived practical significance of the potential *fact* the sentence signifies, then we would expect it to be present in *assertions* of that sentence, being precisely the uses that pragmatically communicate that the speaker believes the world to be that way.' (Finlay ms, chapter 6, pp. 14-15)

In other words: normative sentences are about what is conducive to contextually salient ends. In some cases, the ends are determined by the desires of

participants in the conversation. Given our interest in these ends (and our implicit knowledge of this fact), it makes sense that *asserting* that *X* is conducive to them would have additional informational significance (e.g. would communicate endorsement). If we don't assert it (but merely entertain the possibility, or report on other people's beliefs about the facts), it will typically disappear.

This pragmatic explanation derives support from the fact that the link between endorsement (recommendation, etc.) and normative words is not as robust as might appear from a narrow focus on moral uses (even in assertoric contexts). Consider the sentence 'This tree has good roots'. Is one endorsing the roots, or merely saying that they are conducive to the tree's survival? Similarly: 'I know what I ought to do, but I have no interest in doing so' does not involve endorsement.

Finlay's pragmatic explanation of the occurrence of certain kinds of illocutionary force predicts the absence of it in these two examples. The explanation makes that force depend on (implicit knowledge of) the speaker's commitment to the end. Imagine that John wants the tree to blow down, and says 'It's got good roots, however'. John doesn't communicate endorsement of the roots (if that is possible at all). And that is what we should expect: in this context, we know that John does not desire the tree's survival. Similarly for 'I know what I ought to do, but I have no interest in doing so'. Because we know that the end to which the 'ought' is indexed is no commitment of the speaker's, s/he does not communicate endorsement.

So Finlay can explain why normative sentences have certain types of force only in certain situations. Its defeasibility weakens our reason to suppose that the *semantics* of normative words involves speech acts like endorsement or recommendation.

This does not mean that Finlay believes the "practicality" of normative judgment poses no challenges to reductive analyses. He identifies two principles that require explanation. The first is:

> '*Indispensability:* Every actual agent, no matter what their desires and preferences may be, employs concepts that are expressed by normative words such as 'good', 'ought' and 'reason' in their practical thought and speech.' (Finlay ms, chapter 6, p. 16)

And the second is:

> '*Evidentiality:* An agent's sincerely asserting or judging that something is "good", or that some action "ought" to be performed, or that he has a "reason" to perform it, is by itself and independent of any information about his desires and preferences sufficient though defeasible evidence of corresponding motivation.' (Ibid., p. 17)

Finlay believes these principles are strong enough to restrict a plausible semantics of normative terms. For example, *Indispensability* makes trouble for primitivists who believe 'there are no facts or properties *reducible to the nonnormative* such that seeking and responding to them is constitutive of deliberating' (ibid., p. 17). *Indispensability* would commit the primitivist to either one of two implausible claims: (1) 'that all agents are actually or necessarily disposed to be motivated by some of

these primitive facts and properties' or (2) 'that anybody who reasons towards intentions using only other kinds of facts and properties just isn't deliberating' (ibid., p. 17). The second would be implausible because '[o]n a commonsense and conservative view it is sufficient for deliberation that one has certain desires and preferences, and reasons with aim of satisfying them' (ibid., p. 17).

He also believes that his theory allows him to explain why *Indispensability* and *Evidentiality* are true. With respect to the first Finlay observes that although 'no particular end-relational facts or properties have a guiding role for any agent no matter what their desires might be, nonetheless for any agent there are *some* end-relational facts or properties that have a guiding role – i.e. those facts and properties involving a relation to the agent's own contingently desired or intended ends' (ibid., p. 18). If the conservative hypothesis about deliberation is true, then it is no surprise that all agents use normative concepts in their practical thought and speech: after all, Finlay's theory *reduces* normative concepts to concepts of 'probabilistic relations to ends' (ibid., p. 18). It is also no surprise that normative properties and relations would be "guiding" in practical deliberation (i.e. that agents are disposed to be moved by such properties and relations). After all, they are relations to ends *desired by the agent*.

With respect to *Evidentiality*, Finlay starts with the observation that it 'only holds in general for *unrelativized* normative sentences' (ibid., p. 18). For example, if someone says 'Adultery is good for destroying marriages', we don't infer that s/he expressed approval of adultery or that that s/he is motivated to commit adultery. But we would be inclined to infer this if s/he had said 'Adultery is good'. And the same

holds for 'In order to lose this game, you will have to play very poorly' (no endorsement) and 'You will have to play very poorly' (endorsement).

Of course, Finlay believes that all normative language is implicitly relative. In other words, the sentence 'You will have to play very poorly' abbreviates a more semantically complete utterance. So the challenge for Finlay is to explain why *Evidentiality* holds mainly for superficially unrelativized uses of normative language.

The crux of the explanation is as follows: ordinarily, speakers will not be more explicit than required for accurate understanding.[69] It is not necessary to relativize a normative statement just in case the ends to which it is relative are sufficiently salient. And Grice and Finlay think it is a default assumption that the conversational ends are shared (implicit in Grice's Cooperative Principle).[70] So when a speaker does not explicitly relativize his or her statement to an end (and there are no other clues as to the end's identity), we can reasonably infer that the speaker's statements are relativized to his or her *own* ends, which s/he assumes are sufficiently salient. This explains why we can typically infer from an unrelativized normative statement that the speaker is

[69] Grice captures this in his Maxim of Quantity, which tells us to make our contribution as informative as required but no more informative than required for the purposes of the exchange. See Grice (1975).

[70] Finlay also believes it is defeasible, since not all conversations involve shared ends. But there is at least a presumption of shared ends in many situations. Finlay claims that something like Grice's Cooperative Principle can be derived from the plausible '*Instrumental Law of Pragmatics* (ILP): Speakers always perform the speech act that they believe best for their conversational ends.' (Finlay ms, chapter 6, p. 6). From this we can derive the 'Cooperation Law of Pragmatics (CLP): When people know their conversational ends to be shared, they always perform the speech act that they believe best for the shared conversational ends' (ibid., p. 8).

motivated to promote the ends to which the utterance is implicitly relative. It also explains why unrelativized normative statements tend to have additional informational significance (tend to communicate endorsement). Since we cannot infer from an explicitly relativized statement that the speaker endorses the relevant end (at least not without additional clues), they tend to lack nonassertoric force and to be poor evidence for corresponding motivation.

These pragmatic explanations add further credibility to Finlay's proposed reduction of normative language. Not only does he explain why and when normative language expresses recommendation or endorsement, he also explains why and when it does not.

7.3 Fundamental values

The objection most frequently voiced against Finlay's analysis of 'ought' (at least in my experience) is that it makes 'ought' claims concerning "fundamental values" tautologous (although the only place where this objection is discussed in print is Finlay's own (2009)).

As noted in the previous chapter, not all acts seem to be required in order that some *further* end obtains. A utilitarian might say:

(4) One ought to maximize happiness.

For a utilitarian, the maximization of happiness is the most fundamental or ultimate end in the light of which all action is to be (morally) assessed. But if so, then there is nothing other than the maximization of happiness to which the 'ought' in (4) could be relative. Finlay's theory then predicts that (4) means:

> (4*) Given that one has maximized happiness, it is most likely that one has (first) maximized happiness.

But (4*) is a completely trivial, necessary truth! Surely it is a controversial claim in normative ethics that what it is (morally) right to do is to maximize happiness. Furthermore, it is a claim that (some) philosophers have accepted only after thinking long and hard about it. But why would that be necessary, if it is tautologous?

There are several aspects to this objection. One is that many people dispute (4). Why would that be, if its meaning is (4*)? However, we've already seen the answer to this question in chapter 2. The *point* of moral discourse is not merely theoretical: it is also to protect and promote certain values. These will typically be the values of the *speaker*. So if a hearer does not share the end of maximizing happiness, she will deny (4), not because it is false, but because s/he evaluates moral claims relative to his or her *own* ends.

This thesis about the point of moral discourse is also relevant to another aspect of the objection: why would anyone *say* something trivial like (4*)? Finlay answers this as follows:

> '[I]f the real conversational function of uttering [4] is to demand motivation towards the relevant ends (or at least the prescribed behaviour) rather than to convey [its] semantic content, then [its] significance is quite compatible with [its] being tautologous. We often find that communicative purposes can be served by asserting tautologies: consider 'A fact is a fact,' and 'It ain't over till it's over'.' (Finlay (2009), p. 334)

So Finlay appeals to pragmatic considerations in certain problematic cases. Isn't this a kind of cheating? If the point of moral discourse plays such a crucial role in people's linguistic behaviour, shouldn't it be be incorporated into the semantics of normative words?

Well, not if doing so means incorporating an illocutionary element into the semantics of 'ought'. We've already seen reason to doubt this in 7.2. First, 'ought' occurs in many sentences where nothing appears to be demanded ('I know what I ought to do, but I have no interest in doing so'; 'Either stealing is wrong, or John is not to blame'). Second, 'ought' does not occur in *moral* contexts only ('Maffiosi ought to stick to the code'). Third, 'ought' exhibits the same ambiguity between normative and predictive readings in many languages other than English. So there is likely some common core of meaning. Finlay's semantics is much simpler and more unified than a semantics that gives the moral 'ought' an interpretation unrelated to both other normative uses and its predictive use. So there is good reason to take seriously the invocation of pragmatics to account for some fringe phenomena. For sentences like (4) are, arguably, peripheral. Most everyday uses of 'ought' are indexed to ends

nonidentical to the means ('You ought to buy some shoes', 'You ought to see the doctor', 'You ought to see the exhibition' ...).

But there is a third aspect to the objection. It concerns the nature of moral inquiry. Why do people think long and hard about what we ought morally to do, if the answer is a simple necessary truth?[71]

I would like to approach this problem by considering how to think about the content of the question: 'What ought one morally to do?'. Finlay claims that moral *assertions* are made relative to contextually salient ends. But what about moral questions? Is the question: 'What ought one morally to do?' (always) shorthand for, 'What ought one to do in order that *E* (where '*E* is some particular substantive moral end)? It seems not. In some 'What ought one to do?' questions, it is left open which ends are relevant.

I think Finlay's theory can accommodate this as follows: when the philosopher asks what one ought morally to do, s/he asks: 'What ought one to do in order that the most fundamental moral ends are promoted?', where 'the most fundamental moral ends' is read *de dicto*, not *de re*. This makes sense so long as we understand 'end', 'moral' and 'most fundamental'. This interpretation of the question makes it clear why a philosopher would have to think about the answer. In this interpretation, what these moral ends are is not specified. The philosopher is trying to determine their identity.

Even if the identity of the most fundamental moral ends ultimately depends on the philosopher's own subjective mental states, these needn't be transparent.

[71] This objection was suggested to me in conversation by Krister Bykvist.

Philosophers are typically trying to elicit (their own and other people's) responses by means of thought experiments and other philosophical techniques. This may bring to the surface what we care about most, or most fundamentally.

7.4 Ought all-things-considered

According to the end-relational theory of 'ought', all 'ought' statements are implicitly relative to an end. But isn't there such a thing as ought *simpliciter*, a non-relative 'ought'? Isn't this 'ought' expressed in all-things-considered judgments?

In (2009), Finlay conjectures that the function of 'all things considered' needn't be inconsistent with end-relativity:

> 'If normative 'oughts' are indeed conditional probabilities, then it is natural to read the contrast between prima facie and all-things-considered 'oughts' as the contrast between probabilities conditional on partially specified circumstances and those conditional on fully specified circumstances. The 'things' to be considered therefore need not be ends.' (Ibid., p. 337)

Of course this answers the question what the function of 'all things considered' is supposed to be, but it doesn't answer the question to what end the 'ought' itself is indexed. With respect to this question, Finlay lists two possibilities:

(1) There is no reason why all-in 'ought's could not be relative to a single end. e.g. 'Max ought, all things considered, to mingle with the crowd, in order to evade arrest.'

(2) The 'ought' could be relative to a certain subset of ends, like the ends that the subject deems most important.

He then considers the problem of weighing different ends. It seems that we can deliberate about which of a number of ends to promote. The judgment we arrive at can be expressed by means of an all-things-considered 'ought' judgment. To what end is this 'ought' supposed to be relative?

The first thing he notes is that there is no problem in principle with weighing ends on the basis of a further end. But in cases where there is no further end, it may be strictly false to say that one *ought* to do it. This would misleadingly suggest that there is an end relative to which one ought to. Saying this might of course still serve a useful communicative purpose: it may help to express a demand for a certain action.

I would add that some all-things considered judgments might be relative to the end of maximizing expected utility (see chapter 8). This would explain why the judgment does not appear to be relative to a single (or even several) *substantive* end(s).

Given all these resources, I conclude that the objection from all-in 'ought's does not pose a great threat to Finlay's theory either.

7.5 Epistemic normativity

According to Finlay, all normativity is, in a sense, instrumental: what we must, may and ought to do is relative to some end to which the act is either conducive or at least not detrimental (as in the case of 'may'). But we can also say things like: '*A* has reason to believe the theory of evolution' and '*A* ought to accept the theory of evolution'. Instrumental views of epistemic normativity are sometimes thought to be implausible.

