Objective Epistemic Consequentialism

Amanda Askell BPhil Thesis, University of Oxford, 2011 Supervisor: Scott Sturgeon

Contents

Abstract				3
Introduction				4
1	The	Struc	ture of Consequentialist Theories	6
	1.1	Hedon	istic Act Utilitarianism	6
	1.2	The T	hree Components of Consequentialist Theories	11
		1.2.1	Contributory Final Value	11
		1.2.2	All-Things-Considered Value	15
		1.2.3	A Deontic Theory	23
2	Epis	stemic	Consequentialism	33
	2.1	An Ep	istemic Consequentialist Theory	33
		2.1.1	An Epistemic Theory of Contributory Final Value	33
		2.1.	1.1 The Problem of Trivial Truths	43
		2.1.2	An Epistemic Theory of All-Things-Considered Value	45
		2.1.3	An Epistemic Deontic Theory	55
	2.2	Object	ive and Subjective Epistemic Oughts	61
3	An Application of Objective Epistemic Consequentialism			
	3.1	The P	roblem of Peer Disagreement	66
	3.2	An Ep	istemic Consequentialist Solution	71
Conclusion 7				
Bibliography 7				

Abstract

What should we believe? This is the central question guiding value-driven epistemology. In this thesis I will construct and defend a position that I call *objective epistemic consequentialism*, whose central purpose is to provide an answer to this question. At its core, objective epistemic consequentialism states that we ought to believe a proposition P to a given degree if and only if doing so produces the most *epistemic* value. However, by giving a comprehensive account of objective epistemic consequentialism, I develop a theory which offers a viable response not only to the question of what we should believe and why, but also to which decision procedures we should commit ourselves to, what is of final epistemic value, and what is the nature of epistemic oughts.

In chapter 1 I will identify and explore the three primary components of consequentialist theories in ethics. I will then build on this framework in chapter 2 to offer an epistemic consequentialist account of each of these three components. This constitutes a detailed account of objective epistemic consequentialism. Finally, in chapter 3 I show that the machinery of objective epistemic consequentialism can be used to resolve issues in epistemology that have thus far been deeply problematic. The example I consider is the problem of *peer disagreement*. Not only is objective epistemic consequentialism is able to adequately explain the competing intuitions in these cases, it also produces plausible recommendations for what agents ought to do when faced with peer disagreement. I therefore conclude that objective epistemic consequentialism is an extremely attractive and fruitful normative theory for epistemologists to adopt.

Introduction

The the recent literature on epistemic normativity has approached the topic from two very different directions. On the first approach to epistemic normativity, belief-formation is treated as goal-directed. The goal of belief has standardly been understood as acquiring true beliefs and avoiding false beliefs. But several problems have been raised for solely truth-directed accounts of value. As such, those who take the first approach to epistemic normativity still assume that belief-formation is goal-directed, but they propose alternative and more detailed accounts of epistemic value. Perhaps it is not true beliefs that are of value in epistemology, but true beliefs in important propositions, or justified true beliefs, or knowledge, or some combination of these things.

The second approach, however, comes from those who reject or are even outrightly hostile to the goal-directed account of what it is for something to be epistemically valuable. For example, virtue epistemologists reject the consequentialist account of the value of beliefs described above. Instead, they have argued, epistemic value is produced not by beliefs, but by epistemic agents that display the appropriate intellectual virtues.

The first approach assumes that epistemic normativity occurs within a consequentialist framework, and explores the different possible accounts of epistemic value that are the goal of our cognitive acts. The second approach rejects this framework, and explores alternative accounts of epistemic normativity. In this thesis I will attempt to combine these two approaches. I will explore and defend the *consequentialist framework* that has been assumed by those who take the first approach to the issue of normativity, but with the same kind of rigour that has been employed to virtue-theoretic accounts by those who take the second.

The primary goal of this thesis is to provide a coherent and detailed objective epistemic consequentialist theory. But how are we to go about doing this? The first task is to get clear on the deep mechanics of consequentialist theories in general: what is their account of value? what things do they assess for value? and how do they derive 'oughts' from this account of value? Articulating each of these questions and the issues they raise will be necessary if we want to construct a comprehensive consequentialist theory within epistemology. This will be the main task of chapter 1. In this chapter I will expose the foundational issues driving consequentialist theories more generally, in order to apply them to the epistemic case in the chapters that follow. Once we are clear on the mechanics of consequentialist theories more generally, we can hone in on each of the key variables and components of the theories and attempt to find the best way to articulate these in an epistemic framework: what is of *epistemic* value? and how should the *epistemic* ought relate to this account of value? Providing a clear, detailed and fully worked-out epistemic consequentialist theory that can answer to each of these questions is the main task of chapter 2.

By constructing a detailed account of each of the components of objective epistemic consequentialism, I will show that we are able to capture a host of intuitions that are usually seen to be inconsistent or at odds with a consequentialist normative theory. This will be especially evident in chapter 3, where I argue that even if we adopt a strong externalist view of evidence, epistemic consequentialism can vindicate our conciliatory intuitions in cases of peer disagreement. As such, not only do I seek to give a greater level of depth and clarity to epistemic consequentialism than has thus far been offered in the literature in this thesis. I also seek to offer a defence of the deeper and more clarified formulation of epistemic consequentialism to consequentialism, and its surprising resiliency in the face of common objections to consequentialism, and its similarly surprising ability to offer an intuitive resolution to some deeply problematic cases in epistemiology.

1 The Structure of Consequentialist Theories

There are several very different theories within normative ethics that fall under the general category of 'consequentialist'. At the most basic level, a theory in ethics is considered to be a consequentialist theory if it stipulates that the moral status of an evaluative focal point such as an act or a rule, depends either solely or predominantly on its consequences.¹ In this chapter I will attempt to delineate a general framework that I argue a large proportion of consequentialist theories share. I do so in order to demonstrate how it would be best to characterise epistemic consequentialism in chapter 2. It will helpful to begin, however, by considering an example of a specific consequentialist theory. Therefore in section 1.1 I will offer an analysis of *hedonistic act utilitarianism* from normative ethics. Building on this analysis, I will draw a more general picture of consequentialist theories in section 1.2.

1.1 Hedonistic Act Utilitarianism

Hedonistic act utilitarianism is a position which states that the rightness of an act depends on the amount of happiness produced by that act. In order to differentiate hedonistic act utilitarianism first from non-consequentialist theories and then from other consequentialist theories, let us consider what each would say about the following example:

Forced Organ Donation

Jane is a doctor who recently started work in a large inner-city hospital. The hospital is drastically under-staffed and one evening Jane finds herself alone in the overnight clinic. She begins to contemplate how tragic it is that there are so many patients in the hospital that will die if they don't receive organ transplants within the next few days. A lone hitchhiker enters the clinic to ask if a minor cut has become infected. Jane confirms that it hasn't, while the hitchhiker tells her about his luckless life, his lack of prospects, and laments about his diabetes. Jane notices an organ donor card in the hitchhiker's wallet and realises she has been presented with a rare opportunity. She tells the hitchhiker that she needs to give him a Tetanus shot, but instead she fills the needle with a massive dose of insulin. Jane knows that if she injects the hitchhiker with such a large does of insulin, the man will quickly fall into a diabetic coma and die. She also knows that the hospital is badly administrated, and the missing needle and vials of insulin will go unnoticed. Jane can simply claim that the man entered the hospital showing symptoms of insulin overdose before he quickly lapsed into a coma and died. As a result, the hitchhiker's organs would be donated to patients that need them: saving several lives and improving the quality of several others.

¹The term 'evaluative focal points' is taken from Kagan (2000). I will discuss these in greater detail in section 1.2.2. For now it is sufficient to understand an evaluative focal point as something, such as a rule, act, intention or character trait, which can be assessed in terms of the goodness of its consequences.

Should Jane administer the injection to the hitchhiker? There is a stark difference in the answer given to this question by hedonistic act utilitarianism and the two most prominent families of non-consequentialist theories: deontological theories and virtue ethicist theories. Theories of the latter variety are much more likely to caution Jane against administering the injection even though doing so will save more lives than are lost (i.e. the hitchhiker's), while hedonistic act utilitarians are far more likely to urge Jane to administer the injection, or at least to argue that it would be morally permissible for her to do so.

Deontological and virtue ethicist theories identify several features of the forced organ donation case which they argue make administering the injection morally impermissible. These features include properties of the act and of Jane herself. For example, it can be argued that the act necessarily involves breaching the rights of the hitchhiker not to be used as a means to produce favourable consequences,² and that Jane has a duty not to intentionally kill her patients. These features are not of the same moral significance under hedonistic act utilitarianism. For example, the *rights* of the patient are not inviolable under this theory. If, in the case under consideration, violating a right will lead to a greater outcome than not violating that right, then you ought to violate the right.³ There are also no moral duties that require an agent to fail to maximise the overall amount of happiness in the world, such as a duty *never* to kill, under hedonistic act utilitarianism. Nor, for that matter, are Jane's intentions of moral significance under hedonistic act utilitarianism, since the theory assess an agent's *acts* and not her *intentions*.

How, then, does hedonistic act utilitarianism assess the rightness of administering the injection in the forced organ donation case without reference to features such as rights, duties and intentions? As stated above, the rightness of an act under hedonistic act utilitarianism depends on the amount of happiness produced by that act. Suppose we quantify this into 'happiness utils' (hu), such that one year of healthy life will produce, on average, 1 happiness util. Suppose also that the hitchhiker would be expected to have around 50 years of healthy life if he were dismissed from the clinic unharmed, but that his organs could safe four patients who would each be expected to have around 30 years of healthy life. Given this, we can represent the consequences of the two courses of action available to

²This is prohibited by patient-centred deontological theories. See, for example, Kamm (1989).

³It is possible to give a hedonistic act utilitarian defence of working within a human rights framework most of the time. For example, rights offer people a sense of security that may increase their happiness. Rights may also prevent the occurrence of acts, such as genocide, which in general decrease overall happiness.

Jane in the following manner:

	Hitchhiker	Patients
Inject	0hu	120hu
Don't Inject	50hu	1hu

From this it is clear that the injecting produces 120hu, while failing to inject produces 51hu (assuming that the four patients can expect the equivalent of 1 year of life between them if none of them receive an organ). Since the rightness of an act is determined by the amount of happiness it produces, injecting is clearly the better act under hedonistic act utilitarianism, as it is expected to produce more than twice amount of happiness that not injecting will produce. It is not much of a leap from this to saying that Jane *ought* to administer the injection in this case, since this merely requires that Jane ought to do what is morally better over what is morally worse.

But there is something missing in this account of hedonistic act utilitarianism. While we have said that under this account the rightness of an act is determined by the *happiness* produced by this act, it could be argued that the table above takes into account the possible happiness rather than actual happiness. Suppose that if Jane injects the hitchhiker, his organs will in fact be rejected by all four patients and they will all die very shortly after their respective surgeries. We can represent these actual outcomes in the forced organ donation case as follows:

	Hitchhiker	Patients
Inject	0hu	0hu
Don't Inject	50hu	1hu

If hedonistic act utilitarianism requires that we act such that we produce more *actual* happiness rather than such that we produce more *possible* happiness, then not injecting is the better act in this case, since it will produce 51hu while injecting will produce 0hu. The distinction between actual and possible happiness in hedonistic act utilitarianism occurs primarily at the level of value, where either possible happiness or actual happiness is of value.⁴ I will refer to theories that treat actual outcomes as primary at either the level of value or the level of what one ought to do as 'actualist' theories, while theories that treat expected outcomes as primary will be referred to as 'possibilist' theories.

⁴At the level of what an agent *ought* to do the distinction is instead made between *expected* and *actual* consequences. These are both consistent with an actualist or a possibilist account of value. For example, we might say you ought act such that you expect to produce more actual happiness or act such that you actually produce more possible happiness, though the latter is admittedly somewhat strange. I will discuss both the expected (subjective) and actualist (objective) version of 'ought' in sections 1.2.3 and 2.2 below.

There are three noteworthy components of hedonistic act utilitarianism as it has been outlined thus far. The first is that it identifies something that is of final value, namely happiness. Happiness is of final value here because it is not treated as valuable *for* something else, but rather as an end in itself. For example, Jane's *knowledge* of where the insulin is kept is of instrumental value in the forced organ donation case, since she can use that knowledge to perform an act that produces more happiness. However, under hedonistic act utilitarianism this piece of knowledge isn't of any value except insofar as it contributes to the production of future happiness. It is of no final value.

The second thing to note is that under hedonistic act utilitarianism there is a way of identifying which acts are of greater *total* value than others. We are able to see that in the forced organ donation case where the organ recipients don't die, the total value of the world in which Jane administers the injection is greater than the total value of the world in which she fails to do so. This would be false if we couldn't aggregate happiness: for example, if happiness utils were non-additive, or happiness utils for one agent or time were incomparable with happiness utils for another agent or time. It is arguable that the task of identifying which acts are of greater total value is easier in a theory like hedonistic act utilitarianism, since it posits that only one thing - happiness - has final value. This will be less clear in theories that posit more than one final value.⁵

The third noteworthy component of hedonistic act utilitarianism is that it moves from the identification of total value to the stipulation that, in a choice between more than one act, an agent ought to perform the act that produces the most total value. In other words, this component involves the *assigning of deontic properties* to acts on the basis of the total value they produce. I wish to make three clarifications regarding the assigning of deontic properties before continuing. Firstly, it is worth bearing in mind that a theory can assign more than one deontic property to some act ϕ besides 'being such that you ought to do it'. For example, some acts might be permissible without being obligatory. Or it might be the case that the theory in question does not posit not a binary ought-or-ought-not distinction at all. Instead the hedonistic act utilitarian might argue that you ought to ϕ to degree n, depending on the total value of ϕ -ing. Secondly, although I assumed in the above discussion that you ought to perform the act that produces the most happiness, this is only true if the

⁵The possibility of positing multiple things of value will be extremely relevant the ultimate application of this framework to the epistemic consequentialist theory of chapter 2.

hedonistic act utilitarian theory we are considering is a maximising theory. For example, a satisficing hedonistic act utilitarian theory may stipulate that it is permissible for you to undertake an act if it produces more than 10hu, and that there is no difference in the deontic properties of acts that produce happiness utils above this threshold. Under this account, Jane's choice between injecting and not injecting in the first case is completely up to her, since both acts are equally morally permissible. Thirdly, hedonistic utilitarianism attributes normative properties to *acts* alone. I will argue below that it is possible to attribute deontic properties to many things other than acts.

Selim Berker (2010) identifies the three notable components of consequentialist theories that I have described above as its *theory of contributory value*, its *theory of all-thingsconsidered value* and its *deontic theory* respectively. We can see how the three components combine to form hedonistic act utilitarianism by quoting Berker's own example:⁶

Maximizing Hedonistic Act-Utilitarianism:

i. theory of contributory final value:

Pleasurable experiences have *pro tanto* value as ends (where the degree of value is proportional to the intensity and to the duration of the experience).

Painful experiences have *pro tanto* disvalue as ends (where the degree of disvalue is proportional to the intensity and to the duration of the experience).

Nothing else has *pro tanto* value or disvalue as an end.

ii. theory of all-things-considered value (for evaluative focal point {acts}):

If a_1 and a_2 are two acts available to agent S at time t, then a_1 has more all-things-considered value than a_2 iff [the net balance of pleasure over pain that would be brought about if S performs a_1 at t] is greater than [the net balance of pleasure over pain that would be brought about if S performs a_2 at t] (where net balances are determined additively, somehow weighing intensity against duration, and where intensities of pleasures and of pains are assumed to be positive and negative components of a single scale).

iii. deontic theory (for evaluative focal point {acts} and deontic property being right): Act a available to agent S at time t is right iff no other act available to S at t has more all-things-considered value than a.

In the next section I will give a more precise account of each of these three components of consequentialist theories, and show that this framework is able to accurately capture a variety of standard consequentialist theories within ethics. This will allow me to construct

 $^{^{6}}$ Pg. 11-12, S. Berker, 'Epistemic Teleology and the Separateness of Propositions' [*ETSP* hereafter] (*draft* of July 2010). In order to keep this formulation terminologically consistent with the account of hedonistic act utilitarianism given in this section, 'pleasure' should be treated as interchangeable with 'happy/happiness', and 'painful' should be treated as interchangeable with 'unhappy/unhappiness'.

a fully-detailed epistemic consequentialist theory in chapter 2. Such a theory of epistemic consequentialism will fill each of the key variables within the framework just sketched.

1.2 The Three Components of Consequentialist Theories

In this section I will look more closely at the three components of the consequentialist framework described in the previous section. These are a theory of contributory final value (1.2.1), a theory of all-things-considered value (1.2.2), and a deontic theory (1.2.3). In each subsection I will identify the key variables within the component, and outline the key questions that must be answered before we can formulate an epistemic consequentialist version of the relevant component.

1.2.1 Contributory Final Value

Something that was of *contributory final value* in the forced organ donation case was the future happiness of the hitchhiker, which would have been 50hu if Jane administered the injection, and 0hu if she did not. The hitchhiker's happiness was of *final* value in the calculation because his happiness was an end in itself rather than being valuable because it would achieve some further end, such as increasing the number of smiles in the world.⁷ The hitchhiker's happiness was of *contributory* value because it was only given partial weight in our assessment of the overall goodness or badness of the act of injecting the hitchhiker. Other things given partial weight include the future happiness of the patients, which in the first case presented Jane with an overriding consideration in favour of injecting the hitchhiker despite this resulting in the loss of his future happiness. These two properties are present in all of the contributory final values in hedonistic act utilitarianism.

Contributory values are often called 'pro tanto' values. If an outcome is pro tanto valuable, then that outcome is said to be valuable in *a certain respect*. But this can be read in two different ways. On the first reading, the hitchhiker's happiness is pro tanto valuable because it is valuable for a single agent, whereas when we are trying to work out the allthings-considered value of the situation we will take into account the happiness of *all* of the relevant agents, including that of the patients. But the hitchiker's happiness could also be pro tanto valuable because there is a value other than happiness that ought to be taken into consideration. For example, the loss of his future happiness might be of pro tanto ethical

⁷For an account of the final/instrumental distinction as it is employed here, see Korsgaard (1983).

disvalue, while the loss of his future knowledge might be of pro tanto epistemic value. This second reading of pro tanto value is more controversial, however, since it depends on the thesis that there is more than one kind of value.

Value pluralists argue that there are lots of different kinds of values. A painting might be aesthetically valuable, a happy experience might be morally valuable, and a known belief might be epistemically valuable. In contrast with this position, value monists argue that there is only one kind of fundamental value, and that all different values are instrumentally valuable insofar as they promote the fundamental value. Suppose one is a value monist who believes that the only value is ethical value, and that the only thing of ethical value is happy experiences. Instead of asserting that our intuition about the value of art and knowledge is completely wrongheaded, the monist is likely to argue that a painting can have value, but that what we call its 'aesthetic value' not really a final value. Rather, 'aesthetic value' is something that is instrumentally valuable because it promotes happy experiences. The value pluralist, on the other hand, is able to say that there are various different kinds of non-instrumental values, and that the aesthetic value of a painting is not merely valuable because of the ethical value of the happy experiences it causes.

The above examples refer to value pluralism and monism about different *types* of value. However, one can also be a pluralist or a monist *within* a value-type. For example, we can imagine someone who is a monist about the type of value that exists, such as ethical value, but a pluralist about what is of value within the ethical value-type, such as keeping promises, political freedom and the bonds of friendship. We can also imagine someone who is a pluralist about the type of values that exist, such as ethical, aesthetic, and epistemic value, but a monist about what is of value within each value-type, such as happiness alone being of ethical value, beauty alone being of aesthetic value, and truth alone being of epistemic value. Conventionally the term 'value monist' has been used to denote only those who are a monist about both value-types and about which things are of ethical value within a value-type. Since the topic of this thesis straddles both the ethical and the epistemic value-type as 'value-type monism' and 'value-type pluralism', and monism and pluralism about what is of value within a value-type simply as 'value monism' and 'value pluralism'.

Summary

A theory of contributory final value must specify what things are of final, non-instrumental value. However, the things it specifies should be of contributory rather than all-thingsconsidered value. A thing is of *contributory* value when there are other things that also have value (this can either be the *same* kind of value or of a *different* kind of value), which we need to take into account when we assess the rightness or wrongness of the situation taken as a whole. A thing is of *final* value when it is valuable as an end in itself, and not merely because it promotes something else of value. Finally, a theory of contributory final value may specify that there is only one thing that is of value (value monism) or that there are many different things that are of value (value pluralism). Ideally, a theory of contributory final value will be *exhaustive* such that it specifies all the things that are of contributory final value according to the relevant consequentialist theory, but in some cases it may only be possible to offer a non-exhaustive account of contributory final value.

