Book Proposal
Working Title: Do the Right Thing

Brian Weatherson, July 16, 2010

1 Overview

In recent work in epistemology, and to a lesser degree in ethics, there has been a surprising turn away from the simple idea that believers should believe what is supported by their evidence, and actors should do the right thing in their situation. We see this coming through in a number of ways:

- According to the 'Equal Weight View' of disagreement, if an agent finds out that an 'epistemic peer' disagrees with her, she should move her credences to somewhere between her prior credences and the peer's credences, even if her credences were reasonable before she hear about the peer. This kind of view is defended by Elga (2007, 2010); Christensen (2007c); Feldman (2006); Bogardus (2009); and Matheson (2009), and discussed by many others. Indeed, it is one of the main themes of the volume Warfield and Feldman (2010), the special *Episteme* issue on disagreement, another proposed OUP volume on disagreement edited by Jennifer Lackey and David Christensen, and recent or forthcoming articles in *Mind* (Enoch, 2010), *Philosophical Studies*, (Wilson, 2010), *Philosophy and Phenomenological Research* (Frances, forthcoming) and *Philosophical Quarterly* (Thune, 2010).

- Relatedly, David Christensen (2010; 2007b; 2007a) has argued that if we have evidence about our own fallibility, we should adjust our credences to reflect that fact, even in cases where we had antecedently formed rational judgments. Christensen argues that this kind of evidence about our own fallibility is a distinctive kind of evidence concerning the subject matter of our investigations, what he calls 'higher-order' evidence, even though the evidence does not seem to bear directly on the subject matter of those investigations.

- And several ethicists have argued that in cases where we are unsure what is the right thing to do, we should use an algorithm that computes the appropriate action for us out of the probabilities that one or other ethical theory is correct. See, for example, Lockhart (2000); Guerrero (2007); Sepielli (2009).

I think these views are all mistaken, and propose to defend in their place a form of normative externalism. What I mean by normative externalism is that there are norms governing what an agent should do, and about what she should believe, in various circumstances. And those norms are binding even if the agent does not know they exist, and indeed even if she has a justified belief that they are otherwise. So an agent might be required to believe p in a certain evidential situation, even if she has good reason to believe that a belief in p would be unjustified. And she might be required to ϕ in a situation, even if she has good reason to believe that ϕ is morally

impermissible in that situation. For instance, she might be required to lie to the murderer at the door even if she believes, and even if she fully rationally believes, a Kantian ethical theory that says this is impermissible.

And the externalism I want to defend goes even a touch further than that. Some philosophers might concede that there is some sense in which our agent should lie to the murderer at the door, but there is another sense in which she should not. There is some kind of norm, perhaps of rationality or of coherence, according to which acting in accord with the norms one (rationally) accepts, is the thing to do. This book will argue, primarily through the regress arguments in chapters 4 and 7, that there is no such sense. The only over-arching rule is that we should do the right thing, and believe the rational thing, even when we are confused about what is right and rational.

What's at the heart of normative externalism is a view about the evidential relationship between 'first-order' evidence and the evidence we get from knowing that someone made a judgment. Often we do get evidence that p from the fact that a smart well-informed agent judges that p. I'm going to argue that when we know exactly what evidence an agent has, knowing what they did with that evidence is of no epistemic significance. In the terminology I'll use in the book, the evidence a judgment is based on *screens off* the evidential support that judgment provides.

Normative externalism, of the kind I prefer, rules out a number of popular positions in the three debates mentioned above. Indeed, in most cases it rules out all but one of the actually defended positions in those debates. It implies that the Equal Weight View of disagreement is wrong, and something in the vicinity of the Right Reasons view (Kelly, 2005) is correct. It implies that there is no purely higher-order evidence, in the sense intended by Christensen. And it implies that the epistemic probability of a moral theory being correct is irrelevant to what one should do; what matters is the truth of that theory. So the arguments for normative externalism that I'm going to present in the book are relevant to a number of debates across epistemology and ethics.