In (2003), Thomas Kelly argues against the idea that epistemic rationality is a kind of instrumental rationality. The view under attack is that epistemic reasons are determined by epistemic goals. This view is motivated by the following type of consideration: What gives me reason to raise my hand? That I want to ask a question. This makes it instrumentally rational for me to raise my hand (all else being equal). What gives me reason to believe in accordance with the evidence? That I desire true beliefs (since what is in accordance with the evidence is most likely to be true).

Kelly's main point against this view is that 'what a person has reason to believe does not seem to depend on the content of her goals in the way that one would expect if the instrumentalist conception were correct' ((2003), p. 621). We (can) still say that you have reason to believe *p* even if you couldn't care less about having true beliefs. So at least our ordinary talk seems to indicate that epistemic reasons are categorical (in the sense of not contingent on desires).

Kelly discusses a possible response to this line of thought. The response is that we talk about epistemic reasons as if they were categorical because *everyone* is committed to the goal which provides the reasons. That explanation is compatible with a instrumentalist conception of their nature.

Kelly then starts making trouble for this line. He asks what end that might be to which everybody is committed. I want true beliefs about the location of the university, since I want to get to the seminar. I want to be informed correctly if I watch the news. This latter goal is, in Kelly's terms, *wider* than the former, in that *many* truths are such that my coming to believe them would constitute its better achievement. But I am not committed to a goal *so* wide that believing *any* truth would count as serving it. Kelly, for example, couldn't care less about whether Hubert Humphrey was an only child:

> 'However, from the fact that some subjects are matters of complete indifference to me, it does not follow that I will inevitably lack epistemic reasons for holding beliefs about those subjects.' (Ibid., p. 625).

If the evidence favours that Humphrey was an only child, there is reason for me to believe it, whether or not I have a goal such that believing it would constitute its better achievement. Kelly concludes that the instrumental conception of epistemic rationality is false.

Is Finlay's theory vulnerable to this line of attack? Finlay's theory of 'reason' says that a reason for *A* to *X* is fact which explains why *X*-ing would be (in some way) good. This in turn is analysed in terms of *X* raising the probability that a salient end is realized. Kelly's arguments make it plausible that the salient end (in the case of epistemic reasons) is not (necessarily) an end that is desired by the agent. But the end-relational theory does not require that the end is desired by the agent.

What is required is that there is some epistemic end to which an epistemic reason explains that *X*-ing is conducive. This end is plausibly related to the nature of belief, which is an attitude that (in some sense) *aims* at the truth (Finlay ms, chapter 5, pp. 13-14). This aim might well be constitutive in the sense that '*belief is the attitude of accepting a proposition with the aim of thereby accepting a truth*' (Velleman (2000), p. 252). So Finlay can, and indeed does, say that a reason for *S* to believe that *p* is 'an explanation for *S* of why it would increase the probability that *S* thereby believes that *p* iff *p* is true, if *S* believes that *p*' (Finlay ms, chapter 5, p. 14).

The fact that the end relative to which we assess epistemic reasons is a nonoptional end with respect to the question what to believe explains why epistemic reasons are not contingent on people's desires. Whether or not John wants to believe truly is irrelevant to the question whether there is epistemic reason for him to believe that *p*. That question is assessed relative to the aim internal to believing: i.e. to accept a proposition if and only if it is true.

I conclude that Finlay's theory does not make epistemic reasons instrumental in an objectionable way.

7.6 Conclusion

I think the above constitute some of the most important objections to Finlay's semantics of normative language. But it seems to me that they can be answered. In the next chapter, we will discuss problems arising from the fact that normative

judgments are sometimes made in the light of several different ends. These problems are more difficult to solve.

Chapter 8. Objections to the end-relational theory 2

8.1 Introduction

In the previous two chapters, I have described Finlay's end-relational theory of normative language and explained why it can meet a number of objections. In this chapter, I discuss two important problems that are not so far addressed in print.[72] They are related to the fact that we often assess actions and objects in the light of *several* ends. The first concerns the truth conditions of claims about the weight of reasons (and what the balance of reasons favours). The second concerns 'ought' claims and claims about goodness made in the light of several ends. In this chapter, I will assess possible solutions to these problems. We will see that they are not without their costs.

8.2 'Ought' and the balance of reasons

In his published work, Finlay has remained silent about the weight of reasons. But our assessment of his theory of 'ought' may depend on what the theory can say about

[72] Nor was I aware how Finlay intended to deal with these issues until recently. After this thesis and chapter were substantially written, I received an unfinished draft of chapter 7 of his book manuscript in which some of these problems are discussed. I have tried to incorporate some of this work in the chapter. (Finlay does not yet explicitly address truth conditions for statements about the weight of reasons.)

this issue. For example, there seem to be relations of entailment between '*A* ought to *X*' and 'The balance of reasons favours that *A X*-es' (or 'The collective weight of the reasons for *A* to *X* is greater than the collective weight of the reasons for *A* to not-*X*').

Given Finlay's definition of 'ought' in terms of probability ('*A* ought to *X*' means that given *B* (the modal base) including *E* (some contextually salient end), it is more likely that *A* first *X*-ed than any relevant alternative), it seems natural to think of the weight of reasons in terms of probability as well. In the next section, I will develop a probabilistic theory of weight, but argue that it is inadequate.

We have seen (in chapter 6, section 6.5) that Finlay defines a reason to *X* as a fact which explains why it is in some way good to *X*. This in turn was analysed in terms of probability: '*X* is good for some end *E*' means that *X* raises the probability of *E*. So (in an early paper):

> [...] a fact is a reason for [*X*]-ing, relative to a system of ends [*E*], iff it explains why [*X*]-ing is conducive to [*E*]. (Finlay (2006), p. 8)

But what is it to judge that a reason is stronger than another reason? Even if we grant that judgments about what we ought to do and judgments about reasons are end-relational, some reasons for *X*-ing are stronger than others. What makes this the case? This question is not easy to answer, but it is even more difficult to see what the truth conditions are of judgments about the relative strength of reasons that derive from different ends. For instance, in a situation where one cannot do both, one may

have a reason to keep one's promise and a reason to help someone out, where these derive from different ends. What makes the one stronger than the other?

8.3 A probabilistic theory of weight

In this section, I will explore a natural extension of Finlay's theory of reasons to a theory of weight. According to him, a fact is a reason for *X*-ing, relative to *E*, iff it explains why *X*-ing is conducive to *E*. This in turn means that *X* raises the probability of *E*. So:

> **Reason:** a fact is a reason to *X* (relative to *E*) iff that fact explains why $Pr(E|B\&X) > Pr(E|B\&\text{not-}X)$.[73] [74]

[73] A small problem arises: in a game of pool, the fact *that* chalking the cue tip increases your chances of pocketing the ball itself seems a reason to do it. But the fact that chalking increases your chances of pocketing the ball does not *explain* why chalking increases your chances of pocketing the ball. And so it wouldn't be a reason on Finlay's definition. There are two ways to solve this problem: (1) the fact that chalking increases the chances of pocketing the ball can be a reason in virtue of a further end, like that of winning the game (thanks to Stephen Finlay). (2) In cases where there is no further end, Finlay can take a buck-passing view: the fact that *X*-ing increases the chances of *E* inherits its status as a reason to *X* from whatever facts make this the case (it is a kind of short-hand). After all, the fact that *X*-ing increases the chances of *E* is not a reason to *X in addition to* the facts which explain why *X*-ing increases the chances of *E*.

[74] I will often leave the modal base (*B*) implicit in the rest of this chapter.

What about the strength of reasons? On Finlay's theory, a fact acquires its status as a reason to *X* in virtue of the fact that it explains why *X*-ing increases the chances of *E*. If what matters is the eventuation of *E*, we have most reason to do the act on the supposition of which *E* is most likely. Since 'most reason' means 'weightiest reason' (not 'greatest number of reasons'), the collective strength of reasons seems correlated with how likely the act is to realize the end.

But what about the weight of individual reasons? If we can think of reasons as facts which raise the probability of *E*, we can (try to) correlate the strength of individual reasons with the value to which each explains that the probability of *E* is raised. But at least relative to the same background information, it does not seem possible that there could be one explanation of why *X*-ing raises the probability of *E* to value *n* and another of why *X*-ing raises the probability of *E* to a different value. So if we keep the background information fixed, and the strength of reasons is determined by the value to which they explain that the probability of *E* is raised, then there cannot be a difference in strength between *E*-based reasons for the same act *X*. So, relative to the same background information, we can allow differences in strength only between reasons based on different ends (whether for the same or different acts) and reasons for different acts (whether based on the same or different ends). Is this a problem? Should we allow that two *E*-based reasons for the same act *X* can have different weight?

It is quite hard to find a convincing example. In my (2010) I gave one of two reasons to return a wallet (*X*), both supposedly based on the end of feeling good as a result. The first reason was the fact that returning the wallet is in accordance with

your principles (which raises the probability of feeling good). The second reason was the fact that you might get a reward (which also raises the probability of feeling good). My idea was that the first reason was stronger because the chances of getting a reward were low, so that the probability of feeling good by *X*-ing was not raised as much in light of the fact that you might get a reward as it was in light of the fact that you would be acting in accordance with your principles.

But this example can be disputed: one can plausibly maintain that there is no generic end of feeling good to which both facts explain that the act is conducive. Instead, one might say there are two ends: (1) feeling morally content and (2) feeling flush. Both are species of feeling good, but they are really different ends. So I am no longer convinced that there are cases where two *E*-based reasons for the same act *X* have different weight.[75] I will assume, then, that we need to account only for differences in the weight of reasons based on different ends (whether for the same or different acts) and differences in the weight of reasons for different acts (whether based on the same or different ends).

[75] It is not impossible to account for a difference in strength between two *E*-based reasons for the same act on a probabilistic theory of weight. What is required is a way of comparing the conditional probability of *E* given *X* and background information which includes reason 1 but not reason 2 with the conditional probability of *E* given *X* and background information which includes reason 2 but not reason 1. For details, see my (2010). However, in order to make sense of the idea that the weight of reasons is determined by expected utility (see section 8.3), it is easier not to vary the background information given which we assess a fact's status as a reason (thanks to Ralph Wedgwood).

At least relative to a single end *E*, a probabilistic theory of weight can account for differences in the strength of reasons for different acts as follows:

> **More Reason:** a fact is stronger reason to *X* than another fact is a reason to *Y* (relative to *E*) iff (a) both are reasons[76] and (b) the probability of *E* given *X* is greater than the probability of *E* given *Y*.

This principle entails that there is most reason to *X* (relative to *E*) just in case the probability of *E* given *X* is greater than the probability of *E* given any relevant alternative.

More Reason explains why 'There is most reason for *A* to *X* (relative to *E*)' entails that *A* ought to *X* (relative to *E*). Recall that '*A* ought to *X*' means that, given *E*, *X* is more likely than any relevant alternative. Given symmetry of choice, this is indeed the case if the probability of *E* given *X* is greater than the probability of *E* given any relevant alternative. For symmetry of choice guarantees that if the probability of *E* given *X* is greater than the probability of *E* given *Y*, then the probability of *X* given *E* is greater than the probability of *Y* given *E*.

Conversely, *More Reason* also predicts that if *A* ought to *X* (relative to *E*), then *A* has most reason to *X* (relative to *E*): if, given *E*, it is more likely that *X* than any

[76] This condition is required since a reason was defined as a fact which explains why the Pr(*E*|*B*&*X*) > Pr(*E*|*B*&not-*X*). If a fact meets condition (b), it does not follow that the probability of *E* given *B*&*X* is greater than the probability of *E* given *B*&not-*X*. And so it does not follow that it is a reason in the first place.

relevant alternative, then the probability of *E* given *X* is greater than the probability of *E* given any relevant alternative (again, assuming symmetry of choice).

So *More Reason* generates intuitively plausible results for the weights of reasons based on the same end. But it has an obvious limitation: our reasons derive from many different ends, and a theory of weight should allow for comparisons of the strength of reasons based on different ends. It is not hard to see that *More Reason* (or some appropriate extension) is not suitable for this.

Suppose we have two reasons F_1 and F_2 for an action *X*. And suppose we have two ends, E_1 and E_2. F_1 raises the probability of E_1 given *X*, but F_2 raises the probability of E_2 given *X*. If we want to know which reason is weightier, it won't do to compare the values to which F_1 and F_2 respectively explain that the probability of E_1 and E_2 are raised given *X*. For the probability of E_1 given *X* may be much lower than the probability of E_2 given *X*, even though F_1 is a stronger reason than F_2. This is the case if E_1 is *more important* than E_2.

The obvious response seems to be the following: it's true that when two facts derive their status as reasons to *X* from different ends, their strengths are incommensurable without a further common end to which we can relate them. Once this is supplied (call it '*S*' for *superend*), we are in a position to determine their relative strengths by comparing the values to which they respectively explain that the probability of *S* given *X* is raised.

But there is a problem. It is not *a priori* that the speaker is always committed to a further end *S* such that both F_1 and F_2 raise its probability (given *X* - in the interest of smooth formulation I will sometimes omit this condition). (If the speaker

is committed to a further end at all. Consider a conflict between requirements of prudence and requirements of morality. It is not clear on the basis of what further end we are to assess which one of these to follow.)[77] And even if we could always find a further end, it is not clear that the respective values to which F_1 and F_2 raise its probability will always correspond to our intuitive assessments of their weights.