Questions for Epistemic Consequentialism

The primary questions that this discussion of contributory final value raises for an epistemic consequentialist theory of contributory final value are:

(i) What if, when we consider all value-types, none of the things that are of ultimate final value are epistemic values?

- (ii) Is there more than one type of thing that is of final epistemic value?
- (iii) What are the fundamental bearers of value in epistemic consequentialism?
- (iv) What is of contributory final value under epistemic consequentialism?
- (v) Should we care about *actual* or *possible* epistemic value?

To clarify what is meant by question (i), suppose that true beliefs are the only thing of final value within the epistemic value-type, such that all other things in that value-type like having well-evidenced beliefs - are instrumentally epistemically valuable because they promotes true beliefs. But suppose also that true beliefs are themselves instrumentally valuable when we consider all other value-types, with some ethical or prudential value such as 'survival' or 'pleasurable experiences' being the only thing that is of ultimate final value. Does this make the goal of constructing a viable epistemic consequentialist theory futile? I argue that it does not, and for this reason I will remain largely non-committal with regards to question (i) in the remainder of this thesis. One might be sceptical about value or valuetype monist views that exclude anything epistemically valuable from being of ultimate final value, where by 'ultimate final value' I mean that it is of final value within the set of all value-types and not merely within a given value-type.⁸ Let us suppose that this view is correct with regards to epistemic values, and the things which *seem* to be of ultimate final value in epistemology - such as true beliefs - turn out to be only instrumentally valuable. Surely it would still be useful to find out how one ought to go about acquiring these things that are of final epistemic value, even if these turn out to be of purely instrumental when we expand our field of assessment to all value-types. As such, whatever the answer to question (i) may turn out to be, it is not of particular relevance to the present task of constructing a fully-developed epistemic consequentialist theory.

I will overlook question (ii) in what follows for similar reasons. Although there may be more than one thing of epistemic value, it is difficult to give an exhaustive account of contributory final values in a thesis of this size. In my discussion of contributory value below, I will aim to specify at least one epistemic contributory final value and leave open the possibility that there are a plurality of further values within the epistemic value-type.

I will offer an answer to question (iii), (iv) and (v) in section 2.1.1, where I discuss what is epistemically valuable. Fundamental bearers of value, such as the world in which the hitchhiker is injected, should not be confused with evaluative focal points, such as the act of injecting the hitchhiker. Fundamental bearers of value are the *outcomes* that we are trying to prevent or promote.⁹ In the discussion hedonistic act utilitarianism I treated the world in which Jane injects the hitchhiker and the world in which she doesn't injected the hitchhiker as fundamental bearers of value. I then worked out the value of the act of injecting or not injecting the hitchhiker from the value of the world in which Jane injects the hitchhiker and the value of the world in which Jane injects the hitchhiker and the value of the world in which Jane injects the hitchhiker and the value of the world in which Jane injects the hitchhiker and the value of the world in which Jane injects the hitchhiker and the value of the world in which Jane injects the hitchhiker and the value of the world in which Jane does not inject the hitchhiker. In section 2.1.1 I will argue that the most plausible things to fill the role of fundamental bearers of

⁸This is mainly due to intuitions elicited by cases where we are asked to decide, for example, between a world in which infinite happiness (or some other relevant value) has been achieved, but we have no knowledge, and an identical world in which we have a great deal of knowledge. Should we really be indifferent between these worlds? Or should I refuse to labor over an account of pulsars in order to know more about them if this knowledge won't bring a sufficient amount of future happiness into the world? Both theses seem implausible, but there are enough confounding variables in both cases, including our rational uncertainty about what is of final value, to warrant taking these intuitions with a rather large grain of salt.

 $^{^{9}}$ As such, fundamental value-bearers in ethics must, as Berker argues, be the kind of things that we can prevent or promote (pg. 7, Berker (2010)).

value in epistemology are Savage states and outcomes.¹⁰ I will attempt to answer question (iv) in section 2.1.1 by specifying at least one thing that is of epistemic contributory final value. Finally, I will argue in section 2.1.1 that agent's ought to maximise *actual* epistemic value, rather than possible epistemic value.

1.2.2 All-Things-Considered Value

In the first of the forced organ donation case the all-things-considered value of injecting is the total value of the outcomes for both the hitchhiker and the patients (0hu + 120hu = 120hu). In the same case the all-things-considered value of not injecting is also the total value of the the outcomes for both the hitchhiker and the patients (50hu + 1hu = 51hu). This assumes that there are no further results of injecting or not injecting, however distant or unexpected, that have a non-zero value under hedonistic act utiltarianism. For example, imagine Jane knows that killing the hitchhiker - even doing so is the right thing to do would disturb her deeply and thereby reduce her expected happiness by 0.5 happiness utils per year (from 1hu to 0.5hu) in each of the ten years after she killed him, whereas not killing him would have no effect on her expected future happiness (a net loss of 5hu compared to the life she would have had). Suppose Jane can expect to have another 40 years of healthy life in any case. We can take this into account by presenting the case as follows:¹¹

	Hitchhiker	Patients	Jane	TOTAL
Inject	0hu	120hu	35hu	155hu
Don't Inject	50hu	1hu	40hu	91hu

It is irrelevant whether Jane is rational to be afflicted by the death of the hitchhiker. The lowering of her happiness that it would cause her must taken into account simply because it *will* occur (if we are actualists), or because it is *expected* to occur (if we are possibilists). For the actualist, the same is true of any very improbable, unforeseen result of her killing

¹¹Although I have used the *total* happiness utils gained by each party in these table, I could just have easily have used the total *difference* in their happiness compared would have been the case had injecting occurred or not occurred. Here is what such a counterfactual version of this table would look like, where I treat *not injecting* as the status quo (such that 'Don't Inject' could be left out of the table entirely):

	Hitchhiker	Patients	Jane	TOTAL
Inject	-50hu	119hu	-5hu	64hu
Don't Inject	0hu	0hu	0hu	0hu

¹⁰The framework I discuss is structurally analogous to a modal account, in which the set of *states* are replaced with the set of possible worlds and *outcomes* are replaced the set of possible worlds at which the relevant act occurs. Both states and outcomes can be fundamental bearers of value in Savage's framework, but the value of an *act* is assessed relative to the value of its *outcomes*. See pg. 48-9, Joyce (1999).

the hitchhiker that actually occurs.¹² Suppose that, unbeknownst to Jane, terrorists have implanted the needle she will open to inject the hitchhiker with a deadly pathogen that will eradicate the human race. Under actualism, this outcome ought to be included in our assessment of the overall value of the act of injecting the hitchhiker.

Therefore, in the theories we are considering, the all-things-considered value of an act is the aggregate of all of the pro tanto values and disvalues produced by that act.¹³ We can either give an *unrestricted* all-things-considered value judgment by taking into account all of the pro tanto values from all of the different value-types (moral, epistemic, aesthetic and so on), or we can give a *restricted* all-things-considered value judgment by taking into account only pro tanto values from one value-type.¹⁴ For a value pluralist to do the former in the forced organ donation case would require that he or she take into account not only the net loss or gain of happiness produced by injecting or not injecting the hitchhiker, but also the net loss or gain of all of the pro tanto epistemic, aesthetic and gastronomic values (to name just a few!) produced by injecting or not injecting the hitchhiker. In this thesis I will not be very concerned with unrestricted all-things-considered value judgments, since I will focus on all-things-considered value within the epistemic value-type.

Each of the consequentialist theories considered thus far reaches its account of all-thingsconsidered value by adding up various pro tanto values. One worry that can be raised for theories which use this method is that it is impossible for them to offer an adequate theory of all-things-considered value if some of the pro tanto values they postulate are *incommensurable*. Two kinds of value are incommensurable if there is no single scale against which they can both be measured. If there is no common measurement between, for example, the goodness of knowledge with the goodness of pleasurable experiences, then we cannot add these two kinds of value together to get the all-things-considered value of an act. Such worries need not only affect value pluralists however, since they can occur within monistic theories too. For example, suppose that the hitchhiker's happiness is incommensurable with the patients' happiness because there is no agent-free 'happiness' scale against which we

¹²Under possibilism we should already have minute credences that these outcomes will take place. I focus on the actualist account here, since I will reject the possibilist account of values and oughts in section 2.2.

 $^{^{13}}$ It should be borne in mind that the relationship between an act and the value it produces need not merely be causal. A belief in a falsehood might be bad both because it *causes* further false beliefs, but also because that very false belief *constitutively* produces epistemic disvalue.

¹⁴All-things-considered value judgments can also come in various degrees between these two points. For example, we might consider more than pro tanto value from more than one value-type, without taking into account absolutely all pro tanto values!

can measure them. How are we to give an all-things-considered value of the Jane's injecting or not injecting the hitchhiker? At best we could say that one outcome is better *for* the hitchhiker, and one onecome is better *for* the patients, but neither of these is sufficient to give us the all-things-considered value of the act of injecting or not injecting.

However, even if incommensurability about values is true, we may still be able construct an *ordinal* ranking of different values.¹⁵ Suppose that knowledge and pleasurable experiences are incommensurable pro tanto values. We are still able to say that 'gaining a large amount of knowledge is better than gaining minuscule amount of pleasurable experience', we just cannot say *how much* better, since we can't measure the two values on a common scale. The consequentialist who accepts incommensurability is therefore unable to add up the pro tanto values of the world in which Jane injects and the world in which she doesn't inject in order to tell us how much better the all-things-considered value of injecting is over the all-things-considered value of not injecting. However, she can still give us an ordinal ranking of the all-things-considered value of different acts, e.g. 'the act of injecting has more all-things-considered value than the failure to inject'.¹⁶ Perhaps this ordinal account is sufficient for a theory of all-things-considered value, since we can still make recommendations for which act an agent should perform.

Incommensurability theses about value should not be confused with their much more problematic cousins: *incomparability* theses about values. Two values A and B are incomparable if comparative statements about A and B, such as 'x is better than y', are never true. Unlike value incommensurability, value incomparability prevents us from constructing an ordinal ranking of different values because we have lost the ability to make 'better than' and 'worse than' judgments about those different values.¹⁷ As such, value incomparability

 $^{^{15}}$ Ordinal rankings are those which rank things in terms of 1st, 2nd, 3rd. It is usually distinguished from cardinal measurements, which assigns an absolute number to each of the entities being ranked, so that we can tell how much better one is than another. For example, suppose we want to rank candidates by the number of votes they receive. An ordinal ranking can tell us that candidate A came first, candidate B came second and candidate C came third. A cardinal measurement can tell us that candidate A got 100 votes, candidate B got 15 votes and candidate C got 1 vote.

¹⁶How can she do this? Suppose we know the ordinal ranking of each of the pro tanto values of the different possible acts: inject and don't inject. This allows us use an appropriate voting procedure such as the minimax method to determine the act that is best when we take into account *all* of its pro tanto values. The resultant act is therefore the one that is the best in terms of its all-things considered value. Of course we may want to determine an ordinal ranking and not merely a winner, but I will not attempt work out the details of such a theory here.

¹⁷See Chang (Ed.), 1997. Note that we can still appeal to dominance principles even if two values are incommensurable and incomparable. Suppose that A and B both have non-negative value and that the values of A and B are mutually independent. Then we are still in a position to know that a world in which there are 5 units of x and 4 units of y will better, ceteris paribus, than a world in which there are 3 units of x and 2 units of y, even if we don't know how to compare x with y.

presents a problem for both cardinal *and* ordinal accounts of all-things-considered value. If we have reason to think that epistemic values might turn out to be incomparable, then this would present a much more serious problem for an epistemic consequentialist account of all-things-considered value than would the possibility of their being incommensurable.

The final issue we must address regarding all-things-considered value is the nature of the *evaluative focal points* that have all-things-considered value. Because I have been discussing the forced organ donation cases, I have been assuming that the relevant evaluative focal point is that of *acts*, of which both injecting and not injecting are members.¹⁸ But what are evaluative focal points? The term itself was first proposed by Kagan:¹⁹

'Acts and rules are examples of what I will call *evaluative focal points* (other examples include motives, norms, character traits, decision procedures and institutions). Act consequentialists and rule consequentialists share the foundational consequentialist thought that justification must be in terms of the good, but they differ in their choice of which evaluative focal points to make primary.'

Jane's act of injecting the hitchhiker is a member of the evaluative focal point {acts}, while her intention to help save the other patients is a member of the evaluative focal point {intentions}. Suppose Jane injects the hitchhiker but she in fact does so because she has a secret hatred for people who hitchhike and wants to make them suffer. Although the value of her *action* will remain the same (assuming the organs are all distributed accordingly), the value of her *intention* would, ceteris paribus, be lower under hedonistic act utilitarianism. Why? Because the intention to cause suffering to hitchhikers will probably result in fewer pleasurable experiences in the future than would result from the intention to save patients.

There are several reasons we might want to move to a language of evaluative focal points over the traditional adoption of a single evaluative focal point. Firstly, it provides us with a simple means of explaining *indirect* consequentialist theories, which are theories that do not treat *acts* as their primary evaluative focal point. For example, suppose we were to take *decision procedures* as our primary evaluative focal point, and that we arrive at an all-things-considered value of the decision procedure 'always update on your evidence', and of the decision procedure 'always inversely update on your evidence'. A decision procedure consequentialist assesses the goodness of a cognitive act in terms of how well it 'fits' with

¹⁸One might worry that *not injecting* is technically refraining from an act. But for present purposes we can simply specify *not injecting* as set of actions Jane takes in the nearest world in which she doesn't inject. ¹⁹Pg. 134, Kagan (2000).

the decision procedure with greater all-things-considered value. His assessment of the acts is therefore not arrived at by directly assessing their all-things-considered value.

The second advantage of appealing to evaluative focal points is that we can use this framework to apply our consequentialist theory to *everything*, including acts, intentions and decision procedures. By adopting such a view, we can assess not only the all-thingsconsidered value of Jane's act of injecting the hitchhiker, but also the all-things-considered value of her intentions, character traits and decision procedures. We can do so without treating any of these evaluative focal points as secondary to one of the others, as was the case in the above formulation of decision procedure consequentialism. This might explain why we are less willing to blame Jane in the case in which all of the patients reject their organs, and why we are less willing to praise her in the case in which she just wants to cause the hitchhiker suffering. In the former, Jane has acted in accordance with intentions and decision procedures that usually result in greater all-things-considered value, even though the act itself produces a low amount of all-things-considered value. In the latter, Jane's act produces a high amount of all-things considered value, but she acted in accordance with intentions that usually result in lower all-things-considered value.

Theories that apply consequentialism to all evaluative focal points are called *global* consequentialist theories.²⁰ Within global consequentialist theories it is usually more useful to compare the all-things-considered value of a member of an evaluative focal point, such as the *act* of injecting, with other members of the same evaluative focal point, such as the *act* of not injecting. However, there is nothing obviously fallacious about evaluating the all-things-considered value of members from different evaluative focal points, such as comparing the all-things-considered value of the act of not injecting with the all-things-considered value of the intention to save as many patients as possible.²¹ An obvious concern regarding global consequentialist theories is what they will say when the outcomes recommended by two different evaluative focal points conflict. I will return to this worry in section 2.1.3 below.

 $^{^{20}}$ See Ord (2009).

²¹Perhaps we don't often do this because we are often making value comparisons between two things that are mutually incompatible, since we have to choose to perform one or the other. And presumably it is more common to find mutually incompatible members of the *same* evaluative focal point, than it is to find mutually incompatible members of *different* evaluative focal points.

Summary

A theory of all-things-considered value must take us from the *pro tanto* values resulting from different acts (as specified by our theory of contributory value) to the *total* value of those acts. The simplest way of doing this is by adding up all of the pro tanto values produced by each act. If the pro tanto values are all commensurable and comparable then we can give a cardinal measure of the all-things-considered value of each of the acts. However, if any of the pro tanto values are *incommensurable*, then we can only give an ordinal ranking of the all-things-considered value of each act. And if any of the pro tanto values are *incommensurable*, then we cannot give an ordinal or a cardinal ranking of the all-things-considered value of each act. We can extend our account of all-things-considered value to things other than acts by adopting a global consequentialist view, which holds that we can assess the all-things-considered value of any *evaluative focal point*. Common evaluative focal points include the set of acts, intentions and decision procedures. A specific act, intention or decision procedure is a *member* of each of these evaluative focal points, respectively.

A Worked Example

It will be useful to show that the framework presented here can adequately capture consequentialist theories that differs greatly from hedonistic act utilitarianism. For example, a theory that differs with regards to (i) *what* it considers to be of pro tanto value, (ii) how it *aggregates* those pro tanto values in its theory of all-things-considered value, and (iii) what *evaluative focal point* it treats as primary. One of the more well-known examples of a theory that fits these three criteria is Hooker's *rule consequentialism*.

Hooker's theory of contributory final value employs a welfarist account of pro tanto value, such that it is *well-being* that is of final value.²² Hooker understands an increase in well-being to roughly mean the fulfilment of an objective list of rational desires, rather than merely the 'pleasurable experiences' of hedonistic act utilitarianism.²³ In addition, Hooker does not treat the aggregation of well-being in the theory of all-things-considered value to be straightforwardly additive. Instead his theory of what I am calling all-things-considered value is a *prioritarian* one. Prioritarian views attempt to weight

 $^{^{22}}$ Though this is unclear. It is certainly more intuitive to assess weighted well-being at the level of aggregation, but this may not be what Hooker has in mind.

²³See pg. 37-43, Hooker (2000).

increases in well-being among those agents that are worst off more highly than increases in well-being among those whose levels of well-being are already comparatively high.²⁴ Hooker summarises this view as follows:

Aggregate well-being combined with some priority for the worst off can be expressed as a *weighted sum of well-being*. This brings the priority towards the worst off in the calculation of the sum of well-being.

One reason to adopt a prioritarian view of the aggregation of well-being is that it is better at capturing our more egalitarian intuitions without asking us to 'level down' those who are better off.²⁵ For example, suppose there exists one person who has £100 per day to live off of and five people who each have £1 per day to live off of, and that this has been adjusted for purchasing power parity. If we were to add some fundamental importance to equality, then surely we ought to prefer the removal of £99 per day from the rich person to continuing with the status quo. But this is absurd, since the result is having one additional poor person in the world! On the other hand, if we find ourselves able to distribute a further (PPP adjusted) £100 among any number of people, we surely ought not to be indifferent between giving the rich person an additional £100 per day to live off of, or giving the poor people an additional £20 per day to live off of. The prioritarian attempts to resolve these two intuitions by arguing that instead of maximising the well-being of individuals. Put more simply, well-being is of diminishing marginal value.²⁶

Finally, the evaluative focal point treated as primary under Hooker's consequentialist theory is not the set of *acts*, but the set of *rules*. Rules are characterised as codes of behavior that are *internalised* by agents, and the best rules are those that maximise the weighted sum of well-being. The value of members of other evaluative focal points, such as acts and character traits, can be assessed relative to whether or not they would result from the internalisation of the set of optimal rules.²⁷

I have now given a brief account of the pro tanto values, method of aggregation, and primary evaluative focal point under rule consequentialism. These three components can

 $^{^{24}}$ Notably this means that Hooker's view is not technically a utilitarian one, where utilitarian theories are restricted to those that adopt non-prioritarian welfarism.

 $^{^{25}}$ See Parfit (1997).

 $^{^{26}}$ Providing an adequate account of the function being employed by prioritarians is both beyond the scope of this thesis and unnecessary for present purposes.

 $^{^{27}}$ I will discuss the distinction between *internalising* a rule and *complying* with a rule in section 2.1.3. For a discussion of the interaction between rules and acts, see pg. 93-99, Hooker (2000).

be used to construct a rule-consequentialist theory of all-things-considered value as follows:

Hooker's Rule Consequentialism:

ii. theory of all-things-considered value (for evaluative focal point {rules}): If r_1 and r_2 are two rules that can be internalised by agent S at time t, then r_1 has more all-things-considered value than r_2 iff [the net balance of weighted wellbeing that would be brought about if S internalises r_1 at t] is greater than [the net balance of weighted well-being that would be brought about if S internalises r_2 at t] (where the net weighted well-being is determined additively).

This demonstrates that the framework outlined here is able to encapsulate the all-thingsconsidered value of several different consequentialist theories. These range from hedonistic act utilitarianism, which sums the pleasurable experiences produced by acts, to rule consequentialism, which aggregates the weighted well-being produced by rules. I will now consider the questions this raises for an epistemic consequentialist theory.