1.1 Intended Readership

The primary audience for the book will be philosophers and philosophy students. The book would be very useful for seminars on epistemology, especially for seminars on disagreement. As noted above, there have been a flood of recent articles on disagreement, so I think a (hopefully high-quality) book-length treatment of the topic will be noticed. To the extent that there's a trend in this debate, it is toward 'moderate' positions, which say that the Equal Weight View is not right, but that rational agents should make some concessions to disagreement from apparent peers. I want to resist this trend. Strictly speaking, peer disagreement is of no epistemic significance at all, though in practice it is hard to tell just who is a peer.

So I think the book will be widely read inside academic philosophy as well. But I think it will have some resonance outside philosophy as well. My impression is that debates about disagreement, and about the appropriate response to disagreement, are interesting to at least other academics in a way that a lot of other topics in epistemology are not. I've found non-philosophers on blogs to be more interested in debates

about interest-relativity than in, say, debates about the right logic for vagueness. And I've seen citations to philosophers (e.g., Plantinga) writing on religious disagreement in popular theology books in my local bookshop. So there's some chance this work will have some resonance outside traditional philosophical circles.

And I think debates about moral uncertainty could also, with the appropriate marketing, be of interest to a wider audience. One of the things we see in the existing literature is that considerations about moral uncertainty suggest simple, and apparently powerful, arguments against the permissibility of abortion and of meat-eating. There are various ways to put the arguments; here is the simplest. Premise 1: You should do what is least likely to be immoral. Premise 2: Bringing a child to term/Refraining from eating meat is almost certainly not immoral (at least in normal circumstances). Premise 3: Unless we are unreasonably certain that the arguments by pro-life and pro-vegetarian theorists are wrong, there is a non-negligible likelihood that having an abortion/eating meat are wrong. Conclusion: You shouldn't have an abortion/eat meat. I think the only real way to respond to these kinds of arguments is to attack premise 1 head on. Given the widespread interest in the conclusions of these arguments, and the accessibility of the arguments, I think a lucid discussion of them would be attractive to a wide readership.

1.2   Title

The title here is really a working title; I think it's a little too removed from the content of most of the book to be quite right. (I probably only thought of it because I'm writing this in a sweltering New York summer's week.) Hopefully a more accurate and succinct title will come up over the course of writing.

1.3   Timeline

I currently have about 20,000 words written between a couple of working papers that would be turned into chapters of this book. In particular, much of chapters 1, 2, 3, 4 are written, as is at least the first part of chapter 6, and a sketch of chapter 8. I was originally writing these with the thought that chapters 1-4 and 6 could be covered in a journal article sized paper, so I went by some points fairly abruptly. Versions of that paper have been presented at MIT, Oxford, St Andrews and Konstanz, and discussed with several people around New York, so I've got quite a bit of feedback on the paper. The upshot is that turning those parts into a book manuscript draft would be reasonably quick. Much of chapter 8 is based on a different paper I presented at St Andrews, so it could be turned at least into a draft reasonably quickly as well. Chapters 5 and 7 will take a little more time, and if there needs to be an appendix to chapter 3, that will take a little time getting done carefully as well. But I think it is realistic to think I'd have a full draft ready within 12 months.

2   Chapter Outline

Chapter 1   Introduction: Judgment, Screening and Norms

When an agent we know to be rational makes a judgment, the judgment is usually evidence for the truth of what is judged. Some odd things happen though when we take the agent who comes into possession of that evidence to be the very agent who

made the judgment. It seems odd to say that an agent who rationally judges that p on the basis of E automatically has two pieces of evidence in favour of p: her original evidence E, and the fact that a rational agent, namely herself, has judged that p. That smacks of dubious 'double counting'. If we don't want to say that, we have two options:

1. We could say that her judgment 'screens off' her initial evidence. I call this position JSE, short for **J**udgments **S**creen **E**vidence.
2. Or we could say that her evidence screens off her judgment. I call this position ESJ.