So there are two ways to save our probabilistic theory of weight: we can either deny that reasons with comparative weights ever derive their status as reasons from different ends, or we can prove that whenever two reasons F_1 and F_2 (derived from two ends E_1 and E_2) have comparative weights, then there is always a further end S such that the respective values to which F_1 and F_2 respectively explain why its probability is raised always correspond to our intuitive assessments of their weight. Neither option seems promising.

It will not work to combine E_1 and E_2 into a complex superend (a conjunctive end), for the following reason: E_1 and E_2 may be incompatible, in which case the probability that this impossible end is realized given anything whatsoever is 0. But a

[77] Some people may object to the idea that the speaker needs to be *committed* to a further end. Perhaps it would suffice if such a further end exists. Two points in response: first, I take it that a plausible assignment of truth conditions to normative statements will at least in some cases be constrained by the speaker's commitments. It would be odd to assign utilitarian truth conditions to moral judgments by a Kantian. But it is important to bear in mind that the notion of commitment is not supposed to entail *conscious* commitment. Speakers needn't be able to articulate the ends to which their judgments are indexed. Second, it is not at all clear that a further end can always be found, as I point out in the main text.

reason based on an end that is incompatible with another end needn't have zero weight.

I think the above considerations show that our judgments about the relative weight of facts which derive their status as reasons to *X* from different ends cannot (always) be understood as judgments about the relative values to which they explain that the probability of some further end is raised. This result poses a challenge for Finlay's end-relational theory of 'ought'. For that theory is most naturally paired with a probabilistic theory of weight. But such a theory is inadequate: the weight of an *E*-based reason cannot merely be a matter of the value to which it explains that the probability of *E* is raised. By implication, such a theory cannot explain why '*A* ought to *X*' would entail that the balance of reasons favours that *A X*-es.

What appears to be missing from the account is something which determines the *importance* of an end. And it seems this cannot be understood in terms of probabilities. That, however, makes it unclear how '*A* ought to *X* (relative to *E*)' could entail information about relative importance. Suppose, for example, that importance for a subject is a matter of that subject's preferences. Since the meaning of '*A* ought to *X* (relative to *E*)' does not involve anyone's preferences, there is no logical entailment from '*A* ought to *X*' to 'The balance of reasons favours that *A X*-es'.

In the next section, I will develop a different theory of weight as determined by the expected utility of acts. I will argue that this does not allow logical entailments between 'ought' statements that are not themselves indexed to the end of maximizing expected utility.

8.4 Weight and expected utility

In the previous section, I sketched a dilemma for Finlay's theory. In order to account for the weight of reasons, he could either deny that reasons ever derive from different ends, or he could argue that whenever two reasons (derived from different ends) have comparative weights, there is always a further end which accounts for their comparative strength.

The former seemed implausible, but so did the latter. It did not seem plausible that there would always be some further end to which two reasons can be related in order to account for their comparative strength. But this thought may depend on a particular *conception* of the further end. If we think of ends as particular, *substantive* aims (like acquiring money or avoiding pain), then we have reason for scepticism. But there is another option. This option was suggested to me in personal communication by Finlay himself. The idea is to think of the further end as the maximization of expected utility (I already mentioned it in chapter 5, section 5.6). This end is not substantive in the sense that what it amounts to itself depends (at least in part) on other "lower-level" ends. I will call it a *formal* end. In what follows, I will develop this suggestion.

To make headway, I will explain in slightly more detail than I did in chapter 5

what expected utility theory is (still ignoring various details).[78] Expected utility theory (EUT) can be used as a normative and as a descriptive theory. In its normative guise, it is a theory of rational preference (a theory of which preference ordering it would be rational to have).[79] Its essence consists in an expected utility function which assigns an expected utility value to each act in a set of alternatives. Normative EUT claims that it is rational to prefer one alternative to another just in case the first's expected utility is greater than the second's. In other words: it is rational to maximize expected utility.

What is expected utility? This notion is defined in different ways in different approaches, but we needn't go into the debate between the rivals.[80] The problems we will find arise on most versions of EUT. So I will confine myself to an influential one developed by Richard Jeffrey in (1965).[81]

Most versions of EUT assume that *rational* preferences satisfy a number of axioms (first formulated by Von Neumann and Morgenstern in (1947)).[82] Examples are the following (where '$X \geq Y$' means that one either prefers X to Y or is indifferent between them):

---

[78] The reason I did not go into great detail in chapter 5 was that expected utility theory could not save the standard-relational theory from its problems in accounting for the difference between 'must' and 'ought'.

[79] I will return to the distinction between normative and descriptive uses of EUT in section 8.5.

[80] For a detailed overview, see Joyce (1999).

[81] This version is commonly known as Evidential Decision Theory, as opposed to Causal Decision Theory (see, e.g. Joyce 1999, chapters 4 and 5). My discussion of Jeffrey is based on Joyce's book.

[82] Or at least, most versions assume that there is some way of fleshing out the subject's rational preferences so as to make them satisfy the axioms (Joyce 1999, p. 45).

Transitivity: for every X, Y and Z with $X \geq Y$ and $Y \geq Z$ we must have $X \geq Z$.

Reflexivity: for every X, $X \geq X$.

Completeness: for every X and Y either $X \geq Y$, or $Y \geq X$.

This list is not exhaustive, but that is not important for our purposes. Von Neumann and Morgenstern proved that, provided the relevant axioms are satisfied, there is always some expected utility function which represents a subject's rational preferences.[83] This also holds for Jeffrey's function.[84]

In Jeffrey's version of EUT, the expected utility of an act is determined by two factors: (1) the value of each of its possible outcomes and (2) the probability of each outcome conditional on performance of the act. More specifically, the expected utility of an act is the *sum* of the values of each possible outcome multiplied by that outcome's conditional probability.

In formal notation:

Let X be an act in a set of alternatives.

Let O_1, O_2, (etc.) be possible outcomes of X.

[83] The sense in which such a function *represents* a subject's rational preferences is that it can be proven that X is preferred over Y if and only if X's expected utility is greater than Y's (see Joyce (1999), p. 43).

[84] For a discussion of Von Neumann and Morgenstern's work, see Joyce (1999), chapter 1. For Jeffrey, see chapter 4. One difference between Jeffrey's version of EUT and earlier ones is that Jeffrey defined expected utility in terms of conditional probabilities.

Let $P(O_i|X)$ be the conditional probability of O_i given X.

Let $V(O_i)$ be the value of O_i.

$EU(X)$, the expected utility of X, is then determined by the following equation (called *Jeffrey's Equation* in Joyce (1999), p. 4):

$$EU(X) = \sum_i V(O_i)*P(O_i|X)$$

Both the value of outcomes and their conditional probabilities are determined subjectively. The value of an outcome represents the degree to which the subject desires that outcome.[85] An outcome's conditional probability can (in principle) represent several things, depending on one's theory of probability. For example, it can represent ratios of regions within a possibility space or (rational) degrees of belief. What matters is that the assignment obeys the laws of probability calculus and that the values represent *subjective* probabilities (probabilities relative to the state of information of the subject). After all, normative EUT is a theory of *rational*

[85] Or at least I will assume this in the context of my thesis. John Broome (1993) presents a theory in which utility values represent the relative goodness of various options. I cannot do this, because in the end-relational theory goodness is a matter of conduciveness to ends. But we are invoking EUT to deal with the *importance* of an end, which cannot be reduced to conduciveness to substantive ends.

preference, and the rationality of a mental state is a matter of its relation to other mental states.[86]

Leonard Savage distinguished between acts, outcomes and states of the world (see Joyce (1999), chapter 2), but we can take outcomes to be conjunctions of acts and states of the world. Following Joyce (who in turn follows Jeffrey), I will take acts and outcomes to be propositions. An example of an act proposition is: *I buy a lottery ticket*. The outcome proposition describes a possible state of the world, like: *(I have bought a lottery ticket and) I win a million dollars*. Outcome propositions should be 'detailed enough to supply a definite answer to *every* question the decision maker might care about in the context of her decision' (ibid., p. 52). I might (strongly) desire to win a million dollars only if it does not make my brother miserable. If that matters to me too, then the possible outcomes should include information about my brother's welfare. For example, two relevant outcomes are: *I win a million dollars and my brother is happy* and: *I win a million dollars and my brother is miserable*. If the expected utility function is to represent my rational preferences, then the outcome propositions ought to settle everything that matters to me.

Now for the application of EUT in the context of Finlay's semantics. We're interested in the question what the weight of reasons is determined by and are considering the idea that it is (somehow) determined by expected utility. To see how

[86] Or at least one important notion of rationality, the one that plays a role here, is a matter of internal relations between mental states.

this might go, think of outcome(-proposition)s as end(-proposition)s.[87] This makes sense because we're interested in the weight of end-relational reasons. End-relational reasons are facts which explain why an act is conducive to an end. So it makes sense to make the weight of such reasons depend (at least in part) on the value or importance of the end which makes them into reasons. So the required kind of EUT is one where the expected utility of an act is a function of (1) the value of the ends to which it is conducive and (2) the probability that they are realized conditional on performance of the act.

It should be clear that the notion of value here is not end-relational (not a matter of conduciveness to substantial ends, like avoiding pain or getting richer). That would invite the problem raised in section 8.3. So what is it determined by? For a subjectivist like Finlay, it will be determined by the subject's relation to the end (his or her mental states with respect to it). We can therefore use the idea from EUT that the value of an end for a subject is determined by the degree to which that subject desires the end. So we will be working with a theory according to which the expected utility of an act (for a subject *S*) is determined by (1) the degrees to which *S* desires the ends to which it is conducive and (2) the probability of their realization conditional on performance of the act.

This brings us to the application of this theory to the problem of the weight of reasons. The idea is to make the weight of a reason to *X* depend, in some sense, on

[87] This means that the end-propositions will have to specify everything that matters to the subject, just as outcome-propositions did.

X's expected utility. But, of course, we shouldn't make the weight of an individual reason to *X* depend on the sum of the products of the probabilities and importance of all the (relevant) ends to which *X* is conducive. Rather, we should make the weight of a reason to *X* depend only on the product of the probability and importance of the end *to which the reason explains that* X *is conducive*. The simplest way of doing this is as follows:

> **More Reason (simple):** an E_1-based reason F_1 is stronger reason to *X* than an E_2-based reason F_2 iff $V(E_1)*P(E_1|X)$ is greater than $V(E_2)*P(E_2|X)$.[88]

But this account faces a problem (which I owe to John Broome).

Suppose we have two ends E_1 and E_2. E_1 is that John can afford a ticket to Wimbledon, E_2 is that Sue can afford a ticket to Wimbledon. These ends are equally important (John and Sue are equally friends of ours). The act *X* is that we send £100 to each of them. John is well off, and can probably afford to buy a ticket to Wimbledon even without our money. Nevertheless, *X* slightly increase the probability of E_1, say from 98% to 99%. Sue is not well off, and even if she receives our money, there is only a 60% probability that she can afford the ticket. But without our money, the probability is 0%. So X raised the probability from 0% to

[88] Of course this principle would have to be extended to account for differences in strength between reasons for different acts (whether based on the same or different ends).

60%. *More Reason (simple)* entails that the E_1-based reason to send £100 to each of them is stronger than the E_2-based reason to do this. But that is not intuitively right.

This problem can be avoided by making the weight of an individual reason depend on the degree to which it increases the expected utility achieved from an end. We can determine this degree by subtracting the value of $V(E_1)^*P(E_1|X)$ in the light of background information which *excludes* F_1 from the value of $V(E_1)^*P(E_1|X)$ in the light of background information which *includes* F_1. (This would also in principle allow for differences in strength between reasons for the same act that are based on the same end, although we may not need this, as I have argued in section 8.3)

What we need to account for the weight of reasons is a principle like this:

> **Preliminary More Reason (differential):** an E_1-based reason F_1 is stronger reason to X than an E_2-based reason F_2 iff, for some n and m, F_1 explains why $V(E_1)^*P(E_1|X\&B$ including $F_1)$ minus $V(E_1)^*P(E_1|X\&B$ excluding $F_1)$ is n and F_2 explains why $V(E_2)^*P(E_2|X\&B$ including $F_2)$ minus $V(E_2)^*P(E_2|X\&B$ excluding $F_2)$ is m, and $n > m$.

Our definition does not cover cases where two reasons based on different ends are each reasons for a different act and cases where two reasons based on the same end are reasons for different acts. A definition that covers all these cases is this:

> **More Reason (differential):** an E_1-based reason F_1 is stronger reason to X than an E_2-based reason F_2 is a reason to Y iff, for some n and m, F_1 explains why

$V(E_1)$*$P(E_1|X\&B$ including $F_1)$ minus $V(E_1)$*$P(E_1|X\&B$ excluding $F_1)$ is n and F_2 explains why $V(E_2)$*$P(E_2|Y\&B$ including $F_2)$ minus $V(E_2)$*$P(E_2|Y\&B$ excluding $F_2)$ is m, and $n > m$ (where X and Y are either the same or different acts in R and E_1 and E_2 are either the same or different ends).

This principle says that an E_1-based reason F_1 is stronger reason to X than an E_2-based reason F_2 is a reason to Y iff the degree to which F_1 increases the expected value (= utility) achieved from E_1 is greater than the degree to which F_2 increases the expected value achieved from E_2 (whether X and Y are the same or different acts in R and E_1 and E_2 are the same or different ends).