Questions for Epistemic Consequentialism

The primary questions that this discussion of all-things-considered value raises for an epistemic consequentialist theory of all-things-considered value are:

- (i) Are any contributory epistemic values incommensurable or incomparable?
- (ii) What are the evaluative focal points assessed by epistemic consequentialism?
- (iii) How do we work out all-things-considered value under epistemic consequentialism?

As I've emphasised above, worries about question (i) affect all consequentialist theories of value and not epistemic consequentialism in particular. However, I will avoid getting bogged down in incommensurability and incomparability issues by stipulating only one final epistemic value in section 2.1.1. This removes the worry that multiple epistemic values will turn out to be incommensurable or incomparable. There may still be the worry that inter-agential comparisons are incommensurable or, worse, incomparable. But I see no independent reason to suspect this will be more likely to occur in the epistemic case than in the ethical case, and so I will shelve these worries for the remainder of this thesis.

Answering question (ii) will be the first aim of section 2.1.2. In this section I will argue that epistemic consequentialism is best interpreted as a form of global consequentialism. This stands in contrast to those who believe that an epistemic normative theory can only assess certain focal points such as credences and beliefs. Epistemic consequentialism can, I argue, assess any evaluative focal point. This includes not only credences and beliefs, but also decision procedures, intentions, character traits and major food groups. The aforementioned focal points have simply been the main concern of epistemologists, and not without reason. But we are also able assess the epistemic value of an *act* of going to the library instead of staying in bed, or an *intention* to go to university, or the *character trait* of conscientiousness. In sum, the evaluative focal points assessed by epistemic consequentialism need not be restricted to the set of cognitive acts. During this discussion will also consider the possibility of placing an *ought implies can* restriction on what counts as an 'available' member of an evaluative focal point.

Finally, answering question (iii) will the second aim of section 2.1.2. I will propose that the all-things-considered value of a member of an evaluative focal point under epistemic consequentialism, is the summation of the epistemic contributory final values produced by that member.²⁸ I will argue that an epistemic consequentialist theory of all-thingsconsidered value ought to sum *all* of the contributory final values resulting from the relevant act, intention or decision procedure being assessed. Some of the epistemic contributory final values produced by the adoption of that belief may also be very unexpected, or they may occur at a very distant time and place. For simplicity, however, I will often assume that there are no distant effects brought about by the thing we are considering.

It is worth bearing in mind that the sum of all of the contributory values resulting from an act, intention or decision procedure will include both the *constitutive* and *causal* contributory final values produced by that act, intention or decision procedure. Suppose we were to argue that true beliefs are of contributory final value. When assessing the allthings-considered value a given true belief, we would need to sum not only the value of *having* that belief (its constitutive value) but also the value of the various true and false beliefs that result from having the true belief in question (its causal value).

1.2.3 A Deontic Theory

Once it was discovered in the first of the forced organ donation cases that injecting the hitchhiker produced the greatest number of happiness utils, it was argued that Jane *ought* to inject the hitchhiker because she ought to perform the act that is right. But in order to move from the all-things-considered value of the act of injecting to an assertion that

 $^{^{28}}$ Such an account is sufficient for present purposes, since I'm attempting to give an account of only one contributory final value within one value-type. Constructing a more exhaustive normative theory might require giving different weight to different contributory values even *within* the epistemic value-type.

Jane ought to perform that act, we have to have some way of assigning *deontic properties*, such as rightness and wrongness, to the act in question. This is the role of a deontic theory. A deontic theory gives us a function from the all-things-considered value of an act to the deontic properties of that act.²⁹ In the forced organ donation case this function assigned the deontic property of 'being right' to the act available to Jane had the greatest all-things-considered value; that is, to the act of injecting the hitchhiker.

In this thesis I will focus solely on the deontic properties of 'being right' and 'being wrong' and will go on to look at how these properties relate to what agents ought to do.³⁰ But there are two main ways that one can model these deontic properties. The first is what I will call the *outright* account of rightness and wrongness, and the second is what I will call the *gradational* account of rightness and wrongness.

Under the outright account of rightness and wrongness, an act is either *right*, *wrong*, or *neither right nor wrong* (assuming there exist such 'deontic gaps'). Which of the three deontic properties an act has depends on the all-things-considered value of that act. Let us take three acts, x, y and z. Suppose we are assessing only the total amount of *happiness* produced by each of these acts. We discover that the all-things-considered value of x is 8hu, the all-things-considered value of y is 0hu and the all-things-considered value of z is -12hu. I will assume that we are using a scale in which acts that have a *positive* all-thingsconsidered value are right, acts with *zero* all-things-considered value are neither right nor wrong, and acts that have a *negative* all-things-considered value are wrong. Given this, we can sort each of the acts x, y and z by their deontic properties as follows:

Right	Neither Right Nor Wrong	Wrong
(act x)	$(act \ y)$	(act z)

But there is a problem with the outright account of rightness and wrongness. If our pro tanto values (in this case 'happiness utils') are commensurable and comparable, then we can say not only that x is right, we can give an account of *how much more* right it is than other 'right' acts, such as those that only produce 4hu. It is possible for us to do this because we can rank the all-things-considered value of different acts on a cardinal scale and use this to

²⁹For simplicity, I will often assume that the evaluative focal points of our deontic theory (e.g. 'acts') correspond to the evaluative focal points of our all-things-considered theory of value.

³⁰Other than being right or being wrong, the most common foundational deontic properties that are assigned to acts are *being obligatory, being optional*, and *being impermissible*. I will not explore these deontic properties here, since I do not believe they are necessary for an adequate epistemic consequentialist theory. Furthermore, these deontic properties could easily be grafted on to the account of being right and being wrong outlined here. For a survey of these deontic properties, see McNamara (2010).

make comparisons between them. But this subtlety is lost at the level of assessing rightness and wrongness if we simply group all acts of positive, zero or negative all-things-considered value together. Should we not be dissatisfied with accounts of rightness and wrongness that fail to reflect the different degrees of all-things-considered value that an act can have?

A gradational account of rightness and wrongness is one which denies that rightness and wrongness are all-or-nothing concepts, as was assumed by the outright account above. Instead, an act is right or wrong to degree n (where $n \in \mathbb{R}$). If we use $W_P(n)$ to mean Pis wrong to degree n and $R_P(n)$ to mean P is right to degree n then we can stipulate that $R_P(n) = W_P(-n)$. This means that if something is right to degree 1 then it is wrong to degree -1, and if something is wrong to degree 0 then it is also right to degree 0. Under the consequentialist theories considered here, the degree to which an act is right or wrong can be directly correlated with the all-things-considered value it produces. This means that if the all-things-considered value of x is 8hu then $R_x(8)$ and $W_x(-8)$, and if the allthings-considered value of z is -12 then $R_z(-12)$ and $W_z(12)$. This gradational account of rightness and wrongness can therefore be represented as follows:

←Right -			Wrong→
	$R_x(8), W_x(-8)$	$R_y(0), W_y(0)$	$R_z(-12), W_z(12)$

Notice that there will be no upper or lower bound on the values that R and W can take *if* there is no upper or lower bound to the amount of all-things-considered value that members of evaluative focal points can have. For example, suppose that heaven is possible, and that in heaven there are an infinite number of agents experiencing an infinite amount of happiness at each instance, and that they do so for eternity. So for each agent at each instance, heaven produces ω -many happiness utils. This means that over an infinite number of agents, heaven produces $\omega \cdot \omega$ -many happiness utils for each agent. So for the infinite number of agents, heaven produces $\omega \cdot \omega \cdot \omega$ -many happiness utils! Given cases like this, it is difficult to stipulate an upper or lower bound on the possible amount of all-things-considered value something can have without being prey to counterexamples. As such, we should assume that possible degrees of rightness may also be unbounded in both directions. I have now given two different accounts of the deontic properties of *being right* and *being wrong*: an outright account and a gradational account. Before moving on, however, we need to show how these deontic properties relate to the question of what an agent *ought* to do.

Within consequentialist theories, it is often argued that what an agent ought to do is

the act available to them that will produce the most all-things-considered value. However, this doesn't mean that all acts which aren't the act we 'ought to do are equally right or wrong. Imagine that Adam has £10 and is deciding what to do with it: either he can get a leg wax, or he can give the money to a charity that builds swimming pools for veterans, or he can give the money to a charity that treats neglected tropical diseases. It seems likely that the leg wax will cause Adam a great deal of pain, so suppose it will produce a total of -0.02hu. Building swimming pools for veterans will certainly cause some happiness, but it is costly. So suppose that the total output of Adam's £10 will be 0.01hu. Giving to the charity that treats neglected tropical diseases will cause happiness and is very cost-effective. So suppose that, when put to this cause, Adam's £10 will produce around 3hu. Clearly Adam *ought* to give his money to the charity that treats neglected tropical diseases. But this doesn't mean that giving £10 to the charity that builds swimming pools for veterans and spending it on a leg wax are *equally wrong*, even though neither of these acts is what Adam ought to do. Indeed, if these were the only two acts available to Adam, then it's clear that he *ought* to do the former and not the latter.

Which acts count as 'available' to an agent, and why do we restrict ourselves to this set of acts when determining what an agent ought to do? In the above case, all three acts were assumed to be available to Adam. But why don't we say that what Adam *ought* to do is give £100,000 to the charity that treats neglected tropical diseases, regardless of the fact that he only has £10? The answer is that when we speak of 'available' acts, we assume that there is a fundamental principle of 'ought implies can', and that the acts that are available to an agent are the acts that the agent can perform. Under this view, it is never the case that someone ought to perform an act that they are unable to perform.³¹ But it is unclear what is meant by 'can' in these accounts. It needs to be something stronger than 'If 'Adam ϕ -ing' is within the realm of logical or metaphysical possibility, then Adam can ϕ '. If 'can' is meant in this sense, then not only can Adam give £100,000 to the charity treating neglected tropical diseases, he can also fly. But it also needs to be something weaker than 'If it is true that [were Adam to intend to ϕ then Adam would ϕ with objective probability 1] then Adam can ϕ '. If 'can' is meant in this sense, then not only is it untrue that Adam can donate £10, he also cannot walk, talk or do very much at all.

 $^{^{31}}$ This is, of course, one among many interpretations of 'ought implies can'. For an overview, see pg. 1265-6, Becker & Becker (2001).

Giving an intuitive account of 'can' in these cases is far from easy, but it is not even clear that we need to rely on there being some fundamental 'ought implies can' principle to begin with. Perhaps we need not restrict our deontic theory to members of an evaluative focal point that an agent can perform. Let us assume that actualism, the thesis that what agent's ought to maximise is *actual* all-things-considered value, is true, and that there are various facts about the world and about the capabilities of the agent. These can be used to work out what actual value will or has resulted, over a series of cases, from an agent *S* doing something, *trying* to do something, *internalising* some decision procedure, or *cultivating* some character trait. How do we work out the all-things-considered value of S's actualising each of these relevant things? A highly plausible answer is that we calculate the sum of the pro tanto values that are actually produced by each thing.

Now suppose that Jones has always enjoyed having fun experiences, and that there is no additional benefit to Jones internalising the decision procedure 'always try to have fun' other than the fun experiences it results in. But Jones discovers that when he internalises the decision procedure 'always try to have fun', he starts to feel a great deal of pressure to constantly have fun, he begins to consciously check whether he's having fun or not, and is constantly worried that he is not trying hard enough to have fun. Suddenly he finds himself having fewer fun experiences than he did before he internalised the decision procedure.

In this case, having fun experiences produces much more all-things-considered value than internalising than the decision procedure 'always try to have fun'. So, given how we calculate all-things-considered value, it may be the case that Jones *ought* to have fun experiences but he *ought not* to internalise the decision procedure. This is because Jones' having fun experiences produces a great deal of overall value, and internalising the decision procedure reduces the number of fun experiences Jones will have. Similarly, in the case of Adam, he *ought* to donate £100,000 to the charity that treats neglected tropical diseases, since this would produce the most all-things-considered value. But he would presumably produce more more all-things-considered value by *trying* to give £100,000 to the charity that treats neglected tropical diseases than by *trying* to give £100,000 to the charity that treats neglected tropical diseases. After all, the former will produce no actual value (because Adam will not successfully perform the act), whereas the latter will produce 3hu.

This means that in cases where it very difficult for an agent to do A (such as give $\pounds 100,000$) but easy for her to do B (such as give $\pounds 10$) then, even though the rewards of

successfully doing A would be greater than the rewards of successfully doing B, she still ought to try to do B instead of trying to do A. This is the kind of result we wanted from an 'ought implies can' principle, since the action an agent ought to perform will, in most cases, be one that they are more likely to achieve successfully. As such, we don't need to appeal to an 'availability' condition when determining what an agent ought to do, since this falls naturally from global consequentialism and a plausible theory of overall value.³²

What about cases where an agent *ought* to perform an act, but it seems to her that she ought not to perform that act. For example, suppose that the day before she was left alone in the clinic, Jane was considering which drug to give a patient. The patient displayed several symptoms, which were all consistent with two diseases: one extremely common and the other extremely rare. She knows that Drug A will treat the common disease, giving the patient 30 more years of life, while Drug B will treat the rare disease, also giving the patient 30 more years of life. If given together, Drug A and Drug B would create a deadly cocktail that would certainly kill the patient. Sadly, both conditions are soon fatal if left untreated. We can represent these options and outcomes as follows:

	Rare Disease	Common Disease
Drug A	0 years	30 years
Drug B	30 years	0 years

Suppose that Jane knows that in 99/100 cases a patient showing these symptoms has the rare disease. What should Jane do? It might seem obvious that she ought to prescribe Drug A, since the *expected* value of doing so, given the information Jane has, is higher than the expected value of not doing so. But suppose that, as it turns out, this patient has the rare disease. Given this, there is a sense in which Jane *ought* to prescribe Drug B, since the *actual* value of doing so in this case is greater than the actual value of prescribing Drug A. The former kind of 'ought' is the 'subjective ought', while the latter kind of ought is the 'objective ought'. There is a great deal of disagreement about which of these 'oughts' should be treated as primary. A deontic theory takes us from the all-things-considered value of an act to claiming that an agent *ought* to perform that act. We must be explicit, however, about which kind of ought is being referring to in our deontic theory.

Finally, it is worth noting that a deontic theory presents us with a criterion of rightness

 $^{^{32}}$ Isn't there something paradoxical in saying that an agent ought to do A and not B but they ought to try to do B and not try to do A? I will provide a response to this objection in section 2.1.3.

for assessing members of evaluative focal points. This criterion may be very different from the *decision procedure* that an agent ought to adopt. For example, imagine there is an agent, Dr. Goodcalc, who always tries to maximise happiness utils. A natural way to imagine she might do this is by calculating the happiness utils that will be produces by all of her actions. Luckily, Dr. Goodcalc - true to her name - is able to calculate the happiness utils that will be produced by each of her actions perfectly. However, Dr. Goodcalc knows that it takes her exactly 9 seconds to perform a utility calculation on any finite set of data. One day, Dr. Goodcalc is walking past a shallow pond, when she notices that a small child is drowning in the water. She therefore begins to assess the data relevant to the utility calculation, which includes the fact that the child will certainly be dead in 8 seconds if she fails to intervene prior to that time. Upon realising this, what should Dr. Goodcalc do? It should be clear to anyone that Dr. Goodcalc will not maximise the number of actual happiness utils in the world by continuing with her utility calculation. Instead, she ought to intervene and save the drowning child, since this will produce more happiness utils than letting the child die in order to complete the calculation. There is nothing paradoxical in this situation. A deontic theory provides us with a means of assessing the rightness of things in terms of how much value they produce. It should *not* be confused with the decision procedure that fallible agents ought to employ in order to produce the most value.³³

Summary

A deontic theory is a function from the all-things-considered value produced by the member of an evaluative focal point, to an assignment of deontic properties to that member. The two deontic properties I have focussed on are those of *being right* and *being wrong*. Under an *outright* account of rightness and wrongness, an act is either right, wrong or neither right nor wrong. Which of the three deontic properties it has is determined by the amount of all-things-considered value it produces. Under a *gradational* account of rightness and wrongness, an act is right or wrong to a certain degree, and the degree of rightness and wrongness is determined by the amount of all-things-considered value the act produces.

 $^{^{33}}$ As I argued in section 1.2.2, under global consequentialism decision procedures can, themselves, be assessed for their degree of rightness. It is far from clear that the consequentialist criterion of rightness would do well under such an assessment. But the rejection of 'maximise epistemic utility' as the decision procedure agents ought to always commit to and the assessment of decision procedures for their *causal* value as well as their *constitutive* value, also offers a solution another species of problems: epistemic bribery cases (Greaves (2011 *draft*)). However, exploring these problems would require a diversion into how causal and evidential decision theory both relate to objective epistemic consequentialism, which is beyond the scope of the present thesis.

Once the rightness of several members of an evaluative focal point has been determined, we need to specify which of those members an agent ought to actualise. I have argued that if the agent ought to actualise the act that produces the most overall value, then we need not constrain ourselves to the set of acts that an agent can perform, by which I mean the set of acts she can actualise consistently. We must also be careful to distinguish between the *objective* and *subjective* ought in our formulation of a deontic theory. The subjective ought is relative to the information a subject has, whereas the objective ought is absolute and non-relative. Finally, we should be careful not to confuse our deontic theory, which is a *criterion of rightness*, with the *decision procedure* an agent ought to internalise.

A Worked Example

Alistair Norscross has proposed a version of *scalar consequentialism* that has many of the features of the gradational account of rightness given above. However, Norcross argues that utilitarians should assess acts in terms of their *goodness*: that is, by the utility they produce. He rejects the view that act is right insofar as it is good, and wrong insofar as it is bad. The first reason Norcross offers for this view is relevant to the above discussion:³⁴

"...it seems to collapse the concepts of right and wrong into those of good and bad respectively, and hence, to make the former pair redundant."

A further difference between Norcross' theory and the account sketched above is that he rejects the move from assigning deontic property of gradational goodness to an act, to the view that an agent *ought* to perform that act. Maximising consequentialism, he argues, classifies only those acts that maximise utility as 'right' and all other acts as wrong, such that the agent ought to do whatever act maximises utility.³⁵ But scalar consequentialism only provides a framework for assessing the goodness or badness of an act, without stipulating that an agent ought to undertake the act that produces the most good. Scalar consequentialism thereby avoids the *demandingness* objections often levelled at maximising

 $^{^{34}}$ Pg. 223, Norcross (2006). I find this thought compelling, since the virtue of simplicity is lost to a theory that identifies rightness with goodness while attempting to retain both. However, rightness and wrongness play a non-identical role from goodness and badness in most theories, in that rightness, unlike goodness, is used to determine what act an agent *ought* to undertake. I will therefore retain this distinction in the chapters that follow. If the reader is happy to derive the action that an agent ought to perform from the goodness of that act instead of rightness of that act, then she can assume that rightness and goodness are being used synonymously within the gradational account of rightness I adopt in the chapters that follow.

³⁵Pg. 216, Norcross (2006).

theories, though it arguably gives up much of the benefits of action-guidance that result from maximising theories of what an agent ought to $do.^{36}$

The basic premise of scalar consequentialism is that an action's goodness or badness depends entirely on the amount of value or disvalue that the act produces, and they are therefore not binary concepts.³⁷ The deontic theory of Norcross' scalar consequentialism can by captured by the framework employed in this chapter as follows:

Norcross' Scalar Consequentialism:

- iii. deontic theory (for evaluative focal point {acts} and deontic property goodness): Act a undertaken by agent S at time t is good to degree n, where n is the all-things-considered utility produced by a and n is positive.
 - deontic theory (for evaluative focal point {acts} and deontic property badness): Act a undertaken by agent S at time t is bad to degree n, where n is the all-things-considered utility produced by a, and n is negative.

This demonstrates that the framework is able to capture the deontic theories of both standard and non-standard consequentialist theories, such as hedonistic act utilitarianism and scalar consequentialism. I will now consider what questions this discussion of deontic theories and their relation to oughts present for an epistemic consequentialist theory.

Questions for Epistemic Consequentialism

The primary questions that this discussion of deontic theories raises for an epistemic consequentialist theory of all-things-considered value are:

- (i) Which *deontic properties* will an epistemic consequentialist deontic theory assign to members of evaluative focal points on the basis of their all-things-considered value?
- (ii) How does the epistemic ought relate to different deontic properties?
- (iii) Is the epistemic ought objective or subjective, or are there both objective *and* subjective epistemic oughts?
- (iv) Will the *decision procedure* recommended by epistemic consequentialism be identical to the *criterion of rightness* given by its deontic theory?

In response to question (i), in section 2.1.3 I will focus on the deontic properties of *rightness* and *wrongness*, where these are determined by the all-things-considered value of the relevant member of an evaluative focal point. Since the all-things-considered value of different

³⁶Pg. 218-220; 228-230, Norcross (2006).