This book will be an extended attack on JSE, and a defence of ESJ. In the initial chapter, I will do several things to set up the main storyline.

- I'll say a fair bit about the notion of 'screening' that I'm using. The notion is taken in large part from Reichenbach (1956), though it needs several technical tweaks to make it applicable to contemporary epistemological debates. Roughly, the idea is that A screens off the evidential support that B provides to C iff B is evidence for C, but $A \wedge B$ is no better evidence for C than A alone is.
- I'll say a little about why the debates about ESJ and about JSE, are relevant to a number of epistemological issues. I'm going to argue in chapter 2 that JSE is necessary and sufficient to support Christensen's views on higher-order evidence, and in chapters 3 and 4 that JSE is necessary and sufficient to support the Equal Weight View. Here I'll set up the structure of these arguments.
- I'll describe the kind of externalism about normativity that's motivated by ESJ, and the picture of theoretical and practical inquiry that it leaves us with. As noted above, defending this picture is really the primary aim of the book.

Chapter 2   Higher-Order Evidence

In a series of recent papers, David Christensen (2010; 2007b; 2007a)has argued for the existence of a distinctive new kind of evidence: 'higher-order evidence'. Here is an example of the kind of thing he has in mind. An agent knows that the number of αs is 42+29. She concludes from this that the number of αs is 71. She is then given strong (but as it turns out misleading) evidence that someone in her state is highly unreliable at basic arithmetic. (For instance, she is told that she has taken a drug that makes most people bad at basic arithmetic.) Christensen says this is evidence against the proposition that the number of αs is 71, and that finding out about her (apparently) unreliability is a form of 'higher-order evidence'.

I'm going to argue for two conclusions. First, Christensen's views on evidence stand and fall with JSE. Second, Christensen's views are either unmotivated or false. The argument for the second conclusion turns on cases where the agent rationally concludes that p is probably, but not certainly, true, and then gets evidence that people in situations like hers are usually too cautious. In some of those cases, the agent should not act as if p is true, especially if acting as if p is true threatens the well-being of other people. Given some premises Christensen needs for his own argument for

higher-order evidence, that implies that the evidence suggesting she is too cautious is not really higher-order evidence, although it should be given Christensen's picture. So Christensen's conclusions are inconsistent with his premises, which means that either his conclusion is false, or his main argument for his conclusion is unsound.

Chapter 3   Disagreement and Defeat

The Equal Weight View says that when we find out that someone we (reasonably) thought was a peer has the same evidence as us, but different views on p to us, we should alter our credences in p to something like the mid-point between our prior credence in p and our peer's credence in p. The idea is that in such circumstances, the weight of evidence for p and $\neg p$ is (something like) evenly balanced.

But actually it doesn't seem like the evidence is balanced at all. If S and T are peers, or seemed to be before this encounter, and S has rationally judged p on the basis of E, then S's evidence for and against p looks something like this:

Evidence for p	Evidence against p
S's judgment that p	T's judgment that $\neg p$
E	

Unless S and T's judgments are given exceedingly high weights, on the face of it we'd say that the evidence is still overwhelmingly in favour of p, since E is still strong evidence in favour of p. There are two moves that the Equal Weight theorist might make here. They might adopt JSE and say that E is **screened off** by S's own judgment. I'll deal with that possibility in chapter 4, though of course we'll have already seen some reason in chapter 2 to be suspicious of it. The other move is to say that the support E provides for p is **defeated** by T's judgment that $\neg p$.

This chapter is dedicated to showing why the defeat move won't work, so the Equal Weight theorist needs JSE. I have two main kinds of arguments. First, there's no plausible way to make this story work in cases where S finds out about a number of different 'peers' over time. Second, this view rests on implausible views about how defeat could work. It's vaguely plausible that T's judgment that $\neg p$ could defeat the support that S's judgment that p provides for p. It's much less plausible that T's judgment could normally defeat the support that E provides for p.