If the individual weight of a reason is determined by the difference it makes to the expected value of an act with respect to the end to which it owes its status as a reason, then the *collective* weight of reasons should be determined by the difference they collectively make to the expected value of the act. So we could say:

Most Reason (differential): there is a most reason to X iff, for some n and m and all Y in R nonidentical to X, the reasons to X explain why

$\sum_i V(E_i)*P(E_i|X\&B$ including $F_i)$ minus $\sum_i V(E_i)*P(E_i|X\&B$ excluding $F_i)$ is n and the reasons to Y explain why $\sum_j V(E_j)*P(E_j|Y\&B$ including $F_j)$ minus $\sum_j V(E_j)*P(E_j|Y\&B$ excluding $F_j)$ is m, and $n > m$.[89]

Most Reason (differential) says that we have most reason to do that act the reasons for which make the greatest (positive) difference to its expected value (i.e. we have most reason to X just in case the degree to which the expected value of X is raised by all the reasons to X is greater than the degree to which the expected value of any other act is raised by the reasons for these alternatives).

If it were possible for the collective difference made by the reasons for X to X's expected utility to be greater than the collective difference made by the reasons for Y to Y's expected utility *even though* Y has greater expected utility than X, then *Most Reason (differential)* would be dubious. Or at least it would be dubious on the plausible assumption that we have most reason to do the act which maximizes expected utility. So we need the following to be true:

[89] This principle may seem wrong for the following reason: it only takes into account the extent to which an act is conducive to *desired* states of affairs. After all, I've said that the *importance* of an end was determined by the degree to which that end is desired by the subject. But what if X also has highly *un*desirable outcomes? Surely this matters to what there is most reason to do. However, I think we can take these into account by assuming that each undesirable state of affairs corresponds to a desired state of affairs consisting of the *avoidance* of the undesirable state of affairs.

(N) Necessarily: if the reasons for an act *X* collectively make the greatest difference to *X*'s expected utility (compared to the reasons for relevant alternatives in *R*), then *X* maximizes expected utility.

If there could be differences in the conditional probabilities of ends given acts *prior* to factoring in the reasons for the acts, then (N) would be false. For then a scenario would be possible in which there are two acts, *X* and *Y*, and two ends E_1 and E_2, which are equally important. F_1 is an E_1-based reason to *X* and F_2 is an E_2-based reason to *Y* and they are the only relevant reasons. In the scenario, $P(E_1|X\&B$ *excluding* the reasons to *X*) = 0.7 and $P(E_1|X\&B$ *including* the reasons to *X*) = 0.8. So the difference made to the expected utility of X by the reasons to *X* is very small. Further assume that $P(E_2|Y\&B$ *excluding* the reasons to *Y*) = 0.2 and $P(E_2|Y\&B$ *including* the reasons to *Y*) = 0.6. This means that the difference made by the reasons to *Y* to the expected utility of *Y* is greater than the difference made by the reasons to *X* to the expected utility of *X*. But since the expected utility of *X* is greater than the expected utility of *Y*, we intuitively have most reason to *X*.

This scenario would be possible if (N) were false. But there is reason to think it is true.[90] For suppose that there are some facts which explain why the conditional probability of *E* given *X* is greater than the conditional probability of *E* given any alternative act. Then these facts must be reasons to *X*, because a reason was *defined* as a fact which explains why $P(E|X) > P(E|\text{not-}X)$, where 'not-*X*' is a variable for any

[90] Thanks to Ralph Wedgwood.

relevant alternative to *X* (see section 8.3). This is a reason to think that (N) is true and I will assume so in what follows.[91]

We have now arrived at a theory according to which the weight of reasons is determined by the difference they make to the expected utility of an act. But this theory seems to require certain psychological assumptions which might be problematic. The next section is devoted to this issue.

8.5 Controversies over EUT

In the previous section, I said that EUT comes in a normative and in a descriptive guise. In its normative guise, it is a theory of rational preference. In its descriptive guise, it is a theory of people's actual preferences. Which one is relevant to Finlay?

The following line of thought seems to support the idea that it is really descriptive EUT that could be of use to Finlay: although it is plausible to correlate a speaker's actual preferences with his or her judgments about the weight of reasons, it isn't plausible to correlate them with his or her ideal preferences (the preferences s/he would have if only s/he were rational). So the truth conditions of judgments about the weight of reasons should consist in facts about the speaker's actual preferences. Since descriptive EUT is supposed to describe actual preferences, it is descriptive EUT that is relevant to Finlay.

[91] Should (N) turn out to be false, then many of the points made in this chapter can also be made with respect to *More* and *Most Reason (informal)* discussed below. These principles do not require the truth of (N).

If this is correct, then there seems to be a problem. Empirical studies appear to show that people's actual preferences don't always conform to EUT (see for example Yaqub, Saz & Hussain (2009)). People seem to violate various of the axioms formulated by VonNeumann and Morgenstern at least some of the time. For example, it does not seem plausible that people's mental states (desires) are always determinate enough for precise assignments of values to $V(E)$ (which means that the completeness axiom is often violated; see Joyce (1999), pp. 43–46).

This creates the following problem: if people's judgments about the weight of reasons is plausibly correlated with their actual preferences, but people's actual preferences cannot be described by an expected utility function like EU(X), then why would *Most Reason (differential)* give the truth conditions of judgments about the balance of reasons?

One way to respond to this problem is piecemeal: one could question, for each axiom, whether empirical studies really show that it is violated. Or one could try to amend or change the axioms. For example: the problem of indeterminate mental states might be solved by weakening the completeness axiom. Instead of demanding that the subject is either indifferent between *X* and *Y*, or prefers one over the other (for all relevant *X* and *Y*), we could demand that 'it [...] be possible in principle to "flesh out" [an] agent's preferences (usually in more than one way) to obtain a complete ranking that does not violate [the completeness axiom]' (ibid., p. 45). So long as the subject clearly prefers one option over the others, it can be indeterminate exactly how much a subject desires some of the other options.

But that still leaves other violations. For example, it seems that people's preferences differ when the same options are presented in a different way (violating an axiom of consistency, not listed above). People also seem to violate transitivity, although this is perhaps more controversial: some apparent violations may result from insufficiently finegrained descriptions of the outcomes (see Broome (1993), pp. 100-104).[92]

Since I don't want to try and fix all apparent violations, let us move on to the "wholesale" (as opposed to piecemeal) solution. One could also argue as follows: even if people violate axioms, it may still be plausible to think of *Most Reason (differential)* as stating the truth condition of statements about the balance of reasons. This would be the case if, in judging that *X* is favoured by the balance of reasons, people implicitly commit themselves to thinking that their preference for *X* is rational as specified by EUT. This may be plausible. Suppose, for example, that someone becomes conscious of the fact that s/he both prefers *A* over *B* and *B* over *A*. Would s/he not think that something was amiss (in fact, would s/he not cease to *have* these preferences)? Our current theory predicts that this person would become unsure about what is required by the balance of reasons. And that seems correct.[93]

[92] Broome discusses normative EUT, but his reflections are also useful in our context.

[93] It is sometimes noted that values can be incommensurable, and that this is incompatible with expected utility theory (e.g. Broome (1993), pp. 92-93). But if values are incommensurable, people will likely be unsure about what to prefer. So this theory predicts that incommensurable values will make people unsure about what the balance of reasons requires. That seems correct as well.

Still, one may worry that the requirement of a mathematical representation of preferences is too restrictive. For example: acts that are promotive of ends that matter greatly to the agent may still be weak if the act promotes the end only a little. And *vice versa*: reasons for acts that promote certain ends only a little may nonetheless be strong if the end matters greatly to the agent. A function like EU(X) seems to entail that a particular low (but positive) utility value - the utility of an end, not an act - can be compensated for by raising the conditional probability of this end to some particular higher value (and *vice versa*). But is the relationship between utility values and conditional probabilities really linear for actual subjects? In other words: will raising the conditional probability of an end with very low (but positive) utility always lead the agent to prefer it at the very same value? I'm not sure (nor am I sure that a nonlinear function couldn't be developed).

But, if required, we can bypass these worries by ignoring technical ways of thinking about expected utility. What remains plausible (at least for a subjectivist end-relational theory) is that the weight of reasons is determined not just by the extent to which an act is promotive of ends, but also by the extent to which the subject desires the ends in question. This idea explains why the weight of an *E*-based reason to *X* cannot merely be a matter of the value to which it explains that the act raises the probability of *E*. We also need to take into account the utility of *E*. Now add the premise that (the relevant) preferences with respect to various acts are informed by these two factors (even if not in some linear way). If this is plausible, we can give the following simple truth condition for claims about comparative weight:

> **More Reason (informal):** an E_1-based reason F_1 for A to X is weightier than an E_2-based reason F_2 for A to Y (from the perspective of a subject S) iff S would prefer that E_1 is promoted to the extent it is by A's X-ing to E_2's being promoted to the extent it is by A's Y-ing (in a choice between X and Y such that F_1 and F_2 are the only relevant reasons).[94]

And the following would work as a truth condition for 'The balance of reasons requires that A X-es':

> **Most Reason (informal):** there is most reason for A to X (for a subject S) iff S prefers that the ends to which X is conducive are promoted to the extent they are by A's X-ing to other ends being promoted to the extent they are by other acts in R.

In short: there is most reason for A to X (for a subject S) iff S prefers that A X-es to A's doing any alternative in R.

One may wonder whether the empirical problems with descriptive EUT aren't equally problems for *More* and *Most Reason (informal)*. If people sometimes both prefer A over B and B over A (for example, under different descriptions of A and B), then *Most Reason (informal)* predicts that it is both true that the balance of reasons

[94] This principle bypasses mathematical representations of preferences but implicitly factors in both conditional probabilities and utility values.

favours *A* and that it favours not-*A* (assuming that *A* and *B* are incompatible). That seems absurd.

But is this really a problem? It would be only if the subject's preferences with respect to various acts do not correspond to his or her judgments about the best option (what is favoured by the balance of reasons). This seems unlikely. Of course it is possible to prefer *A* in the light of one's self-interest and to prefer *B* in the light of moral ends, but it does not seem possible to prefer *B* in the light of moral ends yet to judge that *B* is not morally the best option. If that's right, then awareness of inconsistent preferences should lead the subject to think s/he has inconsistent views about the best option. But that is exactly what *Most Reason (informal)* predicts.[95] So I am not convinced that observed violations of preference axioms pose a serious threat to *More* and *Most Reason (informal)*, or even *More* and *Most Reason (differential)*.[96] In what follows, I will proceed as if *Most Reason (differential)* specifies the right truth conditions for claims about the balance of reasons, but I hope it will be clear that my points can also be made with *More Reason (informal)* and *Most Reason (informal)*.[97]

---

[95] So on this account, the 'balance of reasons' operator does not obey the so-called *D* principle of deontic logic, according to which '*Op*' entails '$\neg O(\neg p)$' (thanks to Ralph Wedgwood).

[96] Let us also not forget that subjects often arrive at consistent preferences in actual choice situations, because the number of options considered is often very limited. Furthermore, people rarely encounter situations in which two options are equivalent but differently described.

[97] Or at least most of them can.

8.6 Weight, expected utility and the entailments

Suppose that *Most Reason (differential)* gives us the truth condition of 'There is most reason to *X*'. Can Finlay now explain why '*A* ought to *X*' entails that there is most reason for *A* to *X*? This primarily depends on what ends '*A* ought to *X*' is relative to. Suppose that it is a single (substantive) end E_1. From the fact that out of a set of alternatives, *X* is most conducive to E_1, it does not follow that *X* maximizes expected utility. After all, we may not desire E_1 all that much (nor need it be very probable even given *X*). But since we have assumed that the truth condition for 'There is most reason to *X*' is satisfied iff *X* maximizes expected utility, it doesn't follow from '*A* ought to *X* relative to E_1' that there is most reason for *A* to *X*.

Of course, the same holds if '*A* ought to *X*' is relative to two or more (substantive) ends. From the fact that out of a set of alternatives, *X* is most conducive to E_1 and E_2 (etc.), it does not follow that *X* maximizes expected utility. After all, we may not desire E_1 and E_2 (etc.) all that much. And so it does not follow that there is most reason for *A* to *X*.

What happens if we make '*A* ought to *X*' itself relative to the end of maximizing expected utility? From the fact that out of a set of alternatives, *X* is most conducive to maximizing expected utility, it follows that *X* maximizes expected utility (relative to that set). It may not be obvious that this follows, since being conducive to something means raising the probability that it transpires (which is compatible with its not transpiring). But remember that Finlay's direction of conditionalization is reversed. '*A* ought to *X* in order that *E*' means 'Given *E*, it is most likely that *A X*-ed'.

Now substitute for '*E*' the maximization of expected utility. It is obvious that, given that *A* has maximized expected utility, the act which has greatest expected utility is the one most likely performed by *A*. And since 'There is most reason for *A* to *X*' is true iff *X* maximizes expected utility, it follows from '*A* ought to *X*' that the balance of reasons favours that *A X*-es.