³⁷Pg. 220-3, Norcross (2006).

members of evaluative focal points come in degrees, and we want a way of distinguishing different grades of rightness and wrongness, I will adopt a gradational account of rightness. I will remain silent about whether actions are obligatory, optional or impermissible, since combining this with a gradational account of rightness is beyond the scope of this thesis.

In response to question (ii), I will propose a maximising deontic theory in section 2.1.3, such that an agent ought to actualise the member of the relevant evaluative focal point (e.g. perform the act, internalise the decision procedure, or form the intention) that will produce the most all-things-considered value. I will not place an availability constraint on the member of an evaluative focal point that an agent ought to actualise, since the intuitions supporting an 'ought implies can' condition are adequately captured by the theory of all-things-considered value under global consequentialism, as was argued above.

Question (iii) will be the focus of section 2.2. In this section I will distinguish between different formulations of the subjective ought. I will argue that while there may be both an objective epistemic ought and subjective epistemic ought, the theory of epistemic value outlined in 2.1.1 removes any requirement for appealing to the most common formulations of the subjective ought: that is, belief-relative or *luminous* epistemic oughts.

Finally, in response to question (iv), the deontic theory I outline in section 2.1.3 will often be very different from the decision procedure that agents follow. The reasons for this have already been outlined above, but will be emphasised my application of epistemic consequentialism in chapter 3. In response to this objection I will show that the decision procedure recommended by my account of epistemic consequentialism is often non-identical to its criterion of rightness: that is, to the epistemic consequentialist deontic theory. I will therefore provide an example to show that this is the case in section 2.1.3, before differentiating these two things in the objections of chapter 3.

2 Epistemic Consequentialism

Chapter 1 outlined a general framework for consequentialist theories using examples from ethics. In this chapter I will build on the discussion of this general framework. In the first part of this chapter (2.1) I will construct an epistemic consequentialist theory that provides an answer to all of the key questions identified at the end of each section in chapter $1.^{38}$ In the second part of the chapter (2.2) I will outline the nature of the 'ought' that I argue should be employed within the resultant epistemic consequentialist theory.

2.1 An Epistemic Consequentialist Theory

In this section I will outline an epistemic consequentialist account of contributory final value (2.1.1), an epistemic consequentialist account of all-things-considered value (2.1.2) and an epistemic consequentialist deontic theory (2.1.3). When combined, these three components will form a complete epistemic consequentialist theory. I will then consider, in section 2.2, how this theory relates to the question of what an agent *ought* to do.

2.1.1 An Epistemic Theory of Contributory Final Value

A theory of contributory value specifies the things that are valuable in themselves and not instrumentally, but without aggregating these values. It is worth opening this discussion with a reminder of the key questions that an epistemic consequentialist theory of contributory final value must answer:

- What are the fundamental bearers of value in epistemic consequentialism?
- What is of contributory final value under epistemic consequentialism?
- Should we care about actual or possible epistemic value?

I will answer each of these questions in turn and use it to formulate an epistemic theory of contributory final value at the end of the section.

³⁸Epistemic consequentialism should not be confused with epistemic *instrumentalism*: the account of epistemic rationality objected to by Kelly (2003). Instrumental rationality is usually constructed on the basis of hypothetical goals ('if you want G then (you have reason to) do H') on the basis of subjective preferences. Utilitarianism is not a theory of instrumental value, and nor is epistemic consequentialism.

Bearers of Epistemic Value

The first question we must answer is what are the bearers of contributory final values. Within utilitarianism itself, the bearers of pro tanto value are typically formulated as complete states of affairs.³⁹ As such, the bearer of the contributory final value that results from an act in ethics is the *state of affairs* at which the act occurs. Within decision theory, a similar structure for the assessment of the utility of outcomes was first developed by Savage (1954/1972). The main components of Savage's theory are *states, actions* and *consequences.* Joyce provides an excellent summary of Savage's framework as follows:

'Savage's model envisions a *decision maker*, or *agent*, who chooses among risky or uncertain prospects, called *actions*, whose consequences or *outcomes* depend on the *state of the world*. One describes the agents decision problem by specifying a set \mathbf{A} of possible acts among which she must decide, a class \mathbf{O} of outcomes [consequences] that provides an inventory of all the desirable or undesirable things that could befall the agent as a result of her choice, and a list \mathbf{S} of states of the world that describe possible external conditions that determine what outcome each act in \mathbf{A} will produce.'

It will be useful to adapt Savage's model so that it is consistent with the terminology of this thesis. I will continue to refer to *states* as the fundamental bearers of value. However, these will be understood, following Jeffrey, as sets of propositions.⁴⁰ I will follow Joyce in treating states as a set of propositions completely describes a possible state of affairs. They do so to such a degree of accuracy that all apparent contingencies are provided with a determinate outcome at a state, and an agent's uncertainty can be construed as a lack of certainty about which state obtains.⁴¹ I will, however, replace the use of *acts*, each of which is a function from states to a unique set of consequences according to Savage, with the terminology of the previous chapter.⁴² The structure of Savage's treatment of acts need not change greatly, but since I will employ a global consequentialist framework in section 2.1.2, it will be useful to refer to 'members of an evaluative focal point' (or EFP members) instead of acts. As I stated above, evaluative focal points can be literally anything under global consequentialism, and not merely actions. Finally, I will use Savage's terminology of *consequences* to refer to the set of propositions that are produced when a member of an

³⁹Pg. 58, Vallentyne (1987).

 $^{^{40}}$ Pg. 58-85, Jeffrey (1964/1983).

⁴¹Pg. 15; 50-1, Joyce (1999).

⁴²Pg. 50, Joyce (1999).

evaluative focal point obtains at a given state.⁴³

As Joyce highlights, in cases where there are finitely many states, EFP members and consequences, we can represent the outcomes of a given decision problem in the following kind of matrix. Here I follow Joyce in using C[M, S] to refer to the consequence that an EFP member would produce if state S were to obtain in finitely many cases:⁴⁴

			STATES		
		S_1	S_2	S_3	 S_n
	M_1	$C[M_1, S_1]$	$C[M_1, S_2]$	$C[M_1, S_3]$	 $C[M_1, S_n]$
	M_2	$C[M_2, S_1]$	$C[M_2, S_2]$	$C[M_2, S_3]$	 $C[M_2, S_n]$
EFP	M_3	$C[M_3, S_1]$	$C[M_3, S_2]$	$C[M_3, S_3]$	 $C[M_3, S_n]$
MEMBERS	M_4	$C[M_4, S_1]$	$C[M_4, S_2]$	$C[M_4, S_3]$	 $C[M_4, S_n]$
	÷	÷	÷	÷	:
	M_n	$C[M_n, S_1]$	$C[M_n, S_2]$	$C[M_n, S_3]$	 $C[M_n, S_n]$
			CONSEQUENCES		

More may need to be said about each of these aspects of Savage's model as issues arise, but this overview will suffice for present. The answer to the first question is therefore that the fundamental bearers of value within epistemic consequentialism are *states* and *outcomes*.

Final Epistemic Value

The debate about what is of final value in epistemology has developed largely out of the James-Clifford debate between those who defend *veritism*, and those who defend *eviden-tialism*.⁴⁵ Veritism is the position that only true beliefs are of contributory final value, while evidentialism is the position that only well-evidenced beliefs are of contributory final value. In contemporary discussions of evidentialism it is more likely to hear advocates of the position argue that 'beliefs proportioned with the agent's evidence' are what are of final value in epistemology, and this is what I will mean by the term in what follows. There is a stronger position has also been defended, which I will call *knowledgism*. This is the view that only *knowledge* is of final epistemic value.⁴⁶

Each of these three positions - veritism, evidentialism and knowledgism - has been widely discussed in the literature on outright beliefs. However, in this thesis I will be

⁴³Pg. 52-3, Joyce (1999).

⁴⁴Pg. 49, Joyce (1999).

 $^{^{45}}$ For the two major texts in this debate, see James (1896) and Clifford (1879/1999).

⁴⁶A defence of this view can be found in Hattiangadi (2010). For further discussion see Part 1 of Haddock, Millar & Pritchard (Eds.) (2009).

concerned primarily with the epistemology of partial beliefs or *credences*. Credences are subjective degrees of belief that follow the standard Kolmogorov axioms of probability theory. I will assume here that credences are sharp, such that they map a proposition onto a real number $0 \le n \le 1.^{47}$ Finally, this discussion of credences will presuppose an objective Bayesian framework, such that there is a constraint on prior probabilities that determines the rational prior probability distribution over any given set of propositions.⁴⁸

Within the literature on partial belief, the predominant view is that *truth* is the primary goal of belief. This is evident when we look at how these theories tend to assess the epistemic value of an agent's credence function. A credence function assigns a real number $0 \le n \le 1$ to all of the propositions within the set of Savage states. A given credence indicates the agent's degree of confidence in the proposition in question. Therefore an agent will have credence 0 in all and only those propositions she is *certain* are false, and will have credence 1 in all and only those propositions she is *certain* are true. The standard way to assess the epistemic value of a given credence function is by employing a *proper scoring rule*. A proper scoring rule is something that is supposed to penalise an agent if she makes inaccurate predictions, and reward her if she makes accurate predictions. But accuracy and inaccuracy here are measured only in terms of 'closeness to truth'. For example, arguably the most common scoring rules employed by decision theorists is the quadratic or Brier score. The standard Brier score (BS) is calculated as follows:

$$BS = \frac{1}{N} \sum_{t=1}^{N} (f_t - o_t)^2$$

Where N is the number of cases, f_t is the credence that the relevant subject has in the relevant proposition or propositions P, and $o_t = 1$ where P is true and $o_t = 0$ where P is false. As such, a better Brier score is closer it is to 0, and a worse the Brier score is closer it is to 1. Also, if we were to ask an agent whether P would be true or false over 100 cases, and she were to toss a fair coin and adopt credence 1 in P if it landed heads, and credence 0 in P if it landed tails, then we would expect her Brier score to be around 0.5 regardless of the probability that P or $\neg P$ will occur in each case.

 $^{^{47}}$ I believe that many of the substantive points of this thesis are consistent with the possibility of 'mushy credences', but will not discuss these here. For more on this topic, see White (2010).

⁴⁸I assume that prior to any empirical information about a proposition, a subjects credal function should satisfy the principle of maximum entropy. See Jaynes (1957). For a more recent approach to determining objective priors, see section 3 of Hutter (2007).
To illustrate the Brier sore with an example, imagine we ask an investor, Carol, whether she thinks Stock A will go up or down in price on a given day. She responds that there is a 70% chance Stock A will go up on Monday, a 45% chance Stock A will go up on Tuesday, and a 10% chance Stock A will go up on Wednesday. Suppose we also ask for the predictions of a second investor, Dianne. She thinks that there is in fact a 30% chance Stock A will go up on Monday, a 40% chance Stock A will go up on Tuesday, and a 95% chance Stock A will go up on Wednesday.⁴⁹ As it turns out, Stock A goes down on Monday and Tuesday, but it goes up on Wednesday. We can represent this as follows:

Stock	Down (Mon)	Down (Tue)	Up (Wed)
Carol's Credence in 'Up'	0.7	0.45	0.1
Dianne's Credence in 'Up'	0.3	0.4	0.95
Carol's Error	0.3	0.55	0.9
Dianne's Error	0.7	0.6	0.05

We can use the data on this table to work out Carol's Brier score as follows:

$$BS = \frac{1}{3}((0.7 - 0)^2 + (0.45 - 0)^2 + (0.1 - 1)^2) = 0.5008\dot{3}$$

Since the Brier score is better if it is lower, this is clearly a poor score. We can use this data to work out Dianne's Brier score in the same manner:

$$BS = \frac{1}{3}((0.3 - 0)^2 + (0.4 - 0)^2 + (0.95 - 1)^2) = 0.0814\dot{6}$$

From this we can see that Dianne was better at predicting whether Stock A would go up on each of the days than Carol was, since Dianne's score of $0.0814\dot{6}$ is much better than Carol's score of $0.5008\dot{3}$. This is what we would expect given their respective predictions.

But the Brier score, and other proper scoring rules, can be objected to on the same basis that veritist theories of epistemic value have been criticised in the literature on outright belief, since it only scores people on how close their estimates are to the truth.⁵⁰ These scoring rules fail to take into account how an agent acquired her credence, or how likely it is that she would, given similar circumstances, adopt a similarly accurate or similarly inaccurate credence in the relevant proposition. Because it only scores an agent on the

⁴⁹I assume that both investors also think there's a 100% chance that the Stock will *either* go up or down on each of these days, such that her credence in 'Down' is one minus her credence in 'Up'.

⁵⁰Indeed, these criticisms apply no less to linear, logarithmic and spherical scoring rules. This is because the critique of the Brier score offered here is not a critique of any particular scoring rule, but of the veritistic treatment of value in this and similar scoring rules.

basis of a series of actual cases, the efficacy of a given method can only be derived from how well it does in those very cases. But these cases may represent a non-representative sample or atypical sample. Consider the following two examples:

Lucky Lucy

Lucy is a scientist at a famous research institution. Over the course of her career she has won many accolades and prizes, and is considered by many to be the best scientist of her generation. But Lucy has a secret. She was never actually that good at science. So when her science grades started to slip in high school, she found a fair coin on the ground and decided to try a little experiment. Every time she was presented with a proposition, she would flip the coin. If it landed heads, she would adopt credence 1 in that proposition. If it landed tails, she would adopt credence 0 in that proposition. She'd certainly make sure that she comprehended the meaning of the relevant propositions, but the coin would completely govern the confidence she would place in each. As luck would have it, the coin only ever landed heads for the propositions that were true, and tails for the propositions that were false. Lucy started to see her grades improve, and eventually used the coin to get her through graduate school and went on to publish papers full of complicated results that were all true. The coin wasn't rigged in any way and if, in the future, it ever lands heads for a false proposition, Lucy will adopt credence 1 in that false proposition.

Diligent Dilbert

Dilbert is a scientist working at a small but well-respected university. He has had a somewhat uneventful career, but everyone agrees that he is a competent scientist whose work has consistently been of a good quality. He has always been good at science. He has a great regard for the best available scientific methods, and is very meticulous at checking and rechecking his results using these methods. When considering the hypotheses within his field, Dilbert never adopts a credence that is either too high or too low given the evidence he has. For this reason, he rarely has full confidence in propositions that aren't either mathematical truths or logical tautologies.

It seems clear that Lucy's behaviour is epistemically reckless, and that Dilbert is a much better epistemic agent. After all, Lucy does't base any of her beliefs on her evidence, and by continuing to use the coin flip method she could get things incredibly wrong at any moment. But the Brier score of these agents tells a very different story. Since high school, Lucy has believed all true propositions with credence 1 and all false propositions with credence 0. This means that her Brier score for her credal function for all these years has been a perfect 0. Dilbert, on the other hand, has rarely has credence 1 in any empirical proposition. Instead he has adopted a credence in a proposition between 0 and 1 that is proportional to the evidence supporting that proposition. Because of this, no matter how accurate Dilbert's credences are, his Brier score is *guaranteed* to be worse than Lucy's.

This case shows that the Brier score is liable to fail in at least two ways. Firstly, it doesn't adequately capture our intuitions about what makes for a better or a worse epistemic agent. This is not a trivial error, since much of the evidence supporting the use of the Brier score as a measure of predictive accuracy is its ability to capture our intuitions about good and bad predictive agents. But it also shows a failure to capture good methods. Lucy's method is poor and liable to fail at any moment, and yet this won't be reflected in her Brier score until she actually *does* fail. Surely what we want from a scoring rule is an ability to consistently reward credences that are closer to the truth because they are founded on reliable methods, and to reward these more than lucky guesses. If both of these criticisms hold, then it seems likely that the thing we should treat as our contributory final value in epistemology is not merely high credences in true propositions and low credences in false propositions. If these were the only things of contributory final value in epistemology then Lucy would be a perfect epistemic agent. But Lucy is not a perfect epistemic agent.

There have been similar objections within the epistemology of outright belief to Gettiered beliefs: beliefs that are justified and true, but whose justification was highly dependent on luck.⁵¹ This has led to a wide acceptance of the thesis that there must be an anti-luck condition on knowledge. One example of an anti-luck condition is the safety condition on knowledge defended by Williamson. He argues that an agent believes P safely only if that agent could not easily have failed to believe P in sufficiently similar circumstances using a sufficiently similar method. This means that an agent forming her beliefs reliably is a necessary but not sufficient condition for knowledge. As Williamson states:⁵²

'...in given circumstances, something happens reliably if and only if it is not in danger of not happening. That is, it happens reliably in a case α if and only if it happens (reliably or not) in every case similar enough to α . In particular, one avoids false belief reliably in α if and only if one avoids false belief in every case similar enough to α . When the danger is a matter of degree, reliability involves a trade-off between the degree to which the danger is realized and the closeness of the case in which it is realized. A very high degree of realization in a not very close case and a lower degree of realization in a closer case both make for unreliability.'

 $^{{}^{51}}$ See Gettier (1963).

⁵²Pg. 124, Williamson (2000).

Could a safety condition be applied within the epistemology of partial belief? I argue that it could. Doing so would require that we create a similarity measure for Savage consequences in order to talk about which consequences are sufficiently similar to the consequence in question to be included in the safety condition. But this could be done in a manner entirely consistent with the similarity measures employed in the safety conditions within the epistemology of outright belief. I am thinking here of Lewis' similarity measure.⁵³ As such, I will not focus on articulating a similarity measure for states here.⁵⁴ Instead, I will follow Williamson in stipulating that a consequence is sufficiently similar to another consequence only if these consequences share the same history from the beginning of the case until the point that the relevant EFP member is actualised.⁵⁵ Beyond this, we must use good judgment in determining which consequences are sufficiently similar to be included in the credal safety condition. Which consequences are sufficiently similar will also be partially determined by context. In what follows, I will only use consequences that we can judge to be unambiguously similar enough to the actual consequence, such that if the agent believes falsely in this similar consequence then she is clearly in danger of being wrong. Let \mathcal{L} be the set of consequences that are sufficiently similar to a consequence c^* for those members of \mathcal{L} to be relevant to the safety of an agent's credences at c^* . Finally, since c^* is always sufficiently similar to itself, it is necessarily the case that $c^* \in \mathcal{L}$. We can use this to formulate a version of the Brier score that incorporates a credal safety condition as follows:

$$SBS(c^*) = \frac{1}{N} \sum_{c \in \mathcal{L}}^N BS_c$$

where

$$BS_c = \left(\frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2\right)$$
 at consequence c

In this 'safety Brier score' (SBS), N is the number of outcomes in the set of sufficiently close consequences $(c \in \mathcal{L})$ to the actual consequence c^* , while BS_c is the Brier score at the relevant consequences under assessment. Within each sufficiently similar case (BS_c) , N is the number of cases at c, f_t is the credence that the relevant subject has in P. In BS_c , as in the original Brier score, $o_t = 1$ where P is true and $o_t = 0$ where P is false.

⁵³Pg 48-52; 91-95, Lewis (1973).

 $^{^{54}}$ Specifying a closeness relation for Savage consequences will be an interesting component of a larger exploration of this topic, but it is beyond the scope of this thesis.

⁵⁵Pg. 123-4, Williamson (2000).

Credal functions that receive a score that is close to 0 on the safety Brier score are more likely to both approximate truth *and* be the result of a reliable method. Unlike strict evidentialist views that give *no* weight to a belief being true, the safety Brier score takes the truth or falsity of the proposition into account. But it will also reward a high credence in a false proposition that is formed on the basis of a reliable method, and punish a high credence in a true proposition that is formed on the basis of an unreliable method. Therefore, unlike strict veritist views that give *no* weight to a belief being formed in proportion to one's evidence, the safety Brier score rewards credences that are formed using such an evidentialist method insofar as that method is a reliable one.

If we assume a knowledge-first approach in the epistemology of outright belief, then the safety Brier score is not analogous to knowledgist accounts in the epistemology of outright beliefs. However, it is able to capture many of the intuitions driving both veritist and evidentialist accounts of final epistemic value. It will also become evident, in section 2.2, that using the safety Brier score has the virtue of removing any need to appeal to more than one kind of ought, and that it is able to capture our intuitions in the higher-order evidence cases of chapter 3. For these reasons, I will treat good safety Brier scores as being of contributory final value within epistemic consequentialism. I will remain open to the possibility of there being further things of contributory final value or disvalue within epistemic scores in this thesis.