However, one might wonder whether Finlay can or wants to make 'ought' statements relative to the end of maximizing expected utility. There are at least two worries: first, it seems to collapse the distinction between 'ought' and 'must' in at least some cases.[98] Remember that '*A* must *X* in order that *E*' means that no worlds where *E* obtains are such that *A* does not *X*. But if there is only one act that maximizes expected utility, then there are no worlds in which that end obtains in which the act is not performed. (Notice that the distinction would still exist for cases where more than one act is optimal in terms of expected utility.)

But I'm not sure that this is a serious problem. The word 'ought' can mean 'most probable' even if what grounds the probability is necessity (truth in all possibilities). So there would still be a difference in meaning.[99]

---

[98] Finlay expressed this concern to me in personal communication.

[99] In ms, chapter 7, Finlay does mention the need to explain *why* speakers would be using the weaker 'ought' if the truth conditions for 'must' are also satisfied. But we have some options here. For example, although there is only one act that maximizes expected value, other less preferred options may still be conversationally salient and acceptable to the speaker. Using the weaker 'ought' could help to communicate this fact (Finlay ms, chapter 7, pp. 18-22).

The second reason why Finlay may want to avoid making 'ought's relative to the maximization of expected utility is that it would introduce information about people's mental states (and the relation in which they stand to certain ends) into the meaning of (some) 'ought' statements. That Finlay wants to avoid this is suggested in a forthcoming article where he writes:

> 'I maintain that our moral claims are relativized to our standards or ends, but this is to be read *de re*, not *de dicto*; i.e. if my relevant moral end is *E* then my moral claim is to be interpreted as ought-relative-to-*E*, not as ought-relative-to-my-ends.' (Finlay, (forthcoming b))

If we think of *E* as the maximization of expected utility, it seems we have introduced information about some subject into the meaning of moral statements, because being such as to maximize expected utility is to stand in a certain relation to some subject's desires. This relation cannot be eliminated from our conception of expected utility, since we cannot make the value or importance of ends depend on conduciveness to substantive ends.

The point can be put more simply: the expected utility function represents a subject's preference ordering (itself based on the degrees to which that subject desires certain ends and their conditional probabilities). So '*A* ought to *X* in order that

expected utility is maximized' can be unpacked as meaning '*A* ought to *X* in order that *A* acts in accordance with [some subject's] preferences'.[100]

So it seems that making 'ought's relative to maximizing expected utility would make unavailable to Finlay his response to an objection by Richard Joyce (forthcoming). Joyce complains about relativistic analyses of moral language as follows:

> '[Suppose] the judge at the Nuremberg trials kept relativizing his condemnation of the war criminals with the suffix "…by our moral standards." This would not just be weird and irritating; it would be scandalous; there would be protests. [...] The relativist Nuremberg judge [...] will be interpreted not as adding something unnecessary, but as revealing himself, in adding the suffix, to be saying too little.' (Joyce (forthcoming))

Finlay retorts that his end-relational analyses do not include positive determiners (like 'my', 'your', 'our', etc.):

[100] It isn't essential that the preference ordering is always the speaker's. As indicated in chapter 5, section 5.6, some judgments of expected utility may be estimates of what has greatest utility for someone else. Whether such a judgment is true would then depend on that person's mental states. However, it is quite plausible that the speaker's preferences are involved in certain interesting cases, like the case of moral language.

> ‘Suppose for example that the [...] end is promoting general human wellbeing. In demanding respect for general human wellbeing and asserting that the Nazis acted in ways detrimental to that end, the judge would not be directing our attention to his/our own attitudes at all, but simply to the ideal of general human wellbeing, and its relation to the actions of the Nazis. [...] On this kind of relativist view, it is no essential part of what we as moral speakers communicate that we demand concern or respect for these ends because they are our ends.’ (Finlay (forthcoming))

But if Finlay made ‘ought’s relative to maximizing expected utility, then this response might be unavailable to him. For this end is intrinsically relational: it is the end of acting in accordance with someone’s preferences, or if not this, then acting so as to maximize the value of the function of degrees of (someone’s!) desires and the conditional likelihood of ends.

One might suggest that expected utility itself is not a matter of preferences. It is whatever values are determined by the function. And so the end of maximizing expected utility can also be thought of as the end of acting in accordance with the greatest (abstract) value. No mental states involved! However, I don’t think this is an option. Expected utility values have to be interpreted (they are the values of something which they represent). What they represent is a relationship between desires and the likelihoods of ends. That would already bring mental states into the truth conditions of statements about the balance of reasons. But it makes sense to interpret this relation further as representing preferences.

However, there might be a way to avoid positive determiners even if we concede that expected utility values represent mental states. Perhaps it needn't be explicit in the truth conditions of the 'ought' judgment *whose* preferences are measured by the function. Perhaps they can be thought of indexically, as *those* preferences.[101] This may make it less unattractive to index 'ought' claims to the end of maximizing expected utility (of course it would still introduce mental states into the truth conditions, which Finlay wanted to avoid as well).

If we deny that 'ought' statements are ever relative to the end of maximizing expected utility, then the entailments between '*A* ought to *X*' and 'The balance of reasons favours that *A X*-es' would not be semantic or logical entailments. Or at least they would not be, if *Most Reason (differential)* is correct. For '*A* ought to *X* in order that life is protected', say, does not entail that *X*-ing has greater expected utility than doing something else (not matter whose preferences are involved).

In my view, it would not be a disaster to deny that '*A* ought to *X*' logically entails that the balance of reasons favours that *A X*-es. For if *Most Reason (differential)* is correct, there is still an intimate connection between 'ought' and the balance of reasons. But this relation is pragmatic: speakers will be prepared to say that *A* ought to *X* only if they judge that *X* has greatest expected utility (for the relevant individuals). So we can deduce that the balance of reasons favours *X* (at least from the point of view of the speaker) if we understand what motivates 'ought' statements in the first place.

[101] Thanks to Ralph Wedgwood.

This chimes in with Finlay's pragmatic explanation of the connection between recommendation and normative assertion (see chapter 7, section 7.2). That explanation required implicit knowledge of the fact that speakers are committed to the ends to which their utterances are indexed. If we are implicitly aware that the strength of reasons is not simply determined by the conditional probability of the end, but also on the end's importance to a subject, we will be able to infer from a sincere assertion of '*A* ought to *X*' that *X*-ing has greatest expected utility for the speaker.

There is a second reason why it would not be disastrous to deny that the relation between '*A* ought to *X*' and 'The balance of reasons favours that *A X*-es' is logical: phrases like 'the balance of reasons favours *X*' are hardly colloquial. So we don't have many data from everyday language to suggest that the entailments are logical, rather than pragmatic.

It is worth noting that Finlay would not be the first to deny that these entailments are logical. In (1984) David Wong analysed '*A* ought morally to *X*' in terms of the requirements of a system of moral rules. Relevant systems were ultimately determined by the speaker's mental states. But Wong was a Humean about reasons. And Humeans think that the truth of '*A* has a reason to *X*' depends on *A*'s desires. Since what is required by moral rules to which the speaker is committed does not entail anything about the agent's desires, Wong denied that '*A* ought to *X*' logically entails that there are moral reasons for *A* to *X*.[102] Instead, he offered an explanation of

[102] Similarly, Richard Boyd's naturalistic theory of 'good' in (1988) entails that statements about goodness do not logically entail statements about reasons.

why we normally expect other people to have reasons to act in accordance with the moral systems to which we are referring. Finlay can do something similar (and more persuasive): he can explain why '*A* ought to *X*' seems to entail that the balance of reasons favours that *A X*-es by means of people's implicit understanding of the motivation for normative assertions.

So I doubt that it would be lethal to deny that the entailments between '*A* ought to *X*' and 'The balance of reasons favours that *A X*-es' are logical.

8.7 'Ought' and multiple ends

It seems hard to avoid information about mental states in the truth conditions of claims about the weight of reasons. But thinking that weight is partly determined by the utility of ends is not incompatible with thinking that 'ought' statements themselves are indexed only to substantive ends. For example, a subject might judge that act *X* (which is conducive to *E*) has greater expected utility than other acts which are conducive to other ends. S/he may then say: '*A* ought to *X* in order that *E*'. This statement would be motivated by the fact that *X* has greatest expected utility for the subject, but would not report this fact.

Still, what we ought to do is not always relative to a single end. Sometimes an act is best because it is both conducive to E_1 and E_2: moving to Berlin might be best because it is both close to my family and affordable. Had it been close to my family but unaffordable, it may not have been best (and *vice versa*).

In some cases, we can simply index the 'ought' to the conjunction of both ends: I ought to move to Berlin because it is most conducive to E_1&E_2. But we cannot (always) make 'ought-in-the-light-of-multiple-ends' relative to the conjunction of all the ends that matter to us. One reason is that some may be incompatible (in a context). Finlay gives the example of Letty, who is trying to decide where to live (ms, chapter 7, p. 9). She has three criteria: close proximity to some members of her family (E_1), affordable housing (E_2), and a moderate climate (E_3). It may be that all her options are conducive to different pairs of these ends, but none is conducive to all three. If so, none of the options is at all conducive to the conjunction of all three ends. But, intuitively, there might still be an option that she ought to choose. Finlay proposes that this 'ought' is indexed to the most preferred conjunction of simultaneously realizable ends (ms, chapter 7, p. 12). In this sort of case, the information that this conjunction is most preferred is not part of the truth condition of '*A* ought to *X*'. The 'ought' is indexed to the conjunction of ends, and the fact that this is the most preferred combination is only pragmatically communicated.

This "preferred ends solution"[103] can be used to account for some 'ought' judgments relative to multiple ends. It takes into account expected utility in the sense that our preference for some achievable combination of ends might be informed by the utility of ends and their conditional probabilities. But I think there are still some cases that are not obviously covered.

[103] Referred to by Finlay as the "preferential selection solution" in his ms, chapter 7.

Suppose I have two ends: finding employment (E_1) and finding an affordable apartment (E_2). I have two options, with the following probabilities attached:

A = 0.7 likely to find employment, 0.3 likely to find an affordable apartment.

B = 0.5 likely to find employment, 0.5 likely to find an affordable apartment.

It seems that an agent may prefer A over B (judge that s/he *ought* to do A), although B is more conducive to the conjunction of E_1 and E_2.[104] If so, then it seems we cannot understand 'I ought to do A' (in this context) as meaning 'Given E_1&E_2, it is most likely that I do A'.[105] Or at least, we cannot understand this judgment in this way if E_1 and E_2 are, respectively, the end of *finding employment* and *finding an affordable apartment*. For A is not most conducive to the conjunction of these two ends (B is).

As far as I can see, there are two ways to understand the judgment that the agent ought to do A in this sort of case. One is suggested by a remark of Finlay's in (2009). With respect to 'ought' judgments made in the light of multiple ends, he says the following:[106]

[104] I take it that the conditional probability of a conjuction P&Q cannot be higher than the lowest conditional probability of one of the conjuncts (assuming that P and Q are independent).

[105] Nor can we understand it as meaning that A is most likely given just one of E_1 or E_2.

[106] Finlay does not make these remarks in response to the sort of problem I have raised. Rather, he suggests it to deal quite generally with cases where more than one end is relevant to an 'ought' judgment. In the draft of chapter 7 of his ms, he seems to have abandoned this idea in favour of the

'We have multiple ends and desires, and when we evaluate a means to one end, others often sit in judgement too. Such judgements will not be strictly instrumental evaluations, but rather decision-theoretic evaluations.' (Finlay (2009), pp. 325-326)

'The simplest way of interpreting these [decision-theoretic evaluations] friendly to the end-relational theory is as judgements of which means ought to be adopted in order that expected value is maximized.' (Ibid., footnote 26, p. 326)

One way to interpret this passage is as follows: '*A* ought to *X*' judgments are relative to the end of maximizing expected utility read *de dicto*, not *de re*. In other words: some judgments are not relative to some substantive end or a combination of substantive ends, but relative to the formal end of maximizing expected utility.

Of course, indexing 'ought' to the end of maximizing expected utility might mean giving up on the idea that the relation in which the speaker (or other people) stand(s) to ends is never part of the truth conditions of 'ought' claims. This is a cost, but I think that Finlay could afford it. Richard Joyce's objection to analyses in which that relation is part of the truth conditions has limited weight. The objection was that

---

preferred ends solution. However, that solution does not (seem to) extend to the kinds of case I describe above.

a judge who said that some hideous crime was wrong according to our standards would be saying "too little". But even if some 'ought' judgments are relative to the end of maximizing (the speaker's) expected utility, not all judgments will be. The Nuremberg judge's claim can plausibly be thought of as relative to the single end of protecting human life, or even the expected utility for the victims. The latter does not seem nearly as objectionable.

Of course there might be some cases where the judgment is best construed as relative to the end of maximizing our or the speaker's expected utility. But if this seems like saying "too little", we might be able to explain that appearance. Perhaps it seems too weak because of social processes that lead people to believe that morality is absolute. This is compatible with the view that the best theory of moral language is nonetheless relational, even to people's preferences. So this objection is not particularly strong. And even if none of the above responses work, the truth conditions of the relevant 'ought' judgments might be neutral about whose preferences are measured by the expected utility function (see also section 8.5). In that case, there is no objectionable relationality at all.

There is, however, a different problem.[107] The problem is that ends are ordinarily identified via people's desires or intentions. But it is not plausibly an object of anyone's desires or intentions to maximize expected utility (or, for that matter, to act in accordance with one's preferences). According to Finlay, the problem isn't that people cannot articulate something as complex as EU(X). It might still be plausible to

[107] Thanks to Ralph Wedgwood. Finlay also mentions it in his ms, chapter 7.

assign complex contents as the objects of mental states even if people cannot articulate those contents.[108] Rather, the problem is that we don't usually have higher-order desires or intentions, and don't deliberate with the aim of satisfying them (desires only play a role in the background of practical reasoning). It seems implausible to index 'ought' judgments to an end that nobody has (ms, chapter 7, p. 14).

It's not clear to me how strong this problem is. The interpretation of utterances is subject to a constraint of making good sense of them. It might be plausible to understand certain 'ought' judgments as relative to the end of maximizing expected utility (read *de dicto*) not because people desire to do so, but because they behave and reason in such a way as to end up maximizing expected utility.[109]

Should this argument fail, however, there is another way of thinking about cases where what one ought to do is not plausibly relative to the most preferred achievable conjunction of ends. The case I discussed was one where option *A* gave me a 0.7 chance of finding a job and a 0.3 chance of finding an affordable apartment, whereas *B* gave me 0.5 and 0.5 respectively. In this case, we couldn't relativize 'I ought to do *A*' to the conjunction of finding employment *and* finding an affordable apartment, since *B* was more conducive to that conjunction than *A*.

The way to avoid incorporating information about expected utility (and therefore mental states) into the truth conditions here, is to think of the end as

[108] We've seen in chapter 4, section 4.4, that David Copp uses a similar consideration to defend his standard-relational theory of normative judgment.

[109] Finlay uses this line of thought to defend the preferred ends solution in contexts of uncertainty (i.e. where conditional probabilities as well as utility values matter) (ms, chapter 7, p. 14).

involving probabilities, as follows: *having a 0.7 chance of finding employment & having a 0.3 chance of finding an affordable apartment*. Option *A* is most conducive to this conjunction.

The latter move seems *ad hoc* if the ends to which 'ought' judgments are indexed are not otherwise specified as giving one a certain *chance* of a state of affairs. And I don't really know how to answer this concern.[110]

8.8 'Better' and multiple ends

Whether we take the first or second solution to our problem concerning 'ought' and multiple ends, we find some difficulties with judgments of comparative goodness. Let's give these approaches a name for ease of reference. According to the first, some 'ought' judgments are relative to the end of maximizing expected utility. Let's call this

[110] Invoking such an end also makes trouble for other normative judgments, although this worry *can* be answered. Consider an option which gives me *no* chance of finding an affordable apartment, but a reasonable chance of finding a job. Since this option is *not* conducive to a chance of finding an affordable apartment, it is also not conducive to *having a 0.7 chance of finding a job & having a 0.3 chance of finding an affordable apartment*. Yet we still have *some* reason to choose this option.

This reason, of course, exists because the option is still conducive to *one* of our desired ends (that of finding employment). That suggests the following solution to the problem: although the 'ought' judgment is relative to the end of having a 0.7 chance of finding employment and a 0.3 chance of finding an affordable apartment, judgments of reasons are relative to *any* salient end. As long as the end of finding employment is still plausibly salient, we can explain why we judge that there is a reason to choose this suboptimal option.

the *expected utility approach* (or EU approach). According to the second, some 'ought' judgments are relative to a conjunctive end which specifies the degrees to which the act is conducive to the conjuncts (e.g. *having a 0.7 chance of X & having a 0.3 chance of Y*). Let's call this the *probabilistic conjunction approach* (or PC approach).

What one ought to do is the best out of a set of alternatives. But some suboptimal acts are still better than others. However, since being such as to maximize expected utility is not gradable (something either maximizes expected utility or it doesn't), no act other than the best is at all conducive to maximizing expected utility. That means that, relative to the end of maximizing expected utility, no suboptimal act is better than any of the others. That seems a bad consequence of the EU approach.

But the same problem arises on the PC approach. If the act which maximizes expected utility is the one given which the chance of finding employment is 0.7 and the chance of finding an affordable appartment is 0.3, then the PC approach says that the relevant end is *having a 0.7 chance of finding employment and a 0.3 chance of finding an affordable apartment*. Now suppose we have three options:

A is conducive to a 0.7 chance of *X* and a 0.3 chance of *Y*.

B is conducive to a 0.7 chance of *X*, but not to any chance of *Y*.

C is conducive to a 0.4 chance of *X*, but not to any chance of *Y*.

Intuitively, *A* is better than *B* and *B* is better than *C*. But if neither *B* nor *C* raises the probability of having a 0.7 chance of *X* and a 0.3 chance of *Y*, then *B* and *C* would be

equally good (after all, '*B* is better than *C*' was analysed as '*B* raises the probability of *E* more than *C*').

This problem can be solved for the PC approach by shifting the contextually salient end from comparison to comparison.[111] When we judge that *B* is better than *C* we implicitly switch our attention to the end of having a 0.7 chance of finding employment. *B is* more conducive to this end than *C* is, and thus better.

Notice that switching ends to account for comparative judgments is not only required in cases where the ends themselves involve probabilities (like in our example). It also arises in ordinary cases of multiple ends. Suppose that *A* is best because it is most conducive to finding a job (E_1) and finding an affordable apartment (E_2). If *B* is not conducive to E_2, but conducive to E_1, it is intuitively better than *C* which is neither conducive to E_1 nor to E_2. But neither *B* nor *C* are conducive to the *conjunction* of E_1 and E_2. So relative to this end, *B* is not better than *C*. In order to account for the sense that *B* is better than *C*, we have to switch to a different salient end (presumably E_1) and compare *B* and *C* in the light of this.

I think that the EU approach allows for a slightly different solution. In this case, we can keep the end constant (that of maximizing expected utility), but should restrict our focus on the options. Although suboptimal option *B* may not be any more conducive to maximizing expected utility than *C* if we compare it to *A*, *B* maximizes expected utility when we restrict ourselves to *B* and *C*.

[111] Finlay suggests this in the context of the preferred ends solution in ms, chapter 7, p.17.

The necessity of such manoeuvres are costs for the end-relational theory. They reduce the neatness of an otherwise powerful and simple approach. And it seems that similar mechanisms are required even in the case of single ends. Suppose that buying a cheap car is my only end. We have three choices: the first is the cheapest, the second cheap, but the third expensive. We want to say that 2 is better than 3, but 1 is best. But if 'better' means 'raises the probability of the salient end more', than 1 does not raise the probability that we buy a cheap car more than 2 (after all, being the cheapest is a way of being cheap). So if the end is to buy a cheap car, then 1 and 2 are equally good.

If we instead think of the end as that of buying the cheapEST car, then neither 2 nor 3 are conducive to doing so (if we know that 2 is not the cheapest). So 2 would not be better than 3. Either way, then, we get some bad results.

This problem can be solved by restricting our focus on options, in the same way as I suggested in the context of the EU approach. But it is disappointing that we have to invoke such manoeuvres even in the case of single ends.

Finlay proposes a different solution, which seems to me more drastic (personal communication; also ms, chapter 7, pp. 16-18). Finlay suggests that there is no single end at all (like buying a cheap car), but different ends, corresponding to the different prices of the cars on offer. Let us suppose that the first car costs 250 pounds, the second 500 and the third 1000 pounds. When we say that buying 2 is better than buying 3, we say that 2 is more conducive than 3 to the end of buying a car for 500 pounds. But when we compare 1 to 2, we say that 1 is more conducive than 2 to buying a car for 250 pounds. This solution strikes me as less good than restricting our

focus of attention, because there is no clear sense in which the agent has desires or intends to buy a car for 500 pounds (it is not plausibly an end of the agent's if s/he is comparing different options with an eye to buying a cheap car).

8.9 Conclusion

In this chapter, I have raised some difficulties for Finlay's end-relational theory. Most of these arose because normative judgments are often made in the light of several ends. We've seen that claims about the strength of reasons cannot be understood simply as claims about the extent to which an act raises the probability of an end. It also matters how *important* the end is. But importance cannot be understood merely in terms of probability. Following a suggestion by Finlay, I said that the weight of reasons can be understood in terms of expected utility (or alternatively, the extent to which an act is or would be preferred under certain conditions).

Of course, the expected utility of an act is partly determined by the degree to which a subject *desires* an end. But 'ought' statements were supposed to involve no information about the speaker's mental states. If so, then '*A* ought to *X*' cannot logically entail that the balance of reasons favours that *A X*-es. However, I have argued that there is still a strong pragmatic relationship between these claims: a subject will be prepared to say that *A* ought to *X* only if *A*'s *X*-ing has (for that subject) highest expected utility.

I have also argued that it is not so clear that information about mental states is never part of the truth conditions of 'ought' judgments. There seem to be cases where

we cannot simply index an 'ought' claim to whatever achievable combination of ends has greatest expected utility: the best option may not be most conducive to the conjunction of these ends.

There are solutions to this problem, but none is free of cost. The first is to index such 'ought' statements to the end of maximizing expected utility (the EU approach). This means that the truth conditions of some 'ought' judgments involve mental states (although not necessarily anyone's in particular). The other is to index such 'ought' claims to a conjunctive end which specifies the degrees to which the act is conducive to the conjuncts (the PC approach). This seemed *ad hoc*.

In addition to this problem, multiple ends also required some complex mechanisms to account for comparative judgments (*A* is better than *B*). Both the EU and PC approach were incompatible with indexing (certain) judgments of betterness to the same end as judgments of what is best. So we either had to restrict our focus of attention or switch the ends when comparing different options.

Despite these costs, however, the end-relational theory is still the best subjectivist theory of normative language, as I will argue in the next and final chapter.

Chapter 9. The best subjectivist theory of normative language

9.1 Introduction

In this thesis, I have discussed five types of subjectivist theory of normative language: noncognitivism, error-theories, Humeanism, standard-relational views and Finlay's end-relational theory. In this chapter, I will explain why Finlay's theory is superior to all the others. I will start by listing the (main) problems for the others and seeing how Finlay's theory fares with respect to them. I will then show that Finlay's theory allows a plausible response to the Problem of Detachment, before concluding that Finlay's is the best subjectivist theory of normative language.

9.2 Noncognitivism

Noncognitivism (as applied to normative language) is the idea that normative predicates do not contribute a property to the proposition expressed by normative sentences. Rather, their function is to express a noncognitive state, like a desire, an emotion or intention (at least in assertoric contexts). The main problem for this theory results from the fact that normative language interacts with nonnormative language is the same way as other nonnormative expressions (chapter 1, section 1.6). We can say, for instance: (1) 'Either the Americans landed on the moon, or they didn't', but also (2) 'Either murder is wrong, or the Americans landed on the moon'. It is not clear how to understand 'or' such that its meaning is the same in (1) and (2).

This problem occurs not just for connectives like 'or' and 'and', but also for propositional operators like 'to know' and 'necessarily'.

Finlay's theory avoids this problem because it is a truth conditional view of normative language. There is no problem in accounting for the sameness of meaning of operators like 'or' and 'and' in normative and nonnormative contexts, because normative sentences express factual propositions just like nonnormative sentences.

In addition, Finlay has a good explanation of why people are often motivated to act in accordance with their moral judgments (a fact sometimes taken to support noncognitivism about morals in particular). Finlay's theory says that moral judgments are about what is conducive to / compatible with certain moral ends. These ends are ordinarily ones that the speaker cares about (sincere moral judgment is not made from someone else's or society's perspective). This explains why John's assertion that we ought not to kill is ordinarily paired with (defeasible) motivation on the part of John to avoid murder (chapter 7, section 7.2).

At the same time, Finlay's theory does not have the vice of overgeneralizing. Many judgments about what we ought to do are not obviously linked with motivation ('In order to kill Bill, you ought to poison his drink'). Someone can make this judgment without being defeasibly motivated to poison Bill's drink. Finlay explains this because we can judge that some act would be conducive to killing Bill without wanting to Kill him ourselves.

In addition, Finlay explains why normative sentences typically communicate recommendation or endorsement in assertoric contexts (another fact sometimes taken to favour noncognitivism). His end-relational theory says that normative assertions

communicate the fact *that* some act or option is conducive to / compatible with certain ends. In those contexts where normative assertions also communicate recommendation or endorsement, the ends are desired by at least certain participants in the conversation. Given that interest in the end, it makes sense that *assertions* of conducivity / compatibility would tend to carry informational significance other than that of stating facts (chapter 7, section 7.2).

As before, Finlay avoids overgeneralization. In addition to explaining why some normative assertions are also recommendations or endorsements, he explains why others are not. This includes some assertoric contexts (ones where the speaker is not committed to the end in question). So while Finlay's theory has the virtues of noncognitivism, it avoids the vices.

9.3 Error-theories

No one I know of is an error-theorist about normative language in general. But there are error-theorists about *moral* language in particular. According to these people, all interesting moral statements are false, because they postulate non-existent moral properties and relations.

I have argued that this can't be a theory about the word 'ought' as used in moral contexts, since 'ought' occurs in many contexts without changing meaning (chapter 1, section 1.8). So the error must lie in the concept of morality, which in turn infects adjectives and adverbs like 'moral' and 'morally'.