Actual vs. Possible Value

The final question that an epistemic consequentialist theory of contributory final value must address, is whether it is *actual* or *possible* consequences that are of value. To make the distinction clear, imagine that Aadila is out walking when she is suddenly approached by an epistemic jinn who offers her the following deal: if Aadila can recite Charles Dickens' *Little Dorrit* word-for-word, then the jinn will adjust her credences so that she will have a safety Brier score that is slightly better than it currently is. But if Aadila chooses this option and makes even a single mistake when she recites *Little Dorit*, then the jinn will constantly meddle with her credences, such that her safety Brier score will be dreadful for the rest of her life. If Aadila feels she is unable to recite *Little Dorrit* word-for-word, then she can choose the second option of having the jinn leave her and her credences alone. Suppose Aadila hasn't read *Little Dorrit* since she was very little, and can hardly remember anything from the novel. Which of the jinn's options should Aadila choose?⁵⁶

Since our account stipulates that *possible* safety Brier scores approaching 0 are of value, then it will recommend that Aadila attempt to recite *Little Dorrit* word-for-word. This is because in doing so produces more possible value than not doing so. After all, there is the possibility of improving her credences slightly if she chooses to try to recite, whereas choosing not to recite offers no possible improvement of her credences.⁵⁷

It is not only problematic that such cases are inconsistent with our intuitions about what Aadila ought to do. In addition, adopting the possibilist account of value means that we either have to employ an independent 'ought implies can' principle, or we have to accept that agents often ought to do things that are extremely difficult, extremely high-risk, and will yield only a small amount of utility if they are successful. As I argued in section 1.2.3, an actualist account of value need not stipulate a fundamental 'ought implies can' principle. The results of the actualist account are already consistent with our intuition that it's not the case that an agent ought to do something that it is extremely difficult for her to do for some small gain. This is because the account appeals to the *actual* consequences that will result from the actions of the relevant agent, and if the action is extremely difficult then the actual consequences will be those that occur when the agent is unsuccessful. In order to retain the theoretical simplicity of the actualist account of the 'ought implies can' intuition, it is an actualist account of value that I will assume in what follows. This means that it is the *actual* scores that an agent does or will have on the safety Brier score that are of value, not the *possible* scores she could have if she managed to overcome various obstacles that result from either facts about herself or facts the world in which she finds herself.

An Epistemic Consequentialist Theory of Contributory Final Value

I have now answered the three main questions for an epistemic consequentialist theory of contributory final value that arose from the previous section. The epistemic consequentialist theory of contributory final value that has been constructed in this section is as follows:

⁵⁶Within veritist accounts of value that do not have a safety condition, we could use the following example: Aadila can either guess whether or not P is true and adopt credence 1 or 0 in it respectively with a 1 in 100 chance of being correct, or she can proportion her credence in P with her evidence, where this supports a credence n such that 0 < n < 1. Since the possible value of doing the former is greater than the possible value of doing the latter, Aadila ought to do the former under a possibilist account of value.

⁵⁷This worry is raised in pg. 57, Ord (2009) in response to Feldman's (1986) possibilist account of value.

Objective Epistemic Consequentialism:

i. theory of contributory final value:

A credence or set of credences that actually score well on the safety Brier score have *pro tanto* epistemic value as ends (where the degree of value is proportional to the degree to which the safety Brier score approaches 0).

A credence or set of credences that actually score poorly on the safety Brier score have *pro tanto* epistemic disvalue as ends (where the degree of value is proportional to the degree to which the safety Brier score approaches 1).

There may exist other things that have *pro tanto* epistemic value or disvalue as ends.

This account of contributory final value in epistemology is consistent with a *threshold* reading, which states that something is pro tanto valuable if it produces any credence or set of credences with a safety Brier score less than n and pro tanto disvaluable if it produces any credence or set of credences with a safety Brier score greater than n. However, it is far more plausible to read the account given above as a *gradational* account of contributory final value. Under a gradational reading, something is valuable or disvaluable to degree m, where m is proportional to the safety Brier score of the credence or set of credences it produces. There will still be thresholds at which something is deemed disvaluable (of negative value), valuable (of positive value), or value-neutral (of zero value), but the value of the things within these categories will come in degrees. This is analogous to the gradational account of rightness and wrongness given in section 1.2.3, and it is this gradational reading of contributory final value that I will assume for the remainder of this thesis.

2.1.1.1 The Problem of Trivial Truths

It has been argued by Grimm (2009) that any veritist consequentialist account of contributory final value faces a dilemma regarding *trivial truths*. Trivial truths include all of the propositions that you would place very low on a ranking of the importance of propositions. Imagine counting all the blades of grass in Port Meadow, or what the 323rd entry in the Wichita, Kansas, phone directory is.⁵⁸ These are propositions in which true beliefs hardly seem worth the effort of acquiring. Here is Grimm's dilemma formulated for the safety Brier score account of epistemic value given above:

⁵⁸The second example is given in Goldman (1999).

GRIMM'S DILEMMA⁵⁹

Safe Credences: Safe credences are of epistemic value. Either you hold an *unrestricted* view of the value of safe credences, such that safe credences in trivial truths are epistemically valuable, or a *restricted* view of the value of true belief, such that only safe credences in propositions hat are interesting or important to us are epistemically valuable.

Epistemic Consequentialism: Beliefs should only receive positive or negative appraisal insofar as they do well or do badly when it comes to the things that are of epistemic value.

Horn 1 - The Restricted View: If a restricted view of the value of true belief holds, then we cannot appraise beliefs in trivial propositions. But we *can* epistemically assess credences in trivial propositions. For example, a lucky guess that there are seven pens in my bag is epistemically worse than a high credence that there are seven pens in my bag after carefully checking. Therefore either epistemic consequentialism is false or we must adopt the *unrestricted* view of the value of safe credences.

Horn 2 - The Unrestricted View: The unrestricted view of safe credences is implausible. It cannot be the case that adopting safe credences in any proposition, no matter how trivial, is worthwhile from the epistemic point of view.

There are several different ways that the epistemic consequentialist can respond to Grimm's dilemma. Let me offer two here: one that is consistent with the first horn of the dilemma, and one that is consistent with the second. Though I hold to the second view, I will let the reader decide which of these accounts of epistemic final value they find more plausible:

(1) The Restricted View

How do we construct a restricted view of the value of safe credences that is consistent with our ability to epistemically assess credences in trivial propositions? One way is to construct an importance ordering on propositions. For example, suppose we take whatever criteria underly our intuitions about what constitute trivial and non-trivial propositions. Suppose also that all propositions we range anywhere from being extremely, or even infinitely, important to being important to degree 0: that is, being unimportant to anyone, anywhere, at any time. Given this, can construct an unbounded importance score, such that all propositions are important to degree n, where $n \ge 0$. The epistemic utility of a given credence is some function of its safety Brier score and its degree of importance. All propositions can therefore be assessed for their epistemic utility, even if they are completely trivial.

⁵⁹Pg. 248-251, Grimm (2009).

(2) The Unrestricted View

How do we defend the thesis that all safe credences are of final epistemic value? Suppose we adopt a pluralist account of value, such that there may be various values that lie outside the epistemic value-type. For example, it may be that things can be morally, aesthetically or prudentially valuable. For example, discovering facts about the world will often be conducive to producing greater epistemic utility. If I have safe credences in propositions about how to assess the cost-effectiveness of a charitable organisation, or how to employ the collective findings of the natural sciences to avert natural disaster, then I can save many more lives than if I have safe credences in how many blades of grass there are in Port Meadow. The safe credences in these sets of propositions will produce more *ethical* utility, and therefore the propositions themselves seem more important to us. But the source of their importance comes from their all-things-considered value across more than one valuetype, not from their being of more or less importance within the *epistemic* value-type. We can therefore explain our intuitions that not all safe credences are equally worth pursuing, without having to appeal to variance in the epistemic value of those propositions to do so.

2.1.2 An Epistemic Theory of All-Things-Considered Value

I will begin this section, as I did with the last, with an overview of the main questions raised by the previous chapter. In section 1.2.2 I proposed that an epistemic consequentialist theory of all-things-considered value must answer the following two questions:

- What are the evaluative focal points assessed by epistemic consequentialism?
- How do we work out all-things-considered value under epistemic consequentialism?

As before, I will answer each of these questions in turn and use these answers to construct an epistemic consequentialist theory of all-things-considered value at the end of the section.

Evaluative Focal Points

The main evaluative focal points that epistemologists attend to are either *outright beliefs* or *credences*. And within this assessment it is primarily the *constitutive* value of the relevant belief or credal state that epistemologists have focussed on - how justified, rational, or evidentially supported the belief or credal state in question is. But is it necessary for an epistemic consequentialist theory to focus solely on what are standardly seen as 'epistemic

focal points'? I argue that it is not, and in this section I want to move away from the treatment of beliefs and credences as the only evaluative focal points that should be of interest to the epistemologist. In the discussion of all-things-considered value that follows, I will also move away from the view that the all-things-considered epistemic value of a member of an evaluative focal point is exhausted by its constitutive value.

As I stated in section 1.2.2, global consequentialism is the view that consequentialist theories can directly evaluate the rightness or betterness of not only acts and rules, but also decision procedures, character traits, physical objects and anything else we might be interested in evaluating. It is a position whose central position has been advocated by several different consequentialist ethicists.⁶⁰ However, arguably the most comprehensive study and defence of global consequentialism in recent years has been given by Ord (2009), and it is this account that I will use to orient this discussion of a global epistemic consequentialism.

There are various reasons one might want to extend the evaluative focal points of a consequentialist theory beyond those of acts and rules. This should be most evident when we consider epistemic consequentialism. After all, even if the cognitive acts of adopting a belief or adopting a credence function could be fitted into an act-consequentialist framework, this still leaves things like decision procedures untouched by our consequentialist theory. But we can intelligibly ask whether decision procedure A is epistemically better than decision procedure B. We can also ask whether an institution like the Bodleian library is epistemically better than an institution like the Playboy mansion. Suppose a newly established university were in the unusual situation of deciding between whether to erect a brick-for-brick copy of the Bodleian library or a brick-for-brick copy of the Playboy mansion. The university wants to erect the institution that is better epistemically, i.e. that will produce the most epistemic value. Surely we can compare the epistemic value that has resulted from both institutions in order to recommend which of their brick-for-brick copies, if erected, is more likely to produce more actual epistemic value in the future.

It might be objected that, unlike global consequentialism in ethics, global *epistemic* consequentialism cannot really offer a *direct* evaluation of decision procedures and institutions. This is because one of the contributory final values according to epistemic consequentialism is a credal state that has a safety Brier scores approaching 0. But the only way an agent can have such a credal state is by performing the cognitive *act* of adopting that credal state.

⁶⁰Most notably, Railton (1988), Kagan (2000), Parfit (1984) and Petit and Smith (2000).

Hence global epistemic consequentialism will always assess evaluative focal points like decision procedures *indirectly* in terms of the cognitive acts they produce. But this objection rests on the mistaken assumption that is valuable under epistemic consequentialism is the *act* of adopting a credal state that has a safety Brier score approaching 0. This is not true. It is the credal state itself that is of value, just as what is valuable in hedonistic consequentialist theories is not the *act* of having a happy experience, but the happiness itself. So the epistemic consequentialist is no more committed to an indirect theory than its ethical counterpart is. As such, I will assume in what follows that an epistemic consequentialist theory can assess different evaluative focal points *directly*.

How are we to assess different evaluative focal points? Giving an answer to this question is difficult, since there are many problems that can be raised for different accounts of how we should assess different evaluative focal points. Rather than try to deal with these issues here, I will adopt the account advocated in Ord (2009), who offers this framework in response to a much fuller discussion of the relevant problems.⁶¹ As such, I will assess the epistemic rightness of an evaluative focal point *within a role*, such that:⁶²

epistemic global consequentialist criterion of betterness within a role

x is better than y in role R iff the epistemic value of x being in role R is greater than the epistemic value of y being in role R

By the 'role' that x is in I simply mean the type of thing that the evaluative focal point is *doing* in an intuitive sense: such as an act being *imagined*, a character trait being *adopted*, or a film being *ignored*. Given this, we can legitimately ask whether 'dancing' is a better act to imagine than 'killing', or whether 'Saw' is a better film to ignore than 'Casablanca'.

I will discuss specific roles for decision procedures below. Before doing so, however, there are two things about the criterion of betterness given above that are worth noting. The first is that although it is consistent with an epistemic version of the scale of rightness given in section 1.2.3, it only produces an ordinal ranking of EFP members in a role. We can use this to compare two acts, for example, to determine which is better. We can also use it to find out which is the best act (or set of acts) withinin a given set, since this will be the one that no other members bear the 'better than' relation to. But we cannot indicate *how much* better one member of an evaluative focal point is than another in a given role, since

⁶¹See chapters 3, 4 and 6 of Ord (2009). This discussion relates primarily to chapter 3.

⁶²This is the epistemic analog of the principle given in pg. 39 of Ord (2009).

it doesn't give a cardinal ranking. However, since Ord has pointed out some difficulties for giving a cardinal comparison of EFP members in a role⁶³ and since an ordinal ranking of EFP members still lets us say which member an agent *ought* to actualise, it is this ordinal account that I will adopt in what follows.⁶⁴

The second thing worth noting is that we want to compare two EFP members A and B that are genuine alternatives, i.e. it cannot be the case that both A and B fill the relevant role. In what follows I will compare EFP members that are what Ord calls 'soft alternatives', which I will simply call 'alternatives'. EFP members A and B are 'alternatives' to be in role R if the following condition holds:⁶⁵

(Alternatives) A and B are alternatives iff 'were x to be in the role R, y would not be and were y to be in the role R, x would not be.'

In the discussion that follows, I will assume that the above criterion of betterness can only be applied to two or more EFP members that are alternatives to one another.

The evaluative focal points that I will focus on in this thesis are those of *cognitive acts*, such as adopting a given credal function, and *decision procedures*, such as 'maximise epistemic utility' or 'update on your evidence in accordance with Bayes' Theorem'.⁶⁶ But what are the roles in which we should assess members of these evaluative focal points? In a way this isn't a substantive question, since we could assess these focal points in *any* role, such as the best cognitive act to forget, or the best decision procedure to write on a chalk board. But we will be more interested in EFP members when they are in some roles than when they are in others. It is easier to identify the role in which we want to compare different cognitive acts: this is the role of *being adopted*.⁶⁷ After all, we are interested in the credence function an agent adopts, not the credence function an agent remembers, formalises or laughs at. Bur decision procedures are more difficult to categorise. An initial response might be that we should assess decision procedures in the role of 'being followed',

⁶³Pg. 28-40, Ord (2009)

 $^{^{64}}$ This is not to suggest that a cardinal ranking of EFP members in a role could not be constructed in addition to this ordinal account. Doing so would, however, require that we address the objections raised for such an account, which is beyond the scope of a thesis of this size.

⁶⁵Pg. 40, Ord (2009).

⁶⁶Chapter 4 of Ord (2009) is concerned with the nature of decision procedures. Since decision procedures are a common topic within the epistemology of partial belief, appealing to the common understanding of what constitutes an epistemic decision procedure is sufficient for the purposes of this thesis.

 $^{^{67}}$ Why not the role of *being performed*? The role *being adopted* treats cognitive acts less like willed non-cognitive acts, since it is more usual to say that I adopt a credence function than that I 'perform the act' of adopting a credence function.

but the same problems for outlining what it is to follow a rule seem to apply to outlining what it is to follow a decision procedure. Does an agent follow a decision procedure when she conforms to it to the best of her ability, when she executes it perfectly, or when she internalises or commits herself to that decision procedure?⁶⁸

In this thesis I will primarily be interested in the best decision procedures in the role of 'being committed to' by the relevant agent. It is very important that we get clear on what is meant by this, since differentiating the best decision procedure to 'commit to' from the best decision procedure to 'execute perfectly' will be crucial in the application of epistemic consequentialism to peer disagreement cases in chapter 3. An agent commits herself to a decision procedure by being disposed to follow the decision procedure in the cases where it applies, even if she doesn't execute it perfectly each time she is faced with a decision problem. My understanding of a decision procedure's 'being committed to' is largely consistent with a modified formulation of Boghossian's account of following an epistemic rule at the sub-personal level. I will require that the three conditions *disposition, correctness* and *explanation* are satisfied if an agent is committed to a decision procedure as follows:⁶⁹

(Disposition) If S has committed to decision procedure D, then S is disposed to make decisions in accordance with D.

(Correctness) If S has committed decision procedure D, then S acts correctly relative to her commitment to D if, in a case where D stipulates she should do A, she does A. She has acted incorrectly relative to her commitment to D if, in a case where D stipulates that she should do A, she does not do A.

(Explanation) If S is complying with D by doing A, then S's commitment to D at least partially explains S's doing A.

Under this account, committing to a decision procedure need not be done consciously by the agent (though it certainly *can* be).⁷⁰ For example, I do not remember ever consciously committing to the decision procedure D_1 : 'if you know that P and you know that P entails Q, then believe that Q'. And yet I have satisfied the above three conditions for D_1 since

 $^{^{68}}$ Ord discusses these three roles for decision procedures at length in chapter 4, Ord (2009).

⁶⁹For the account given by Boghossian, see pg. 482-3, Boghossian (2008). Boghossian goes on to raise worries for both this account of rule-following and the dispositional account of rule-following to which it is related. A more extensive defence of epistemic consequentialism must respond to both Boghossian's worries about rule-following in its account of the role in which we should assess decision procedures, and to Ord's regress problem for decision procedures (pg. 96-9, Ord (2009)).

⁷⁰That is, an agent need not make any conscious decision to commit to a given decision procedure in order to do so, and nor must she be consciously aware that she is committed to the relevant decision procedure.

childhood. If I knew that addition in finite arithmetic is commutative, and I knew that a + b was an instance of finite arithmetic, then I would believe that a + b = b + a. I was therefore disposed to believe in accordance with D_1 , and my belief that a+b=b+a could in part be explained by a commitment to this decision procedure. Furthermore, if I had known the first two facts but had failed to believe that a + b = b + a, then I would have certainly acted incorrectly relative to my commitment to D_1 .

Committing to a decision procedure also doesn't require that one can execute it perfectly in every case. It is entirely consistent with the above conditions that an agent might be committed to a decision procedure without always successfully complying with it. As Ord points out, she might fail to comply with the decision procedure in a given set of cases because of cognitive deficiencies, weakness of will, and so on, without falling foul of her commitment to the decision procedure.⁷¹ As stated above, I will return to this issue in chapter 3 by showing that in cases of peer disagreement, the best decision procedure to commit to is very different from the best decision procedure to execute perfectly.

The best decision procedure to commit to will be assessed answerable to the actualist conception of value discussed in the previous section. As such, it will rarely be the case that the best decision procedure to commit to is one that it is very difficult for an agent to actually comply with. As Ord states:⁷²

'Since [commitment] does not simply assume compliance, any factors that a decision procedure has which are conducive to such compliance will be properly taken into account in the assessment. Thus decision procedures that actually achieve less good because they are computationally difficult or highly demanding or do not excite moral computation will be appropriately penalised for this.'

This is similar to the result that was outlined in section 1.2.3, where it was shown that the best thing for Adam to do was to try to give £10 to charity instead of trying to give £100,000 to charity. Sine Adam didn't have £100,000, the actual value of his trying to give away this sum of money was lower than the actual value of his trying to give away the £10 that he had. Similarly, a decision procedure like 'adopt a credence function such that you always have a safety Brier score of 0' will be incredibly difficult for any non-omniscient agent to follow. As such, the actual epistemic value of committing to this decision procedure will be lower than the actual epistemic value of committing to a cluster of simple epistemic

⁷¹Pg. 68, Ord (2009).

 $^{^{72}}ibid.$

heuristics like 'always increase your credence in P if you perceive that $P'^{.73}$

By assessing the best decision procedures to *commit to* we also avoid the problem faced by Dr. Goodcalc in section 1.2.3. This is because if a good decision procedure 'G' takes 9 seconds to calculate, then there will be a better decision procedure that Dr. Goodcalc can commit to, namely G+ 'If you have 9 seconds to calculate, calculate in accordance with G. If you don't have 9 seconds to calculate, then calculate in accordance with H (where 'H' is a simple heuristic such as 'perform the action that is intuitively the right one')'.

Before moving on, I must briefly attend to the most salient objection to this view, which Ord calls the 'inconsistency objection'.⁷⁴ In the epistemic case, the inconsistency objection arises because the credence in P that the best decision procedure to be committed to recommends might be different from the credence in P that the best credal function to adopt recommends. Brad Hooker expresses this objection as follows:⁷⁵

'Suppose...the best decision procedure for you to accept is one that leads you to do act x now. But suppose also that in fact that the act with the best consequences in this situation is not x but y. So global consequentialism tells you to use the best possible decision procedure, but also not to do the act picked out by this decision procedure. This seems paradoxical.'