An error-theory should be motivated by the way we use moral language. But the evidence does not clearly favour it (chapter 1, section 1.9). There is a more charitable interpretation of our practices. Behind it lies a theory of moral language according to which it serves more than a merely descriptive purpose. It is also used to protect and promote certain values (chapter 2, section 2.7). This explains why we assess a moral statement as false *from our own moral perspective* (relative to our own moral ends).

At the same time, Finlay's particular take on moral language allows many moral judgments to be true. After all, we will often be right that an act is conducive to / compatible with certain moral ends. So Finlay avoids large-scale error in what seems like a respectable practice of evaluation. And since charity requires us to ascribe such error only as a last resort, I think the end-relational theory is preferable even in the case of moral language.

9.4 Humeanism

Humean views of normative language typically tie the truth conditions of normative statements to the desires (or other motivational states) of *the agent*. For example, the directive 'ought' judgment '*A* ought to *X*' is true just in case *A* would be motivated to *X* under certain conditions (Harman's version).

This leads to a number of problems (chapter 3, section 3.3). First, it does not seem as if all our assessments of what reasons there are to act are sensitive to information about the agent's psychology. We often judge that something is wrong

(i.e. ought not to be done) without any attention to the agent's desires. This creates a presumption against the Humean analysis. We should tie the truth conditions of normative judgments as closely as we can to what they seem to be responsive to.

Furthermore, the fact that many of our normative judgments are not sensitive to the agent's desires predicts large-scale falsehood in normative judgment. It may not be true that the agent is motivated in the same way as we are. But, intuitively, the normative assessment may still be true. So (at least some) Humeans take an uncharitable view of normative judgment: they are very often false.

Finlay's theory avoids both problems. Normative judgments are plausibly sensitive to the relation between acts and certain ends. Since the ends needn't be the agent's, Finlay allows that many judgments can be true even if the agent is not motivated towards the ends which they (implicitly) refer to. So the theory is more charitable than Humeanism, and forges a closer link between the truth conditions of normative judgments and what they are responsive to.

Humeanism also had difficulty accounting for the distinction between 'ought' and 'must'. 'Must' seems stronger than 'ought', for example in the sense that 'must' entails 'ought' but not the other way around. It is unclear how a theory that makes 'ought' about an agent's motivations could account for this. Schroeder avoids motivations, but lacked an intelligible account of what the weight of reasons was determined by (this mattered since 'ought' was analysed in terms of the balance of reasons). So there is no way of telling whether Schroeder might make a plausible distinction beween 'ought' and 'must'.

Finlay, on the other hand, has a persuasive model: whereas 'ought' concerns what is most likely given certain ends, 'must' concerns what is necessary given certain ends (chapter 6, sections 6.2 and 6.3). That explains why 'must' entails 'ought', because if *X* is true in all possible worlds compatible with certain ends, then *X* is also most likely given those same ends. But not the other way around: what is most likely is not always necessary. So Finlay's theory can account for the perceived difference in meaning between 'ought' and 'must'.

9.5 Standard-relational views of normative language

In chapters 4 and 5 I discussed standard-relational views of normative language. According to theories of this type, what one ought to do is a matter of the requirements of imperatives. In the most plausible version, there are many imperatives to account for the multiplicity of reasons. What one ought to do is determined by what is entailed by imperatives with the highest priority, whereas reasons are facts which explain why an imperative (not just the ones at the top) entail that the agent acts a certain way. The theory had a number of options for 'better' and related terms like 'good' and 'best'. Two of them defined 'better' in terms of what we ought to choose and one in terms of greater degrees of property exemplification.

This view was not unpromising but struggled with a number of things: first, it had to deny that Buridan's ass had any reason to choose the left stack of hay (and similarly for the right stack). It also had to deny that there really was some reason to choose 10 pounds in a choice between losing 10, gaining 10 or gaining 50. We didn't

lack resources to explain (away) these intuitions, but it seemed preferable to have a theory which allowed that there was some reason in both cases.

Finlay's theory does this. He defines a (*pro tanto*) reason to *X* as a fact which explains why it in in some way good to *X*. This in turn was analysed in terms of being conducive to some contextually salient end. If the relevant end is the survival of the ass, then there is a reason to choose the left stack and a reason to choose the right stack, because there are facts which explain why doing so raises the probability of its survival. Similarly, if the relevant end is making profit, then there is some reason to choose the 10.

Things are not unlike the standard-relational theory if the end is not to make some profit, but to *maximize* your profit. You have equally little reason to choose suboptimally relative to that end, since choosing 10 does not raise the probability of *maximizing* profit if there is a better option.

The main advantage for Finlay lies in the fact that he makes a clear distinction between 'must' and 'ought'. The standard-relational theory, on the other hand, could not do this, since it had to say that the truth conditions of 'ought' are the same as those of 'must'.

Finlay also does a better job of explaining what is common between the normative and the epistemic 'ought'. They probably share a common core, since 'ought' allows both readings in many languages other than English. Since the standard-relational theory defined the normative 'ought' in terms of what is required (entailed) by rules, it seemed natural to analyse the epistemic 'ought' in terms of what the evidence requires. But the evidence may favour *p* without *requiring* (or *entailing*) *p*.

Finlay's probabilistic analysis of the epistemic 'ought' ('It ought to rain tomorrow') explains why the relation between the evidence and *p* can be weaker than necessitation. The relative weakness of being most likely also explains why the normative 'ought' is weaker than the normative 'must': the first says that the performance of an act is most likely (on the assumption that some end has been achieved), whereas the second says that its performance is necessary (on the assumption that some end has been achieved).

So Finlay allows more reasons than the standard-relational theory and explains the difference between 'must' and 'ought'.

9.6 The problem of detachment

Finlay's analysis of 'ought' has a further advantage: it allows a plausible solution to the Problem of Detachment (discussed in Greenspan (1975), Darwall (1983), Broome (1999) and Bedke (2009), among others). The problem is that it seems possible to "detach" intuitively false 'ought' claims from hypothetical imperatives supplemented with suitable premises. Take the following example:

(1) If you want to kill Bill, then you ought to poison his drink.

(2) You want to kill Bill.

(3) Therefore, you ought to poison his drink.

But, as Matthew Bedke puts it, 'you ought not poison Bill's drink. You ought to avoid him and seek counseling' ((2009), p. 673).

Sine I'm not sure that Finlay's response to this problem is adequate, I will develop my own answer instead. In (2009), Finlay says that '[an ought claim's] antecedent doesn't concern an agent's motivations, so the consequent doesn't detach from facts like those' (p. 324). The idea is as follows: 'ought' claims are interpreted as statements of conditional probability. The "antecedent" of such claims ('Given *B* including *E*, ...') does not concern an agent's desires (even if *E* is sometimes *determined* by his or her desires). So there is a sense in which the fact that you *want* to kill Bill does not license the conclusion.

But whether this response really shows that (3) cannot be deduced from (1) and (2) depends on the interpretation of (1). On the end-relational account, 'ought' is always implicitly relational: you ought to *X in order that E*. In some hypothetical imperatives, the end is made explicit, as in 'In order to evade arrest, Max ought to mingle with the crowd'. But the antecedent of (1) is 'If you *want* to kill Bill' and it makes no sense to say 'You ought to poison Bill's drink in order that you want to kill him'.

So Finlay claims that hypothetical imperatives of the type 'If you want to *Y*, then you ought to *X*' are a kind of relevance conditionals: 'If you want a biscuit, there are some on the table' ((2010b), p. 83; ms, chapter 4, pp. 24-25). This is not an ordinary conditional because the antecedent stands in a different relationship to the consequent: the presence of biscuits on the table is not highly probable given that you want some biscuits, nor does it counterfactually depend on your desire. Rather, the

antecedent makes salient something in the light of which the information in the consequent is relevant. Finlay believes that the antecedent of (1) does the same. 'If you want to kill Bill' makes salient that you have as an end that Bill is killed. In the light of this information it is relevant to make an 'ought' statement which is plausibly indexed to the salient end (not the desire). So (1) really means (something like): 'If you want to kill Bill, then in order to kill him, you ought to poison his drink ((2010b), p. 83, ms, chapter 4, pp. 24-25).

If this view of (1) is correct, however, then (3) *does* follow from the premises, if the end to which the 'ought' in the conclusion is indexed is the same end involved in the consequent of (1). For if so, then the argument can be rendered as follows:

(1*) If you want to kill Bill, then in order to kill him you ought to poison his drink.

(2*) You want to kill Bill.

(3*) Therefore, in order to kill Bill you ought to poison his drink.

In my view, it is *good* to allow an interpretation of (1) – (3) under which the argument is valid. For I see the Problem of Detachment as a kind of paradox: on the one hand, (1) – (3) seems like a valid argument. But it also seems invalid. The challenge is to explain why this is so.[112]

[112] If my take on the Problem of Detachment is correct, then it is also important to give premise (2) an indispensable role. Finlay's interpretation of "want-conditionals" as relevance-conditionals does just this.

Finlay's theory of "want-conditionals" (statements of the form 'If you want to *Y*, then you ought to *X*') allows us to explain why the argument seems valid. It is valid just when the 'ought' in (3) is indexed to the same end as the 'ought' in (1). But it also allows us to explain why the argument can seem invalid. This explanation goes as follows: we object to (3) because of an implicit switch in our normative perspective (the ends relative to which we assess whether you ought to poison Bill's drink). We assess the correctness of (1) by focussing on the end of killing Bill. But when we object to (3), we do so because we have implicitly switched to a broader normative perspective (one including moral ends). The switch is plausibly due to the fact that the 'ought' in (3) is surface-categorical. This may suggest that it is all-things-considered (i.e. valid in the light of *all* relevant ends, not just that of killing Bill).

So the Problem of Detachment is essentially defused as follows: although one can "detach" that you ought to poison Bill's drink *in order to kill him*, you cannot detach that you ought to poison Bill's drink in order to achieve any other end. So it does not follow from (1) and (2) that you ought *morally* to poison Bill's drink, or that you ought to do this *all things considered*. Detaching the conclusion is compatible with your being morally required to avoid him and seek councelling. That should reduce the sense that there is anything wrong with the conclusion. For if we accept that (1) is true (i.e. that if you want to kill Bill, then you ought to poison his drink), we shouldn't object to: 'In order to kill Bill, you ought to poison his drink' either.[113]

[113] My response to the Problem of Detachment is possibly implicit in the following remarks by Finlay (although I'm not sure that I quite understand them): '[...] whether the antecedent obtains [whether the 'If you want'-clause is satisfied] is not of relevance to our interest in these propositions

9.7 Summing up

This thesis started with the following question: *On the assumption that there are no objective normative facts, what is the best theory of normative language?*. I believe that Finlay's end-relational theory is the best, for the following ten reasons: (1) it is a truth conditional, (2) non-error-theoretic account that (3) straightforwardly explains why it can be simultaneously true that *A* ought morally to *X* but prudentially to not-*X*. (4) It does not convict normative language of large-scale error, like Humean accounts. (5) It explains why normative judgment is sometimes accompanied by corresponding motivation and (6) why its expression communicates endorsement or recommendation in certain contexts but not others. (7) It provides a simple and unified account not just of normative modals, but also of their nonnormative counterparts. This is important, since the evidence suggests that they share a common core of meaning. (8) It can plausibly account for the meaning of a variety of other terms, like 'reason' and 'good'. (9) It does not make epistemic 'ought's and reasons objectionably subjective. It does not matter whether agents *care* about the truth in

---

[propositions expressed by sentences like 'You ought to poison Bill's drink'], which rather addresses comparative probabilities given merely possible circumstances. By asserting propositions concerning probabilities that are conditional on counter-to-fact circumstances, we succeed in giving practical advice and evaluating actions. To even think about detaching the consequent is therefore to miss the point.' (Finlay (2009), p. 324).

order for them to have epistemic reasons. (10) It allows a plausible response to the Problem of Detachment.

For all these reasons, I believe that Finlay's theory is the best subjectivist theory of normative language. It counts as subjectivist because the ends relative to which we assess whether something is good, wrong or permissible can (at least in certain interesting cases) vary with people's subjective mental states.

Of course, being best does not mean being free of problems. It is a comparative matter. The end-relational theory struggles most with normative assessment in the light of many ends (chapter 8). This might require relativizing at least some judgments to the end of maximizing expected utility. This introduces mental states into their truth conditions and invites an epicycle: options known to have less than the highest expected value do not raise the probability that expected value is maximized. So in order to ground judgments like '*A* is better than *B*' (where *A* and *B* are both suboptimal), we have to restrict the class of alternatives to *A* and *B* alone (or at least we have to keep out options superior to *B*).[114] Once we've restricted the class of alternatives like this, we can explain why *B* is more conducive to maximizing expected utility than *A*: given just *A* and *B*, *B* maximizes expected value (and thus raises its probability to 1). We also saw that such manoeuvres are required in certain cases with a single end (chapter 8, section 8.8).

[114] Or, alternatively, we have to switch the contextually salient ends.

And there is another problem. Finlay's theory allows that '*A* ought to *X*' *logically* entails that the balance of reasons favours that *A X*-es (and *vice versa*) only when 'ought' is relative to the end of maximizing expected utility.

I think the latter problem is not weighty enough to sink Finlay's theory. Given that '*A* ought to *X*' is typically said *because X*-ing maximizes expected value, there is a strong pragmatic link between such statements and claims about the balance of reasons.