Consider an example of this from epistemology. April is in a perfectly normal world that is free from brains in vats, veridical hallucinations, fake barns and the like. Therefore the decision procedure 'always increase your credence in P if you perceive that P' is a very good decision procedure for her to be committed to, and indeed she is committed to this decision procedure. One day, April believes she sees a small insect fly in front of her, and so she increases her credence that an insect flew in front of her. But April is actually suffering from myodesopsia and has only believed she has seen an insect fly in front of her. In fact, a floater in April's eye has cast a shadow on her retina that looks identical to a small insect flying in front of her. Since a good decision procedure doesn't require perfect compliance, and since a decision procedure that asks agents to discriminate between phenomenally identical experiences will be impractical, April may have still complied with the best decision procedure to commit to when she increases her credence. But when we ask what the best credence for April to adopt is, the outcome will be that it is better for

 $^{^{73}}$ This is true even if 'adopt a credence function such that you always have a safety Brier score of 0' is much closer to the *criterion of rightness* of epistemic consequentialism, as I shall discuss in section 2.1.3.

⁷⁴Pg. 91, Ord (2009).

 $^{^{75}}$ Section 5, Hooker (2008).

April not to increase her credence that a small insect has flown past her, since doing so will worsen her safety Brier score. Does this mean that assessing all evaluative focal points directly leads to inconsistent recommendations about what an agent should do?

The simple response we can give to this objection is that April should both commit herself to the relevant decision procedure, and at the same time she should not increase her credence that she is seeing a an insect in the case I outline above. After all, committing to a decision procedure doesn't require compliance with it in *every* case. As Ord points out, the objection that it is inconsistent for the global consequentialist to respond in this way seems to rely on something like the following principle being true:⁷⁶

'If S ought to X, and S cannot X without also Y-ing, then it is permissible for S to Y' Here I take 'Y is permissible' to simply mean 'it is not the case that S ought not to Y'. To see why these different recommendations are not inconsistent under actualist global consequentialist note that, where $O_s(P)$ is taken to mean 'S ought to P', the following is not a principle of deontic logic:

$$O_S(P \land Q) \rightarrow O_S(P) \land O_S(Q)$$

Why would the actualist deny that conjunction elimination is a principle of deontic logic? The standard cases that are given in defence of this view all have the following structure: suppose that Bella has been drinking late into the night and she is now at home deciding what time she should set her alarm clock to. She knows that the best thing to do would be to both wake up at 8am and go to her ethics seminar (10hu). But she also knows that if she wakes up at 8am then she won't go to the seminar but will lie in bed feeling hungover (-15hu). If she sets her alarm to 11am instead, then she'll sleep off most of the hangover (2hu). What time should Bella set the alarm to? Suppose that Bella is correct and that waking up at 8am going to the seminar would result in the most actual value. We can represent the actual values of each of the three outcomes as follows:⁷⁷

	8am	11am
Seminar	10hu	—
No Seminar	-15hu	2hu

It should be clear that Bella *ought* to [wake up at 8am and go to the seminar] according the actualist. Given this, if the conjunction elimination principle holds, then Bella ought

 $^{^{76}}$ Pg. 93, Ord (2009). In the passages that follow I am entirely indebted to Ord's analysis of the inconsistency objections and possible responses to them on behalf of the actualist global consequentialist.

 $^{^{77}}$ This is similar to the case given in pg. 49-50 of Ord (2009).

to wake up at 8am. But if Bella wakes up at 8am then she won't go to the seminar but will instead suffer 3 hours of a hangover when she could have slept through it instead. So out of the options of [waking up at 8am] or [waking up at 11am], Bella ought to chose [waking up at 11am] according to the actualist, since this will create more actual value than choosing the former will. As such, conjunction elimination in deontic logic is inconsistent with the mist basic premise of actualism: that we should consider only the *actual* value that results from an EFP member being in a given role when assessing that EFP member.

How does this conflict with the above principle? Replace S with Bella, X with 'wake up at 8am and go to the seminar' and Y with 'wake up at 8am'. Then we get the following:

'If Bella ought to wake up at 8am and go to the seminar, and Bella cannot wake up at 8am and go to the seminar without also waking up at 8am, then it is permissible for Bella to wake up at 8am (i.e. it is not the case that Bella ought not to wake up at 8am).'

The antecedent of this principle is true according to actualism, since if Bella actually [wakes up at 8am and goes to the seminar] then she maximises actual value. But if Bella actually [wakes up at 8am], this will produce the least amount of actual value out of the three options. So if actualism is true then Bella ought not to wake up at 8am. But if this means that the consequent of this principle is false. Therefore if this principle is true then actualism about value is false. But then this principle cannot be used to pose inconsistency objections to global consequentialist theories with an actualist conception of value without begging the question against those very theories.⁷⁸ Given this rejoinder to the objection, I will now move on to outline an epistemic consequentialist theory of all-things-considered value.

An Epistemic Consequentialist Theory of All-Things-Considered Value

How do we work out all-things-considered value under epistemic consequentialism? Given the discussion of the previous section, the answer to this question should be fairly straightforward. Firstly we must sum *all* of the epistemic contributory final value produced by the relevant EFP member in a given role. This means that the all-things-considered value of a given EFP member will take into account not only the *constitutive* value of that member, such as the Brier score of the credences in a specific case, but also the *causal* value of that

⁷⁸For a far more comprehensive treatment of the inconsistency objections, see pg. 91-6, Ord (2009).

EFP member, such as the Brier score of the credences that those credences help to produce. After all, we can imagine many cases in which a lucky guess could produce a lot of future epistemic value. Suppose, for example, that god exists but that fallible epistemic agents such as ourselves must partake in a Kierkegaardian leap of faith if they are to ever adopt a high credence in the proposition that god exists. There may be many propositions that rest on the premise 'god exists' that the relevant agent can safely come to know once they have made this leap, but that would be inaccessible to them otherwise.⁷⁹

Once we have summed the epistemic contributory final value produced by a given EFP member in a given role, we can use this to construct either a *cardinal* or *ordinal* ranking of the different EFP members that are alternatives to one another. Here I will give a theory of all-things-considered epistemic value that produces an *ordinal ranking*, as follows:⁸⁰

Objective Epistemic Consequentialism:

- ii. theory of all-things-considered value (for evaluative focal point {credences}): If c_1 and c_2 are credences or sets of credences that can be in the role of being adopted by agent S at time t, and c_1 and c_2 are alternatives to one another, then c_1 has more all-things-considered value than c_2 iff [the safety Brier score of the credences that will be brought about if S adopts c_1 at t (including c_1 itself)] is lower than [the safety Brier score of the credences that will be brought about if S adopts c_2 at t (including c_2 itself)] (where the degree of value of a safety Brier scores is proportional to the degree to which it approaches 0).
 - theory of all-things-considered value (for evaluative focal point {decision procedures}): If d_1 and d_2 are decision procedures that can be in the role of being committed to by agent S at time t, and d_1 and d_2 are alternatives to one another, then d_1 has more all-things-considered value than d_2 iff [the safety Brier score of the credences that will be brought about if S commits to d_1 at t] is lower than [the safety Brier score of the credences that will be brought about if S commits to d_2 at t] (where the degree of value of a safety Brier scores is proportional to the degree to which it approaches 0).

This epistemic consequentialist theory of all-things-considered value could, of course, be used to assess things other than credences and decision procedures and in roles other than 'being adopted' and 'being committed to'. However, since I will primarily be concerned with these two evaluative focal points for the remainder of this thesis, I will restrict myself to the account given above in what follows. I will now consider an objection to this account

⁷⁹I therefore reject the 'separateness of propositions' intuition expressed by Berker (2010) and others. This roughly states that agent necessarily does a wrong thing epistemically by adopting a credence function that has a poor safety Brier score in a single proposition, even her doing so produces a much better safety Brier score with respect to a larger set of propositions.

 $^{^{80}}$ I do so to avoid the criticisms of cardinal accounts that have been given elsewhere. I leave open the possibility that a cardinal theory could be constructed from which one could derive this ordinal account.

of all-things-considered value, before turning the final component of objective epistemic consequentialism: that is, its *deontic theory*.

2.1.3 An Epistemic Deontic Theory

The deontic component of objective epistemic consequentialism will be the simplest to outline, since it builds on much of the discussion that has taken place in the two previous sections. Let us begin, as we did above, with a reminder of the key questions that were raised for an epistemic consequentialist deontic theory in chapter 1:

- Which deontic properties will an epistemic consequentialist deontic theory assign to members of evaluative focal points on the basis of their all-things-considered value?
- How does the epistemic ought relate to different deontic properties?
- Will the decision procedure recommended by epistemic consequentialism be identical to the criterion of rightness given by its deontic theory?

I will divide this subsection into two parts. In the first part, I answer the first two questions above. I will give both the deontic properties epistemic consequentialism assigns to different EFP members, and the relation of this to what an agent ought to do. In the second part, I will answer the third question. Here I will re-emphasise what has already been stated above: that the criterion of rightness of objective epistemic consequentialism is very different from the decision procedures it will recommend that agents ought to commit to.

Deontic Properties and Oughts

In section 1.2.3 I focussed primarily on the deontic properties of *rightness* and *wrongness*, where I defended a gradational account of both of these properties. However, I argued in the previous section that when comparing the all-things-considered value of more than one EFP member in a role, we should give an ordinal rather than a cardinal ranking of the relevant EFP members. Can we retain both a gradational account of rightness and a non-gradational account of all-things-considered value? Moreover, can we do so while retaining some of the intuitiveness of the gradational account as it was outlined above?

I propose that we are able to retain the intuitiveness of being able to assign different acts different degrees of rightness. This doesn't mean, however, that we need to construct a gradational theory of all-things-considered value, or use gradational deontic properties with respect to what an agent *ought* to do. We should instead treat rightness as a means of *assessing* an EFP member, and treat *optimality* as that which governs which EFP members an agent ought or ought not to actualise. To see how we can jointly use rightness as a means of assessing an EFP member, and optimality as a means of determining what an agent ought to do, consider the following case. Gertrude is an investigative journalist who has two main sources of information in a government department: A and B. A is very well-informed and B is somewhat well-informed. Gertrude wants to use her sources to check on a rumour she has heard about. She has a credence of 0.5 in the rumour before she approaches either informant. If A thinks the rumour is true (false), Gertrude will raise (lower) her credence by 0.3, and if B thinks the rumour is true (false), Gertrude will raise (lower) her credence by 0.1. We can represent this as follows, where \checkmark means the informant things the rumour is true, and \bigstar means the informant thinks the rumour is false, and neither informant withholds judgment about the rumour:

A (± 0.3)	B (± 0.1)	G's credence in <i>rumour</i>
1	1	0.9
\checkmark	×	0.7
X	1	0.3
×	×	0.1

Now suppose that A says that the rumour is true and B says that the rumour is false. As such, Gertrude adopts a credence of 0.7 in the rumour. But suppose that Gertrude has been wantonly reckless in placing any confidence in B, since he is obviously a mole placed by the government to undermine the confidence of reporters in true rumours and increase their confidence in false rumours. Before checking the with her sources, there were four simple decision procedures that Gertrude could have committed herself to:

- DP_1 : Place some confidence in both A and B
- DP_2 : Place some confidence in. A and place no confidence in B
- DP_3 : Place some confidence in B and no confidence in A
- DP_4 : Place no confidence in either A or B

Let us assume that if Gertrude has confidence in A or B in these decision procedures, then she has the same confidence in them that she did the case above. If she has no confidence in A or B in these decision procedures, then she ignores their view completely. As such, Gertrude committed herself to DP_1 in the above case. Let us assume that A and B are both close to the source of all rumours, such that if a rumour about their department is true or false then they will both know it. We can now table the results were she to adopt each of these decision procedures:

	А	В	DP_1	DP_2	DP_3	DP_4
	1	1	0.9	0.8	0.6	0.5
if R is true	1	X	0.7	0.8	0.4	0.5
if R is FALSE	X	1	0.3	0.2	0.6	0.5
	X	X	0.1	0.2	0.4	0.5

Here I have highlighted what while *actually* happen if the rumour R is true and if the rumour R is false in rows (2) and (4). We know that these will be the actual results because of the relevant facts about Gertrude and her sources. Given the argument of the previous section, these are the *only* two rows that are relevant to the actualist epistemic consequentialist's assessment of Gertrude's credences. Let us assume that the consequence in which the rumour is false is sufficiently similar to the actual consequence (in which it is true) to be included in the safety Brier score of Gertrude's actual credences. We can use this information to assess DP_1 to DP_4 using the safety Brier score from section 2.1.1:⁸¹

$$SBS(DP_1) = \frac{1}{2}((0.7 - 1)^2 + (0.3 - 0)^2) = 0.09$$
$$SBS(DP_2) = \frac{1}{2}((0.8 - 1)^2 + (0.2 - 0)^2) = 0.04$$
$$SBS(DP_3) = \frac{1}{2}((0.4 - 1)^2 + (0.6 - 0)^2) = 0.36$$
$$SBS(DP_4) = \frac{1}{2}((0.5 - 1)^2 + (0.5 - 0)^2) = 0.25$$

Gertrude's safety Brier score is better if her credences are safely closer to 1 when the relevant proposition R is true, and safely closer to 0 when the relevant proposition R is false. It should be clear given the above calculations that were we to rank these decision procedures from best to worst, we would find that $DP_2 > DP_1 > DP_4 > DP_3$.

How does this relate to the *rightness* and the *optimality* of these decision procedures? Well, DP_2 is the *optimal* decision procedure for Gertrude to follow, since there is no decision procedure that produces a credal distribution with a better safety Brier score. Since I

⁸¹Given the reciprocity in this case we could also formulate it using the standard Brier score.

have indicated that optimality should govern what an agent *ought* to do, this means that Gertrude ought to have committed to DP_2 and ought not to have committed to any of DP_1, DP_3 or DP_4 . But although Gertrude has done something she ought not to do in committing to DP_2 instead of DP_1 , this is clearly not as bad as committing to either DP_3 or DP_4 . This fact can be captured by our gradational assessment of the rightness of these decision procedures. Suppose that a decision procedure is right to degree n, where n is the safety Brier score of the credences it produces. Given this, the degree of rightness of a decision procedure increases as n approaches a perfect 0, and decreases as n recedes from 0. Therefore, despite the fact that Gertrude ought not to commit to D_1 , committing to this decision procedure still has a 'rightness' score of 0.09, which is better than D_3 by 0.27 degrees of rightness and is better than D_4 by 0.16 degrees of rightness.

It may be objected that this strips 'rightness' and 'wrongness' of there evaluative content. But this will be true of any theory that assigns uses its outright deontic concepts (like 'optimal' and 'non-optimal') to determine what an agent ought to do, and its gradational deontic concepts (like 'rightness' and 'wrongness') to offer a more fine-grained assessment of EFP members in a given role. There is certainly more that could be said about such a theory, and in defence of some of its more controversial assumptions,⁸² but the minimal version of it given here is all that can be given in a thesis of this size.

I will now consider the relation between the criterion of rightness of objective epistemic consequentialism, and the decision procedures it recommends that agents commit themselves to. After this, I will outline the deontic theory of epistemic consequentialism.

Distinguishing the Criterion of Rightness from Decision Procedures

A criterion of rightness is the thing that we use to assess how right wrong a given EFP member.⁸³ A decision procedure is a set of heuristics or rules that an agent uses to decide what to do or believe. Under actualist epistemic consequentialism, the criterion of rightness will not necessarily be the best decision procedure for an agent to follow, indeed it may never be the best decision procedure to follow. Consider the following two criteria:

⁸²More specifically, a fuller account would need to deal with the objections from Norcross (1997) and Ord (pg. 39-40, 2009) to fine-grained assessments of EFP members in a given role. These criticisms seem to apply regardless of whether or not these fine-grained assessments have any normative import.

⁸³Given the deontic theory of this chapter, the term *criterion of optimality* may be more appropriate.

- COR_1 : Adopt credences such that you maximise epistemic utility.
- COR_2 : Adopt credences such that you maximise expected epistemic utility.

The first of these is the criterion of rightness for the objective ought, while the second is a criterion of rightness for the subjective ought, as I will discuss more in section 2.2. But will either of these maximise epistemic utility if they are employed as a decision procedure? The answer to this depends on the role in which we are assessing COR_1 and COR_2 when we ask whether they are the best decision procedure to 'employ'. Suppose we are assessing them in the role of 'being executed perfectly'. COR_1 and COR_2 may be the decision procedure an agent ought to execute perfectly, since actually executing them perfectly will necessarily maximise epistemic utility (though it may be almost impossible for agents like ourselves to do so). But suppose we are assessing them in the role of 'being committed to'. In this case it is clear that they will be much less effective than a decision procedure that an agent is actually able to comply with given her cognitive limitations, such as:

 DP_1 : Update your credences by conditionalising on your total evidence.

 DP_2 : If you believe you perceive that P, increase your credence that P.

Now we may not *always* be able to execute these decision procedures correctly, but committing to them will doubtlessly lead to more epistemic utility than only committing to COR_1 or COR_2 . Most obviously, the calculations involved in COR_1 and COR_2 are sometimes extremely complicated. Not only will they be time-consuming, but they have a high risk of error. If such actual errors are made by agents that commit to them, then they will turn out to be worse than simpler decision procedures like DP_2 , which are easier to comply with without agents like ourselves making too many mistakes in the process. For example, imagine two agents, James and Joan. James commits only to COR_2 while Joan commits only to DP_2 . James and Joan are both fairly poor at making calculations, so when James tries to work out what to believe, he gets it wrong about 90% of the time and to a degree that varies quite widely. Joan, on the other hand, would rarely be assessed as perfect under the criterion of rightness. However, Joan approximates the safe credence to have in a proposition about her perceptions about 80% of the time, and where she varies she varies to a smaller degree than James. It should be clear that Joan is much more likely to produce more actual epistemic value than James when it comes to adopting credences in propositions about their perceptions. As such, under the epistemic criterion of rightness, James and Joan should both commit to DP_2 rather than COR_2 .⁸⁴

I will provide another example in chapter 3 of a decision procedure which, like COR_1 and COR_2 , is a superior decision procedure in the role of 'being executed perfectly', but it yields worse results when agents commit to it in certain cases involving higher-order evidence. It is very important to recognise that the criterion of rightness, even in its subjective form, will not always be the best decision procedure for an agent to adopt. Given this, we can now provide a formulation of the deontic theory of epistemic consequentialism.

An Epistemic Consequentialist Deontic Theory

Given the above discussion, we are now able to construct an objective epistemic consequentialist deontic theory for the deontic properties of *being right*:⁸⁵

Objective Epistemic Consequentialism:

- iii. deontic theory (for EFP {credences} and deontic property being right): Credence c adopted by agent S at time t is right to degree n, where n is the sum of the safety Brier scores produced by c.
 - deontic theory (for EFP {decision procedures} and deontic property being right): Decision procedure d adopted by agent S at time t is right to degree n, where n is the sum of the safety Brier scores produced by d.

As I argued above, a credence or decision procedure gets more right under epistemic consequentialism as as n approaches 0, and less right as n recedes from 0.

I have argued above that an agent *ought* to adopt the credence or commit to the decision procedure that is optimal, i.e. that produces the most all-things-considered value, and that she *ought not* to adopt any other credence or commit to any other decision procedure. The above deontic theory allows that sub-optimal credences and decision procedures are not all equally right under epistemic consequentialism, but it does not make explicit which EFP member an agent ought to actualise. As such, I will stipulate that an agent ought to actualise the EFP member that is optimal in a given role. This can be made explicit by adding the following two principles to the epistemic consequentialist deontic theory:

⁸⁴Parfit (pg. 40-3, 1984) refers to these as concerns about the *self-effacing* nature of consequentialism.

⁸⁵Why do I not also give an account of *being wrong*? Because it would be difficult to give an account of both rightness and wrongness here without depending on an overly arbitrary threshold at which a given safety Brier score moves from being of positive value to being of zero value to being of negative value.

Objective Epistemic Consequentialism:

- iii. deontic theory (for EFP {credences} and deontic property being optimal): Credence c adopted by agent S at time t is is optimal iff there is no other credence c', such that c' is an alternative to c at t, and adopting c' at t produces more actual all-things-considered value than adopting c at t.
 - deontic theory (for EFP {decision procedures} and deontic property being optimal): Decision procedure d committed to by agent S at time t is optimal iff there is no other decision procedure d', such that d' is an alternative to d at t, and committing to d' at t produces more actual all-things-considered value than committing to d at t.

In the next section I will briefly consider the nature of the *ought* within this epistemic consequentialist deontic theory. More specifically, I will assess whether the ought that epistemic consequentialists should focus on is the *subjective* or *objective* epistemic ought.

2.2 Objective and Subjective Epistemic Oughts

In this section I will focus solely on the question raised in the last section of chapter 1. As a reminder, the question that was raised in section 1.2.3 was:

• Is the epistemic ought objective or subjective, or are there both objective and subjective epistemic oughts?

A good way to answer this question is to outline an epistemic version of a Jackson-case.⁸⁶ These cases are standardly used to elicit the difference between objective and subjective oughts by presenting us with a situation in which what we *ought* to do in the objective sense is not the same as what we *ought* to do in the subjective sense. Consider the following:

Alex is not sure what credence to place in the proposition that his team will win. He knows that they will either win or lose the game, but given their track record and the track record of the opposing team, the pundits all say that there is only a 1 in 2 chance that his team will win: perfect evens. Alex can either place credence 1 that they will win, credence 0 that they will win, or credence 0.5 that they will win. The first credence function is very good at capturing the facts if they win but very bad if they lose. The second credence function is very bad at capturing the facts if they win but very good if they lose. And the third credence function is not great at capturing the facts if they lose or if they win, but nor is it as bad as the worst results of the first two credence functions.

Let $Cr_A(W)$ be Alex's crendence that his team will win, and suppose that his credence that his team will lose is just $1 - Cr_A(W)$. Suppose that we are measuring the value of a

⁸⁶See Jackson (1991) and Parfit (1988, *unpublished*) for examples of these cases in ethics.

credal function using the *standard* Brier score. Then the value of the outcomes that can result from each of the three credence functions available to Alex is as follows:

	$Cr_A(W) = 1$	$Cr_A(W) = 0$	$Cr_A(W) = 0.5$
Team Wins $(W = 1)$	$BS = 1(1-1)^2 = 0$	$BS = 1(0-1)^2 = 1$	$BS = 1(0.5 - 1)^2 = 0.25$
Team Loses $(W = 0)$	$BS = 1(1-0)^2 = 1$	$BS = 1(0-0)^2 = 0$	$BS = 1(0.5 - 0)^2 = 0.25$

Clearly the most epistemic value to be gained is if Alex adopts credence 1 in his team winning and they do win, or if he adopts credence 0 in his team losing and they do lose. After all, both of these options result in Alex having a perfect standard Brier score of 0 over this single case. But the most *expected* epistemic value is gained if Alex adopts credence 0.5 in his team winning, since this way he will never get things disastrously wrong, even though he is guaranteed to never have a perfect Brier score regardless of whether his team wins or loses.⁸⁷ So which credence function should Alex adopt?

The answer that has been given by ethicists is that, *objectively*, Alex ought to maximise actual utility. Therefore, if his team *will* win, then he ought to adopt credence 1 in W, and if his team *will* lose then he ought to adopt credence 0 in that proposition. After all, these are the two options that clearly maximise actual epistemic utility. To get a sense of the objective ought, imagine Alex's friend Raj has inside knowledge that Alex's team has been paid to throw the game, and that they definitely *will* lose tomorrow. When Alex asks Raj which credence he, Alex, ought to adopt, Raj seems to respond truly when he says 'you ought to adopt credence 1 that your team will lose'. What Raj has said is true despite the fact that Raj knows that Alex has no idea about the match fixing. So Raj can't mean that, *given Alex's beliefs or evidence*, Alex ought to adopt credence 0 in W. Instead Raj presumably means that *given all of the facts* (including those facts that Alex doesn't have access to but Raj does) Alex ought to adopt credence 0 in W.⁸⁸ This fact-relative ought is what I will refer to as the *objective ought*.

But something seems wrong with saying that Alex ought to adopt either credence 1 in W or credence 0 in W depending on whether his team will win or lose. After all, this has less expected value than adopting a credence of 0.5 in W. It is standardly argued that this is because there is another, subjective sense of ought under which Alex *ought* to adopt credence 0.5 in W, since he ought to maximise the expected value relative to the facts he

⁸⁷If there is a 1 in 2 chance that his team will win, then the *expected* epistemic value of Alex adopting credence 1 (or 0) in W is 0.5 (BS= $\frac{1}{2}((1-1)^2 + (1-0)^2))$, which is worse than an expected score of 0.25.

⁸⁸There have been attempts to cash out the semantics of this exchange in a non-objective manner, such that Raj is saying 'relative to my (Raj's) evidence, you (Alex) ought to adopt credence 0 in W'. I will not explore the plausibility of such accounts here. For more on the topic see Kolodny and MacFarlane (2010).

has access to. There are two common but very different ways of defining the subjective ought. The first is that it is what an agent subjectively ought to do relative to her *beliefs*, and the second is that an agent subjectively ought to do what is relative to her *evidence*. To see the difference between these two accounts of the subjective ought, suppose that an agent necessarily has access to the beliefs that she holds.⁸⁹ This would mean that what the agent subjectively ought to do would be *luminous* to her: i.e. she would always have access to what is required of her given her beliefs and her . This means that if a subject knew that she ought to maximise expected epistemic utility, then she could never knowingly and without negligence do what it is subjectively wrong for her to do. After all, she has luminous access to all of the information she needs to satisfy the belief-relative ought. But it has been argued extensively within epistemology that an agent does not necessarily have access to what her evidence is. ⁹⁰ Given this, even if an agent knew that she ought to maximise expected epistemic utility, she might not know how to go about doing this, since she doesn't know what her evidence is. As such, she doesn't know what the evidence-relative ought requires of her. Williamson has expressed such a view of epistemic rationality:⁹¹

'The standard conception of rationality depends on a distinction between the aims and methods of cognitive activity. On that conception, truth is an aim. We cannot attain it directly; we cannot follow the rule 'Believe truly!' when we do not know what is true. Therefore we must use methods to reach the truth. Rationality is a method. We can follow rules of rationality because we are always in a position to know what they require. If the argument of section 8.6 is correct, this picture of rationality is mistaken. Just as one cannot always know what one's evidence is, so one cannot always know what rationality requires of one. Just like evidence, the requirements of rationality can differ between indiscriminable situations. Rationality may be a matter of doing the best one can with what one has, but one cannot always know what one has, or whether one has done the best one can with it.'

But does the objective epistemic consequentialism outlined above need to appeal to either the belief-relative or evidence-relative subjective ought? I will argue that there are two reasons we do not need to appeal to either form of the subjective ought.

⁸⁹Though even this thesis is controversial, especially when we are employing sharp credences!

⁹⁰Here I assume a Williamsonian account of evidence, such that Evidence = Knowledge, and even if one knows that P, one does not necessarily know that one knows that P. See Ch. 4 & 9, Williamson (2000).

 $^{^{91}}$ Pg. 179, Williamson (2000). Williamson himself goes on to say that '[w]e can use something as a method in contexts in which one is usually in a position to know whether one is complying with it...in that sense we can use even believing truly as a method in contexts in which on is usually in a position to know what is true'. While I agree with the content of this statement, I would argue that whether or not an agent can *usually* know how to comply with a method is not important (for reasons that are outlined in pg. 36-7, Ord (2009)). What is important? The fact that we find ourselves in a world in which the methods an agent usually knows how to comply with are those that produce the most actual epistemic utility.

The first reason that the epistemic consequentialist doesn't need to appeal to a subjective ought, is that she treats good scores on the *safety* Brier score as being of contributory final value. This dissolves the problem of the epistemic Jackson case, since the *actual* value Alex will produce by adopting credence 1 in W is the sum of his Brier scores at sufficiently similar consequences. But even if his team wins in the actual consequence, there is a sufficiently close consequence in which they lose and in which Alex has credence 1 in W. Therefore the safety Brier score of the three credence functions Alex can adopt are:

	$Cr_A($	(W) = 1	$Cr_A(W)$) = 0
W = 1	$SBS = \frac{1}{2}((1-1))$	$)^{2} + (1 - 0)^{2}) = 0.5$	$SBS = \frac{1}{2}((0-1)^2 +$	$(0-0)^2) = 0.5$
W = 0	$SBS = \frac{1}{2}((1-0))$	$)^{2} + (1-1)^{2}) = 0.5$	$SBS = \frac{1}{2}((0-0)^2 +$	$(0-1)^2) = 0.5$
		$Cr_A(1)$	W) = 0.5	
	W = 1	$SBS = \frac{1}{2}((0.5 - 1))$	$(0.5-0)^2 = 0.25$	-
	W = 0	$SBS = \frac{1}{2}((0.5 - 0))$	$(0.5-1)^2 = 0.25$	-

In this case, the safety Brier score punishes Alex for using a method that recommends extreme credences in a proposition which has a 1 in 2 chance of being false. Therefore, even in the cases where he gets lucky and has a credence 1 (0) in a true (false) proposition, his unsafe credences are not of greater epistemic value than having adopted the safe credence of 0.5 in the proposition that his team will win. By employing the safety Brier score, the epistemic consequentialist is able to appease our intuitions about what Alex ought to do within the framework of the *objective* epistemic ought. Alex objectively ought to maximise safe credences in propositions, and the only way he can do this in the above case is by adopting a credence of 0.5 in the proposition that his team will win.

The second reason that the epistemic consequentialist doesn't need to appeal to the subjective ought, in either its belief-relative or evidence-relative form, is that much of the work of the subjective ought can be done by the global consequentialist component of the theory I have described. Imagine a case in which Alex objectively ought to adopt credence 1 in a proposition, because this will actually produce a better safety Brier score that would be produced by adopting any other credence in P. But the best decision procedure for Alex to commit to is one that recommends he adopt credence 0.7 in the relevant proposition. In such cases, the epistemic consequentialist can appeal to the discussion of the previous section. Alex *objectively* ought to adopt credence 1 in the proposition, but he also *objectively* ought to commit to a decision procedure that recommends not to adopt credence 1 in the proposition. Thus he objectively ought to continue to commit to the decision procedure,

without actually complying with it in this specific case. But if Alex does comply with the decision procedure, however, we still have an intuition that he has done *something* right. The epistemic consequentialist can reply that he *has* in fact done something right: he has committed himself to a decision procedure that produces the greatest epistemic utility over a series of cases, even though it recommends a sub-optimal credence in this individual case. We also have an intuition that Alex has done *something* wrong in cases like this. The epistemic consequentialist can reply that he *has* in fact done something right: he has adopted the sub-optimal credence 0.7 in P instead of the optimal credence 1 in P.

For both of these reasons, I propose that epistemic consequentialism is best understood as a form of *objective* consequentialism. This means that what an agent ought to do is relative to all of the facts in the world, which include, but are not limited to, facts about her beliefs and the evidence she has.⁹² The fact that epistemic consequentialism is able to employ only one 'ought' while still appeasing our intuitions about what an agent ought to do when she is faced with an epistemic Jackson case is a theoretical virtue, since it makes for a more parsimonious account of 'oughts'. This doesn't mean, however, that there won't be further cases that problematise a solely objective account of epistemic consequentialism. Considering such cases is beyond the scope of this thesis. However, given that epistemic consequentialism can do without subjective oughts in a case as seminal as the Jackson case, it is not implausible to suppose that it might be able to resolve other cases used to support the need for a subjective ought in a similar manner.

I have now outlined the three primary components of epistemic consequentialism: its theory of contributory final value, its theory of all-things-considered value, and its deontic theory. I have also outlined the nature of the ought that will be assumed in this epistemic consequentialist theory. The next chapter will provide additional support for the epistemic consequentialist theory I have outlined here by applying it to cases of peer disagreement.

 $^{^{92}}$ This does not mean that there is *no* subjective ought within epistemic consequentialism. I am merely arguing that there is no need for a subjective ought in the account of epistemic consequentialism given here.

3 An Application of Objective Epistemic Consequentialism

In this chapter I will consider the application of objective epistemic consequentialism to a problem in contemporary epistemology. The problem I will consider is that of *peer disagreement*, where I argue that epistemic consequentialism can both explain the intuitions motivating both conciliatory and non-conciliatory responses to this problem, and recommend what an agent ought to do when she is faced with a case of peer disagreement.

3.1 The Problem of Peer Disagreement

Suppose that you watched a football match two years ago and, on the basis of your memory of the game, you have high credence in the proposition that Argentina beat Mexico in that game. Now suppose that you meet with one of your colleagues, who watched that very match with you, and who is equally good at remembering such things. But when you mention the match she insists that Mexico bear Argentina. What should now happen to your credence that Argentina beat Mexico? The crucial question here is how much weight you ought to give to 'higher-order' evidence. The fact that your peer disagrees with you gives you 'higher-order evidence', because it brings into question whether your credence that Argentina beat Mexico is actually supported by your evidence about the game.

It might seem like you ought to considerably reduce your credence that Argentina beat Mexico when faced with peer disagreement. Views that support this intuition are often called *conciliatory views*. Roughly speaking, conciliatory views claim that you ought to give higher-order evidence a great deal of weight. In contrast, *non-conciliatory views* claim that you ought to give higher-order evidence almost no weight at all.

Since the literature on peer disagreement can be technical at times, this chapter will necessarily be more technical than the ones that proceeded it. For example, for the purposes of this chapter I must distinguish between the credence that a subject S actually has in P at time t_i , which will be written as $ActCr_S^i(P) = n$, and the 'evidentially supported' credence for S to have in P is n at time t_i , which will be written as $EvCr_S^i(P) = n$. Here a credence is evidentially supported iff it the the subject adopting it has complied with the decision procedure: 'update correctly on your total evidence by conditionalisation'.

Consider the following case of peer disagreement: two subjects, Ann and Beth, are the top-performing members of their school's history quiz team, having virtually identical track records when it comes to remembering historical events with accuracy. This is because Ann and Beth have the roughly the same evidence with respect to historical questions, and Ann and Beth are equally likely to have updated correctly on that evidence in any given case. They both occasionally make mistakes, but neither makes mistakes more often than the other. At one competition, Ann and Beth are asked in what year the English Peasants' Revolt occurred. The question is presented alongside the year '1381', which can either be right or wrong. Ann and Beth buzz in for their team at exactly the same time. The host gives them a moment to confer. Ann says to Beth 'I have a credence of 0.8 that '1381' is the answer', but Beth responds to Ann 'I have a credence of 0.8 that '1391' is the answer'.⁹³ In such a situation, what should happen to Ann and Beth's credences in the respective propositions, P: '1381 is the right answer', and $\neg P$: '1381 is the wrong answer'?

Within the objective Bayesian framework I have assumed in this thesis, there are three potential causes of credal disparity between Ann and Beth. These are:

- (1) Ann and Beth have different sets of priors.⁹⁴
- (2) Either Ann, or Beth, or both Ann and Beth have updated incorrectly on their evidence.
- (3) Ann and Beth have different evidence with respect to the relevant proposition.

In cases of peer disagreement, (1) and (3) are excluded from being possible explanations of Ann and Beth's credal disparity. This is because it is assumed that Ann and Beth both have the rational set of priors, and that Ann and Beth have the same P-relevant evidence. However, although this means that there must have been a updating error on the part of one or both of the subjects with different credences in the relevant proposition, if they know that they are equals with respect to (1) and (3) then it is no more probable for Ann or Beth that Ann as made an updating error than it is that Beth has made an updating error.⁹⁵

According to the equal weight view of peer disagreement developed by Elga (2007), if Ann and Beth both know prior to the disagreement that they are peers, then conditional upon them finding out that they disagree about P, they ought to split their respective

⁹³For simplicity, suppose there is no phenomenal intermediary of these propositions for Ann and Beth, such that neither has 'it appears that 1381 is the answer' or 'it appears that 1391 is the answer' as part of their evidence. Instead, their credences in the respective answers arose directly from their updating on their relevant background evidence about English history.

 $^{^{94}}$ Given the constraints on priors outlined in section 2.1.1, one of these sets at most can be rational.

 $^{^{95}}$ Note that this means that in cases of peer disagreement, the rational credence for both Ann and Beth to have in the relevant proposition before they compare their credences will necessarily be identical given the exclusion of (3) as the cause of their credal disparity.

credences in P. For the purposes of this thesis, the equal weight view can be represented as requiring the straight averaging of Ann and Beth's credences in the case of peer disagreement described above.⁹⁶ Under this straight averaging interpretation of the equal weight view, the rational credence for Ann (A) and Beth (B) to adopt in P at t_1 (after the disagreement), conditional on Ann and Beth knowing that they are epistemic peers, is:

$$EvCr_A^1(P) = EvCr_B^1(P) = \frac{ActCr_A^0(P) + ActCr_B^0(P)}{2} = \frac{0.8 + 0.2}{2} = 0.5$$

But is it true under an *externalist* account of evidence that it is just as likely on Ann's evidence that she is mistaken about the answer to the question about the Peasants' Revolt? Externalism about evidence is, roughly speaking, a rejection of any conception of first-order evidence that stipulates that a subject's evidence is always exhausted by her non-factive mental states, such as her beliefs.⁹⁷ For example, suppose that Ann and Beth are 'internal twins' iff Ann and Beth are identical with respect to their non-factive mental states. Internalism about evidence stipulates that if two subjects are internal twins, then they have the same evidence. We can therefore construe the externalist reciprocal of this as follows:⁹⁸

Evidential Externalism: It is possible that Ann and Beth are internal twins and Ann and Beth do not have the same evidence.

Adopting externalism about evidence will affect what is considered to be the 'evidentially supported' credence. Given this account of evidence, it is possible that Ann and Beth are internal twins, such that they have all of the same non-factive mental states. And yet while n is the evidentially supported credence for Ann to have in P, n is not the evidentially supported credence for Ann and Beth have different evidence.

Now, as a matter of fact, Ann is correct in this case, since the Peasants' Revolt did indeed occur in 1381. Let H be a set of the relevant first-order evidence that both Ann and Beth have about the historical events of England in the 1300's. Suppose also that $EvCr_A^0(P|H) = EvCr_B^0(P|H) = 0.8$. This means that 0.8 is the credence in P that is evidentially supported by Ann and Beth's evidence about history H at t_0 (that is, before

⁹⁶This straight averaging account of the equal weight view is considered and rejected in pg. 283-5, Jehle & Fitelson (2009). Further criticisms of this account are considered in Wilson (2010).

⁹⁷For my use of non-factive mental states, see Wedgwood (2002).

⁹⁸Here I follow the account given in pg. 376-8, Silins (2005).

the disagreement). This is the credence that Ann did in fact adopt in P at t_0 , but Beth mistakenly adopted a credence of 0.2 in P at t_0 because she updated incorrectly on H. According to the externalist about evidence, Ann and Beth's evidence for and against P at t_1 (that is, after they disagree and compare credences), consists of the following:⁹⁹

Evidence that
$$P$$
 Evidence that $\neg P$
 H
 $ActCr^{0}_{A}(P) = 0.8$ $ActCr^{0}_{B}(\neg P) = 0.8$

Under most accounts of evidentiary externalism, Ann need not be in a position to know that her first-order evidence consists of H and not some $\neg P$ -supporting evidence, H'.¹⁰⁰ As such, even if Ann does not know which column of this table her historical evidence falls under, her evidence still consists of the P-supporting H, and it is *this* that she is rationally required to update on for her credences to be 'evidentially supported'.

I propose that if Ann actually has her historical evidence H and the proposition 'Beth and Ann are peers' (or Peers_{A,B}) as part of her evidence, then this evidence (H and Peers_{A,B}) will fully screen the positive correlation between Beth's credence in P before the disagreement ($ActCr_B^0(\neg P) = 0.8$) and the proposition $\neg P$ for Ann. A piece of evidence 'fully screens' another in relation to a given proposition by nullifying the correlation that normally exists between that piece of evidence and the proposition in question.¹⁰¹ Therefore, Ann's historical evidence H fully screens the positive correlation between $ActCr_B^0(\neg P) = 0.8$ and $\neg P$ for Ann iff the following are rational priors:¹⁰²

(I)
$$EvCr_A(\neg P|ActCr_B^0(\neg P) = 0.8) > EvCr_A(\neg P)$$
, and
(II) $EvCr_A(\neg P|(ActCr_B^0(\neg P) = 0.8) \land (H \land Peers_{A,B})) = EvCr_B(\neg P|(H \land Peers_{A,B}))$

But why would Ann's evidence H and $\operatorname{Peers}_{A,B}$ at t_0 fully screen the positive correlation between her evidence at t_1 that Beth has a high credence in $\neg P$, and $\neg P$? To see why this is the case, note that since Ann has $\operatorname{Peers}_{A,B}$ as part of her evidence, she knows (or

⁹⁹I am indebted to Weatherson's (brian.weatherson.org/JSE.pdf) tabling of evidence in this manner.