The first problem (or set of problems) strikes me as more damaging. Normative judgments are often made in the light of different ends. Of course many can be indexed to individual (substantive) ends or the most preferred achievable conjunction. So many 'ought's can be true in virtue of raising the probability of these ends more than relevant alternatives. But not all can (as argued in chapter 8, section 8.7). Some apparently true judgments will either have to be (1) indexed to maximizing expected value or (2) indexed to ends like *having a certain chance of E*, or (3) simply not be true. Since no other 'ought' judgments involve EUT, the first may seem *ad hoc*, as does the second. But the third is simply wrong. Finally, it seems inelegant to suppose that judgments of comparative goodness often involve restricting the focus of attention or switching salient ends.

I suspect that a superior theory might substitute Finlay's preferred relation to ends (that of probability raising) for a different one. But it won't be easy to say exactly what it is, as chapter 5 suggests. Still, a good alternative would probably retain many elements of Finlay's work (his theory of 'must' and 'may', the logical form of statements about goodness, the response to the Problem of Detachment, the

commonality between normative and epistemic 'ought's, etc.). Certainly compared to its famous competitors, Finlay's is the best subjectivist theory of normative language.

References

Ayer, A. (1936), *Language, Truth and Logic*, London: Victor Gollancz Ltd.

Bealer, G. (1989), "*On the Identification of Properties and Propositional Functions*", *Linguistics and Philosophy*, 12: 1-14

Bedke, M. (2009), "*The Iffiest Oughts: a Guise of Reasons Account of End-Given Conditionals*", *Ethics*, 119: 672-698

Binderup, L. (2008), "*Brogaard's Moral Contextualism*", *The Philosophical Quarterly*, 58: 410-415

Björnsson, G. & Finlay, S. (forthcoming), "*Metaethical Contextualism Defended*", *Ethics*

Blackburn, S. (1988), "*Attitudes and Contents*", *Ethics*, 98: 501-517

Blackburn, S. (1992), "*Through Thick and Thin*", *Proceedings of the Aristotelian Society*, supplementary volume 66: 285-299

Blackburn, S. (1993), *Essays in Quasi-Realism*, New York: Oxford University Press

Blackburn, S. (1998), *Ruling Passions. A Theory of Practical Reasoning*, Oxford: Clarendon Press

Boghossian, P. (2006), "*What is Relativism?*", in Greenough, P. & Lynch, M. (2006), *Truth and Realism*, Oxford: Clarendon Press

Boyd, R. (1988), "*How to Be a Moral Realist*", in Sayre-McCord, G. (ed.) (1988b), *Essays on Moral Realism*, Ithaca: Cornell University Press

Brogaard, B. (2008), "*Moral Contextualism and Moral Relativism*", *Philosophical Quarterly*, 58: 385-409

Broome, J. (1993), *Weighing Goods: Equality, Uncertainty and Time*, Oxford: Basil Blackwell

Broome, J. (1999), "*Normative Requirements*", Ratio, 12: 398–419

Broome, John (2004), "*Reasons*", in Jay Wallace, R. & Pettit, P. & Scheffler, S. & Smith, M. (eds.) (2004), *Reason and Value. Themes from the Moral Philosophy of Joseph Raz*, Oxford: Clarendon Press

Copp, D. (1995), *Morality, Normativity, and Society*, New York: Oxford University Press

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

Davidson, D. (1967), "*Truth and Meaning*", *Synthese*, 17: 304-323

DeRose, K. (1992), "*Contextualism and Knowledge Attributions*", 52: 913-929

DeRose, K. (2010), "*"Bamboozled by Our Own Words": Semantic Blindness and Some Arguments against Contextualism*", *Philosophy and Phenomenological Research*, 73: 316-338

Evers, D. (2010), "*The End-Relational Theory of 'Ought' and the Weight of Reasons*", *dialectica* 64: 405-417

Evers, D. (forthcoming), "*The Standard-Relational Theory of 'Ought' and the Oughtistic Theory of Reasons*", *Australasian Journal of Philosophy*

Egan, A., Hawthorne, J. & Weatherson, B. (2005), "*Epistemic Modals in Context*", in Preyer, G. & Peter, G. (eds.) (2005), *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, Oxford: Oxford University Press

Finlay, S. (2004), "*The Conversational Practicality of Value Judgement*", *The Journal of Ethics*, 8: 205-223

Finlay, S. (2006). "*The Reasons that Matter*", *Australasian Journal of Philosophy*, 84: 1-20

Finlay, S. (2008), "*The Error in the Error Theory*", *Australasian Journal of Philosophy*, 86: 347-369

Finlay, S. (2009), "*Oughts and Ends*", *Philosophical Studies*, 143: 315-340

Finlay, S. (2010a), "*Normativity, Necessity, and Tense. A Recipe for Homebaked Normativity*", in Shafer-Landau, R. (2010), *Oxford Studies in Metaethics Vol. 5*, Oxford: Oxford University Press

Finlay, S. (2010b), "*What Ought Probably Means, and Why You Can't Detach It*", *Synthese*, 177: 67-89

Finlay, S. (forthcoming), "*Errors upon Errors: a Reply to Joyce*", *Australasian Journal of Philosophy*

Finlay, S. (ms), *Confusion of Tongues: a Theory of Normativity* (available upon request at http://www-bcf.usc.edu/~finlay/)

Foot, P. (1961), "*Goodness and Choice*", *Proceedings of the Aristotelian Society*, supplementary volume 35: 45-60

Geach, P. (1960), "*Ascriptivism*", *The Philosophical Review*, 69: 221-225

Geach, P. (1965), "*Assertion*", *The Philosophical Review*, 74: 449-465

Gibbard, A. (1990), *Wise Choices, Apt Feelings. A Theory of Normative Judgment*, Oxford: Clarendon Press

Gibbard, A. (2003), *Thinking How to Live*, Cambridge, Massachusetts: Harvard University Press

Gill, M. & Nichols, S. (2008), "*Sentimentalist Pluralism: Moral Psychology and Philosophical Ethics*", *Philosophical Issues*, 18: 143-163

Gert, J. (2005), "*A Functional Role Analysis of Reasons*", *Philosophical Studies*, 124: 353-378

Grice, P. (1975), "*Logic and Conversation*", in Cole, P. & Morgan, J. (eds.) (1975), *Syntax and Semantics, 3: Speech Acts*, New York: Academic Press

Greenspan, P. (1975), "*Conditional Oughts and Hypothetical Imperatives*", *The Journal of Philosophy*, 72: 259–276

Hare, R. (1952), *The Language of Morals*, Oxford: Clarendon Press

Hare, R. (1970), "*Meaning and Speech Acts*", *The Philosophical Review* 79: 3-24

Harman, G. (2000a), "*Moral Relativism Defended*", in Harman, G. (2000e), *Explaining Value and Other Essays in Moral Philosophy*, Oxford: Clarendon Press

Harman, G. (2000b), "*Relativistic Ethics: Morality as Politics*", in Harman, Gilbert (2000e), *Explaining Value and Other Essays in Moral Philosophy*, Oxford: Clarendon Press

Harman, G. (2000c), "*Justice and Moral Bargaining*", in Harman, Gilbert (2000e), *Explaining Value and Other Essays in Moral Philosophy*, Oxford: Clarendon Press

Harman, G. (2000d), "*Is There a Single True Morality?*", in Harman, Gilbert (2000e), *Explaining Value and Other Essays in Moral Philosophy*, Oxford: Clarendon Press

Harman, G. & Thomson, J. (1996), *Moral Relativism and Moral Objectivity*, Cambridge, Massachusetts: Blackwell

Heathwood, C. (2009), "*Moral and Epistemic Open-Question Arguments*", *Philosophical Books*, 50: 83-98

Horty, J. (2007), "*Reasons as Defaults*", *Philosophers' Imprint*, 7: 1-28

Horwich, P. (1990), *Truth*, Oxford: Oxford University Press

Jeffrey, R. (1965), *The Logic of Decision*, Chicago: The University of Chicago Press

Joyce, J. (1999), *The Foundations of Causal Decision Theory*, Cambridge: Cambridge University Press

Joyce, R. (2001), *The Myth of Morality*, Cambridge: Cambridge University Press

Joyce, R. (forthcoming), "*The Error in "The Error in the Error Theory"*", *Australasian Journal of Philosophy*

Kalderon, M. (2005), *Moral Fictionalism*, Oxford: Oxford University Press

Kearns, S. & Star, D. (2009), "*Reasons as Evidence*", in Shafer-Landau, R. (ed.) (2008), *Oxford Studies in Metaethics Vol. 4*, Oxford: Oxford University Press

Kelly, T. (2003), "*Epistemic Rationality as Instrumental Rationality: a Critique*", *Philosophy and Phenomenological Research*, 66: 612-640

Kölbel, M. (2002), *Truth without Objectivity*, London: Routledge

Kratzer, A. (1977), "*What 'Must' and 'Can' Must and Can Mean*", *Linguistics and Philosophy*, 1: 337–355

Kratzer, A. (1981), "*The Notional Category of Modality*", in: Eikmeyer, H. & Rieser, H. (eds.), *Words, Worlds, and Contexts*, Berlin: Walter de Gruyter

Lycan, W. (2010), "*Direct Arguments for the Truth-Condition Theory of Meaning*", *Topoi*, 29: 99-108

MacFarlane, J. (2005), "*Making Sense of Relative Truth*", *Proceedings of the Aristotelian Society*, 105: 321-339

MacFarlane, J. (2007), "*Relativism and Disagreement*", *Philosophical Studies*, Vol. 132: 17-31

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

Moore, G. (1903), *Principia Ethica*. Cambridge: Cambridge University Press

Nolan, D., Restall, G. & West, C. (2005), "*Moral Fictionalism Versus the Rest*", *Australasian Journal of Philosophy*, 83: 307-330

Parfit, D. (ms), *On What Matters*, University of Oxford

Prinz, J. (2007), *The Emotional Construction of Morals*, Oxford: Oxford University Press

Ragnar, F. (2008), *Metaethical Relativism. Against the Single Analysis Assumption*, doctoral dissertation, University of Göteburg

Railton, P. (1996), "*Moral Realism: Prospects and Problems*", in Sinnott-Armstrong, W. & Timmons, M. (eds.) (1996), *Moral Knowledge?*, New York: Oxford University Press

Ridge, M. (2006), "*Ecumenical Expressivism: Finessing Frege*", *Ethics*, 116: 302-336.

Ridge, M. (2007), ""*Ecumenical Expressivism: The Best of Both Worlds?*", in Shafer-Landau, R. (ed.) (2007), *Oxford Studies in Metaethics* 2, Oxford: Clarendon Press

Sayre-McCord, G. (1988a), "*The Many Faces of Moral Realism*", in Sayre-McCord, G. (ed.) (1988b), *Essays on Moral Realism*, Ithaca: Cornell University Press

Searle, J. (1962), "*Meaning and Speech Acts*", *The Philosophical Review* 71: 423-432

Schroeder, M. (2007). *Slaves of the Passions*. Oxford: Oxford University Press

Schroeder, M. (2008), *Being For: Evaluating the Semantic Program of Expressivism*, Oxford: Oxford University Press

Schroeder, M. (2009), "*Hybrid Expressivism: Virtues and Vices*", *Ethics*, 119: 257-309

Schroeder, M. (2010), *Noncognitivism in Ethics*, London: Routledge

Schueler, G. (1988), "*Modus Ponens and Moral Realism*", *Ethics*, 98: 492-500

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

Sloman, A. (1970), "*Ought and Better*", *Mind*, 79: 385–394

Strawson, P. (1950), "*Truth*", *Proceedings of the Aristotelian Society*, supplementary volume 24: 111–56

Streiffer, R. (2003), *Moral Relativism and Reasons for Action*, London: Routledge

Timmons, M. (1999), *Morality without Foundations. A Defense of Ethical Contextualism*, New York: Oxford University Press

Velleman, D. (2000), *The Possibility of Practical Reason*, Oxford: Oxford University Press

Von Fintel, K. & Iatridou S. (2008), "*How to Say Ought in Foreign: The Composition of Weak Necessity Modals*", in Guéron, J. & Lecarme, J., eds., Time and Modality, Dordrecht: Springer

Von Neumann J., & Morgenstern O., *Theory of Games and Economic Behavior* (2nd edition), Princeton: Princeton University Press

Von Wright, G. (1963), *The Varieties of Goodness*, London: Routledge & Kegan Paul

Wedgwood, R. (1997), "*The Essence of Response-Dependence*", *European Review of Philosophy*, 1: 31-54

Wedgwood, R. (2010), "*Schroeder on Expressivism: For – or Against?*", *Analysis* 70: 117-129

Wiggins, D. (1998a), "*Truth, Invention, and the Meaning of Life*", in Wiggins, D. (1998b), *Needs, Values, Truth*, Oxford: Clarendon Press

Williams, B. (1981a), "*Internal and External Reasons*", in Williams, B. (1981b) *Moral Luck. Philosophical Papers 1973-1980*, Cambridge: Cambridge University Press

Wong, D. (1984), *Moral Relativity*, Berkely: University of California Press

Yaqub, M.Z., Saz, G. & Hussain, D. (2009), "*A Meta-Analysis of the Empirical Evidence on Expected Utility Theory*", *European Journal of Economics, Finance and Administrative Sciences*, 15: 117–133