¹⁰⁰For example, if the externalist accepts the E=K thesis, which equates one's evidence with one's knowledge, then E-Access will fail by standard arguments against the KK thesis of ch. 5, Williamson (2000). ¹⁰¹The term is adopted from Weatherson (2010), who in turn adopts it from Reichenbach (1956).

¹⁰²In this formulation of (I) and (II) I assume for the sake of brevity that the rational prior credence in P is less than 0.8. A more general formulation would (a) generalize (I) and (II) to any subjects S and S', where $S \neq S'$, and (b) stipulate that S''s credence in P is greater than S's prior credence in P.

has as part of her evidence) that Beth doesn't have any P-relevant evidence that she, Ann, doesn't have. Now combine this with the fact that the rational credence to have in P on Ann's first-order evidence H is 0.8, and H is part of Ann's evidence. What is the evidential import of Beth's high credence in $\neg P$ for Ann? Surely all that Beth's high credence in $\neg P$ is evidence of (for Ann) is that one of the following disjuncts is true: either Beth has irrational priors, or Beth has incorrectly updated on the relevant first-order evidence H.¹⁰³ But neither of these disjuncts should result in any increase in Ann's credence in $\neg P$ at t_1 . Therefore Ann's evidence H and Peers_{A,B} fully screens the positive correlation that would otherwise exist between $ActCr_B^0(\neg P) = 0.8$ and $\neg P$. Given (I) and (II) this means that the rational credence for Ann to have in P at t_1 is identical to the rational credence for her to have in P at t_0 before she knew anything about Beth's credences: namely, exactly 0.8.

Christensen has responded to such worries for conciliatory views of peer disagreement by proposing the following about higher-order evidence [HOE]. In this case, the 'higher-order evidence' is Ann's evidence that she and Beth are peers, and that she and Beth disagree about the relevant proposition P:

'In accounting for the HOE ... I must in some sense, and to at least some extent, put aside or bracket my original reasons for my answer. In a sense, I am barred from giving a certain part of my evidence its due.'¹⁰⁴

In the case under discussion, the original reasons for Ann and Beth to adopt their respective credences in P was primarily their first-order evidence H. So Christensen advocates the exclusion of some of the relevant first-order evidence, such as H, from the set of total evidence relevant to the credence that it is 'evidentially supported' for Ann and Beth to have in P at t_1 . I will refer to the practice of excluding first-order evidence in peer disagreement cases as the 'bracketing' of first-order evidence.¹⁰⁵ In defence of bracketing, Christensen outlines several higher-order evidence cases that are, in all their essentials, not dissimilar from the peer disagreement case given above. He then argues that we should be unwilling to bet heavily on our credences in such cases.¹⁰⁶ The real question facing us: is bracketing the right thing to do *epistemically*? In the next section, I will show that

 $^{^{103}}$ This is the case even if Ann doesn't have, as part of her evidence, the proposition that her credence in P on her relevant evidence H was evidentially supported at t_0 .

¹⁰⁴Pg. 13, Christensen (2010) [this reference is to the online version].

 $^{^{105}}$ Bracketing is extremely similar to Elga's (2007) stipulation that equal weight view is conditional on the probability you would place in your being right *prior* to your finding our about the disagreement.

¹⁰⁶Pg. 195, Christensen (2010).

the safety Brier score that is produced by agents that commit themselves to the decision procedure, 'update correctly on your total evidence by conditionalisation in cases of peer disagreement' is worse than the safety Brier score that is produced by agents that commit themselves to the alternative decision procedure 'bracket your relevant first-order evidence in cases of peer disagreement'. Given this, I argue that it is the latter decision procedure that agents should commit themselves to in order to maximise epistemic utility.

3.2 An Epistemic Consequentialist Solution

Here I will argue that the machinery of epistemic consequentialism is able to resolve cases of peer disagreement like the one presented above. It can do so because it enables us to choose the best decision procedure for maximising good safety Brier scores. To see how it does this, suppose that we ask Ann to commit herself to the following decision procedure:

Non-Conciliatory (NC): If your credence in proposition P is evidentially supported given your total evidence (including H), then do not alter your credence in P when faced with peer disagreement. If your credence in proposition P is not evidentially supported given total evidence (including H) then adopt the credence that is evidentially supported.¹⁰⁷

Since Ann's total evidence includes both H and the fact that she and Beth are peers, in order to comply with decision procedure NC, Ann must adopt either a high (0.8) or a low (0.2) credence in sufficiently similar cases to the one presented above. Given the full screening argument of the previous section, if Ann has H and the fact that she and Beth are peers as part of her evidence, then she should only ever adopt the evidentially supported credence in P when she is complying with NC, since her first-order evidence screens the evidential import of the credence of whoever updated incorrectly. But there will presumably be sufficiently similar consequences in which '1382' was the year presented after the question. We can suppose that Ann's method of recalling 14th Century English history is not infallible, and that she would also have a credence 0.8 that this proposition is true, while Beth would retain her now evidentially supported credence of 0.2 that the year presented is the year in which the Peasants' Revolt occurred. In such cases, Ann can

 $^{^{107}}$ I refer to the evidence set prior to the disagreement. But the previous section has indicated that there are no grounds for peer disagreement to be a defeater of H in an objective Bayesian framework. Therefore it would not particularly matter if this were relative to Ann's evidence set before or after the disagreement.

only be said to be complying with NC if, upon finding out that she and Beth disagree, she switches her credence in the relevant proposition from 0.8 to 0.2. Now suppose we ask Beth to instead commit to the following decision procedure:

Conciliatory: When faces with peer disagreement about P, adopt the straight average of your credence and the credence of your peer in P.

This means that in the cases where Ann and Beth comply with NC and C respectively, Beth should have a credence of 0.5 in P and 0.5 in $\neg P$ after finding out about Ann's differing credence in P, whereas Ann should have a credence of 0.8 in P if she is right, and a credence in 0.2 in P if Beth is right. Which of these decision procedures is better?

In what follows I will need to distinguish between the better decision procedure to occupy the role 'being perfectly executed' and the better decision procedure to occupy the role of 'being committed to', as was outlined in section 2.1.3. The question then becomes: which of NC and C in are better *in these two different roles*?

For simplicity I will assume that the relevant year shown in the history competition is the correct year in exactly half of the cases, and that if the year shown is the correct one, then the proposition P that it is correct is such that $EvCr_A(P_n) = EvCr_B(P_n) = 0.8$. Similarly, if P is false then $EvCr_A(P_n) = EvCr_B(P_n) = 0.2$. I will also assume that since Ann and Beth are peers, they are each correct about P exactly half of the time and mistaken about P exactly half of the time. As above, Ann and Beth both access to what their higherorder evidence is (i.e. the fact that their peer disagrees with them) but not what all their first order evidence is, or which propositions it supports. More specifically, Ann and Beth don't have access to the fact that H is part of their evidence and not some similar set of propositions H'. Consider Ann's credences in P if she perfectly complies with NC in all sufficiently similar cases, which will here include c1 and c2:

> (c1) $EvCr_A(P|H) = 0.8$ and A adopts credence 0.8 in P at t_1 (c2) $EvCr_A(P|H) = 0.2$ and A adopts credence 0.2 in P at t_1

We can use the safety Brier score to determine the accuracy of A's credences in each proposition. Here I will calculate the safety Brier score of Ann's credences if NC is in the role of 'being perfectly executed' and 'being committed to', respectively. The former does
not take into account any epistemic failings on Ann's part. We can therefore predict that she will get it right in roughly half of the sufficiently similar consequences, and wrong in roughly half of the similar consequences. We can therefore work out the safety Brier score for NC in the role of 'being perfectly executed' by Ann as follows:

$$SBS = \frac{1}{N} \sum_{c \in \mathcal{L}}^{N} \underbrace{(0.8 - 1)^2}_{(c1)} + \underbrace{(0.2 - 0)^2}_{(c2)} = 0.04$$

As we can see, the safety Brier score of NC in the role of 'being perfectly executed' is extremely good. What about when we evaluate NC in the role of 'being committed to'? We must take into account facts about what Ann is like as an epistemic agent when evaluating decision procedures in the role of 'being committed to', and this will be reflected in Ann's safety Brier score. When Ann attempts to comply with NC, she does not have access to whether her credence in P is evidentially supported or not, since she doesn't know what her total evidence consists of. Ann also doesn't know who has correctly updated in the relevant peer disagreement cases, and has a 0.5 credence that it herself and not Beth. Since she has no extra evidence about which of them has updated correctly in a given case, she will presumably have to guess whether she has updated correctly or not if she is to try to comply with NC.We would expect that, in attempting to conform to NC, Anne would *correctly* guess that her credences were evidentially supported (or unsupported) roughly half the time, and to *incorrectly* guess that her credences were evidentially supported (or unsupported) roughly half the time, given her higher-order evidence. Suppose for simplicity that Ann guessed correctly in exactly half of the cases and incorrectly in exactly half of the cases. The resultant outcomes over sufficiently similar consequences would be as follows:

- (c1) $EvCr_A(P|H) = 0.8$ and A adopts credence 0.8 in P at t_1
- (c2) $EvCr_A(P|H) = 0.2$ and A adopts credence 0.2 in P at t_1
- (c3) $EvCr_A(P|H) = 0.8$ and A adopts credence 0.2 in P at t_1
- (c4) $EvCr_A(P|H) = 0.2$ and A adopts credence 0.8 in P at t_1

Therefore, the safety Brier score of NC in the tole of 'being committed to' by Ann is:

$$SBS = \frac{1}{N} \sum_{c \in \mathcal{L}}^{N} \underbrace{(0.8 - 1)^2}_{(c1)} + \underbrace{(0.2 - 0)^2}_{(c2)} + \underbrace{(0.2 - 1)^2}_{(c3)} + \underbrace{(0.8 - 0)^2}_{(c4)} = 0.34$$

This is clearly much worse than NC in the role of 'being perfectly executed', but how do both of these results fare against Beth's obedience to decision procedure C within each of these roles? We suppose that Beth also has access to the higher-order evidence about the disagreement, such that she knows she is in a case of peer disagreement. She is therefore aware that she must attempt to comply with decision procedure C. Since she also has the credal information about herself and Ann, she is in a position to average their credences in order determine the rational credence for her to have in P at t_1 according to decision procedure C. As such, decision procedure C in the role of 'being perfectly executed' and 'being committed to' will have the same or very close results in sufficiently similar consequences:

- (c5) $EvCr_B(P|H) = 0.8$ and B adopts credence 0.5 in P at t_1
- (c6) $EvCr_B(P|H) = 0.2$ and B adopts credence 0.5 in P at t_1

Consequently Beth's safety Brier score under decision procure C, both the role of 'being perfectly executed' and 'begin committed to, will be:

$$SBS = \frac{1}{N} \sum_{c \in \mathcal{L}}^{N} \underbrace{(0.5-1)^2}_{(c5)} + \underbrace{(0.5-0)^2}_{(c6)} = 0.25$$

Here we see an interesting result regarding the decision procedure NC and the decision procedure C. If we rank these two decision procedures by their safety Brier score in the role of 'being perfectly executed' and 'being committed to', we find the following:

Which decision procedure is better in the role of 'being perfectly executed'? [Decision Procedure NC in the role of being perfectly executed (SBS 0.04)] is epistemically better than

[Decision Procedure C in the role of being perfectly executed (0.25)]

Which decision procedure is better in the role of 'being committed to'? [Decision Procedure C in the role of being committed to (SBS 0.25)] is epistemically better than [Decision Procedure NC in the role of being committed to (0.34)]

Therefore the non-conciliatory decision procedure is better in the role of 'being perfectly executed', while the conciliatory decision procedure is better in the role of 'being committed to'. But if fallible epistemic agents like Ann and Beth attempt to perfectly execute decision procedure NC, they will not produce more actual epistemic utility than NC in the role of 'being committed to. After all, neither Ann nor beth know how to perfectly execute NC in the relevant cases! Therefore, the thing that will maximise epistemic utility in cases of peer

disagreement according to the actualist formulation of objective epistemic consequentialism I have outlined here, is committing to a conciliatory decision procedure.

The conciliatory decision procedure C is very different from the non-conciliatory decision procedure usually employed within Bayesian epistemology, under which rational agents should only adopt the credences that are supported by their total evidence. I have argued that such decision procedures would, if they were executed perfectly, require that agents give no weight to the opinions of their peer. This accords with the intuitions of those who have defended non-conciliatory responses to cases of peer disagreement. However, as I have shown here, itdoes *not* follow from this that fallible epistemic agents ought to commit to these decision procedures, since doing so will not maximise epistemic utility.

Being able to justify conciliatory responses in cases of peer disagreement is not kind of result one would expect to come out of a theory that doesn't give any weight to subjective oughts. Especially when such a view is combined with a strongly externalist account of evidence. The fact that the theory of objective epistemic consequentialism I have outlined here is able to recover a conciliatory account of peer disagreement within a wholly objective, externalist framework is, I propose, a strong point in favour of that theory.

Conclusion

In this thesis I have built up from the most fundamental grounds of the components of consequentialist theories and their account of value. I have applied the understanding of this foundation to the construction and defence of a detailed, actualist form of objective epistemic consequentialism. The version of epistemic consequentialism I have defended has the advantage of incorporating the positive results of global consequentialism, which allows it to evaluate focii beyond those of actions and beliefs. It also draws on several key insights from the work on epistemic value within traditional epistemology and applies these to a more formal account of partial beliefs.

Providing a deep and comprehensive foundation for our epistemic consequentialist theory has, in turn, yielded some extremely worthwhile results. One of the most promising results of the formulation of objective epistemic consequentialism that I have constructed here, is that it is able to provide an adequate response to the problem of peer disagreement. Not only this, but it is also able to overcome a variety of objections that have been posed to consequentialist accounts within both ethics and epistemology.

The objective epistemic consequentialist theory that I have constructed here provides epistemologists with a viable starting point for further work in this area. The benefits of objective epistemic consequentialism that have already been displayed in this thesis can be taken as evidence that such a project is indeed worthy of further exploration.

Bibliography

Becker & Becker, (Eds.), Encyclopedia of Ethics, Vol. 1-3, Routledge, London, UK 2001

Berker, S., 'Epistemic Teleology and the Separateness of Propositions' (draft of July 2010)

Boghossian P., 'Epistemic Rules', *The Journal of Philosophy*, Vol. 105, No. 9, September 2008

Chang, R., (Ed.), *Incommensurability, Incomparability, and Practical Reason*, Harvard University Press, USA 1997

Christensen (2010), 'Higher-Order Evidence', *Philosophy and Phenomenological Research* Vol. 81, No. 1, pp. 185 - 215, 2010

Clifford, W. K., 'The Ethics of Belief' (1879), in *The Ethics of Belief and Other Essays*, Prometheus Books, New York, USA 1999

Elga, A., 'Reflection and Disagreement', Noûs, Vol. 41, No. 3, pp. 478 - 502, 2007

Feldman, F., Doing the Best We Can: An Essay in Informal Deontic Logic, Dordrecht:D. Reidel Publishing Company, 1986

Gettier, E., 'Is Justified True Belief Knowledge?', Analysis, Vol. 23, pp. 121 - 23 (1963)

Goldman, A.I., Knowledge in a Social World, Oxford University Press, UK 1999

Greaves, H., 'Epistemic Decision Theory', (*draft*, 2010)

Grimm, S. R., 'Epistemic Normativity', in Haddock, Millar & Pritchard (Eds.), *Epistemic Value*, Oxford University Press, UK 2009

Haddock, Millar & Pritchard (Eds.), Epistemic Value, Oxford University Press, UK 2009

Hattiangadi, A., 'The Love of Truth', Studies In History and Philosophy of Science Part A, Vol. 41, No. 4, pp. 422-432, December 2010

Hooker, B., – *Ideal Code, Real World*, Oxford University Press, UK 2000 – 'Rule Consequentialism', *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), E. N. Zalta (Ed.), http://plato.stanford.edu/archives/spr2011/entries/consequentialism-rule/>

Hutter, M., 'On Universal Prediction and Bayesian Confirmation', *Theoretical Computer Science*, Vol. 384, No. 1, pp. 33 - 48, September 2007

Jackson 'Decision-theoretic Consequentialism and the Nearest and Dearest Objection', *Ethics*, Vol. 101, pp. 461 - 482, 1991

James, W., The Will to Believe, Longmans, Green, and Co., USA 1896

Jaynes, E. T., 'Information Theory and Statistical Mechanics', *Physical Review, Series II*, 106 (4): 620-630, 1957

Jeffrey, R., The Logic of Decision, 2nd edition, University of Chicago Press, USA 1964/1983

Jehle, D., & Fitelson, B., 'What is the "Equal Weight View"?', *Episteme*, Vol. 6, No. 3, pp. 280 - 293, 2009

Joyce, J., – 'A Nonpragmatic Vindication of Probabilism', *Philosophy of Science*, Vol. 65, 1998

The Foundations of Causal Decision Theory, Cambridge University Press, UK 1999
'Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief', Huber & Schmidt-Petri (Eds.), Degrees of Belief, Synthese Library, Springer, USA 2009

Kagan, S., 'Evaluative Focal Points', Hooker, Mason & Miller (Eds.), *Morality, Rules and Consequences: A Critical Reader*, Edinburgh University Press, UK 2000

Kamm, F. M., 'Harming Some to Save Others', *Philosophical Studies*, Vol. 57, No. 3, 1989

Kelly, T., 'Epistemic Rationality as Instrumental Rationality: A Critique' *Philosophy and Phenomenological Research*, Vol. 66, No.3, pp.612 - 640, May 2003

Kolodny & MacFarlane, 'Ifs and Oughts', *Journal of Philosophy*, Vol. 107, No. 3, pp. 115 - 143, 2010

Korsgaard, C., 'Two Distinctions in Goodness', *Philosophical Review*, Vol. 92, pp. 169 - 95, 1983

Lewis, D., Counterfactuals, Blackwell, Oxford, UK 1973

McNamara, P., 'Deontic Logic', *The Stanford Encyclopedia of Philosophy* (Fall 2010 Edition), Edward N. Zalta (Ed.), http://plato.stanford.edu/archives/fall2010/entries/logic-deontic/>

Norcross, A., – 'Good and bad actions', *Philosophical Review*, Vol. 106, pp. 1 - 34, 1997 – 'The Scalar Approach to Utilitarianism', West (Ed.), *The Blackwell Guide to Mill's Utilitarianism*, Blackwell, Oxford, UK 2006

Ord, T., Beyond Action: Applying consequentialism to decision making and motivation, DPhil Dissertation, University of Oxford 2009

Parfit, D., Reasons and Persons, Oxford University Press, UK 1984

- 'What we together do', unpublished manuscript, March 29, 1988

- 'Equality and priority', Ratio, Vol. 10, No. 3, pp. 202 - 22, December 1997

Pettit & Smith, 'Global consequentialism', Hooker, Mason & Miller (Eds.), Morality, Rules and Consequences: A Critical Reader, Edinburgh University Press, UK 2000

Pritchard, D., 'Recent Work on Epistemic Value', *American Philosophical Quarterly*, Vol. 44, pp. 85 - 110, 2007

Railton, P., 'How Thinking About Character and Utilitarianism Might Lead to Rethinking the Character of Utilitarianism', *Midwest Studies in Philosophy*, Vol. 13, No. 1, pp 398 - 416, 1988

Reichenbach H., The Direction of Time, University of Los Angeles Press, USA 1956

Savage, L., The Foundations of Statistics, 2nd edition, Dover, New York, USA 1954/1972

Silins, N., 'Deception and Evidence', Philosophical Perspectives, Vol. 19, pp. 375 - 404, 2005

Vallentyne, 'The Teleological/Deontological Distinction', *Journal of Value Inquiry*, Vol. 21, No. 1, 1987

Weatherson, B., 'Do Judgments Screen Evidence?', *online paper*, 2010,

 kian.weatherson.org/JSE.pdf>

Wedgwood, R., 'Internalism Explained', *Philosophy and Phenomenological Research*, Vol. 65, pp. 349 - 369, 2002

- 'The Meaning of 'Ought', Shafer-Landau (Ed.), Oxford Studies in Metaethics: Vol. 1, pp. 127 - 160, 2006

White, R., 'Evidential Symmetry and Mushy Credence', Gendler & Hawthorne (Eds.), Oxford Studies in Epistemology: Vol. 3, Oxford University Press, Oxford, UK 2010

Williamson, T., Knowledge and its Limits, Oxford University Press, Oxford, UK 200

Wilson, A., 'Disagreement, Equal Weight and Commutativity', *Philosophical Studies*, Vol. 149, No. 3, 2010