When I've presented this material at various places, I've found that many people mean many different things by the 'Equal Weight View'. And this starts to matter at just this point of the argument. Some people think it just means that S's judgment and T's judgment should have the same epistemic significance. (I accept that, though not everyone does.) Others think that it means S should move her credences to somewhere between her initial credences and T's. I think that's false, if we make all of the assumptions usually made in this dispute, though it will be true in a lot of real-world situations.

So it might turn out to be necessary to survey the possible things people might mean by the 'Equal Weight View', and the things that have been meant in published articles on disagreement. (If this turns into too much like a literature survey from a PhD dissertation, I might have an exegetical appendix to chapter 3.) My sense is

that there isn't a principled theory around that's weaker than implausible views, and strong enough to be worth caring about, but it might take a bit of time going finely through what has been written to establish this conclusion.

Chapter 4 Screening and Regress

Chapter 3 was intended to establish that the Equal Weight View of disagreement is correct only if JSE is true. I'll start this chapter by arguing that the converse is true, and that given JSE various objections to Equal Weight (such as those recently voiced by Thomas Kelly (2010) and Jennifer Lackey (2010)) can be answered.

But JSE is false, as we can see by noting that it falls into a vicious regress. That there's some kind of regress is fairly clear. According to JSE, every time we make a judgment we get a new piece of evidence, namely that judgment, and that evidence screens off our old evidence. So it seems we should be in a position to make a new judgment based on the this new evidence, and this starts an endless regress.

That there is some kind of regress here seems obvious. That it is problematic is far from obvious. But I'm going to argue that it is problematic, and that it's so because the apparent ways to stop it don't actually work. After replying to recent attempts to block the regress by Sherrilyn Roush (2009) and Adam Elga (2010), I discuss a 'ratificationist' response to the regress. The ratificationist move is based on moves made by Richard Jeffrey (1983) in decision theory, and I argue it fails for similar reasons to why Jeffrey's decision theory fails.

Chapter 5 Defending ESJ

I'm going to argue that once we know what evidence a person has, learning what judgment they make has no (further) evidential force. That's one consequence of the view, presented by Frank Jackson (1987), that a person's judgment that p is epistemically valuable because (and only because) it is evidence that they have evidence that p. This kind of view has generally been taken to have many counter-intuitive consequences, and this chapter will defend the view against those challenges.

Many people have argued that ESJ-type views issue implausible verdicts in cases where a large number of apparent peers, with apparently the same evidence, disagree with our judgment. David Christensen (2007c) makes such an argument, and Thomas Kelly (2010) offers it as one of the main reasons he moved away from the ESJ-friendly view he endorsed in his (2005). I'm going to argue that in any plausible version of the case we do not know, at the most important time, that the disagreeing agents have the same evidence as us.

If we endorse ESJ, it isn't clear what should be done in cases where we have reason to doubt our own competences. (These cases are central to the anti-ESJ stories told in Egan and Elga (2005) and Roush (2009).) I'm going to argue first that the anti-ESJ intuitions that have been elicited depend on looking at too few cases. In cases where we have reason to believe that we are overly cautious, ESJ gives the intuitively correct answer. But my main argument will be that when we have reason to doubt our own competence, we should be very hesitant to reconsider judgments that we've made. (The line I'm taking here draws heavily on work by Richard Holton (2009), who

argues that sometimes it is irrational to reconsider our judgments.) If we don't reconsider judgments that are rationally made, we can avoid the alleged counter-intuitive consequences.

Finally, it is sometimes thought that ESJ deprives us of epistemic resources, since it implies that some self-verifying beliefs (such as the belief that there are beliefs) are not well-supported. Following some discussion in Enoch (2010), I'm going to argue that whether a self-verifying belief is justified depends on whether it is appropriate to take a 'third-personal' perspective on our own judgments, and ESJ can explain why this is so.

Chapter 6   Uniqueness and Levels of Justification

Two big picture questions will have arisen over the course of the previous chapters.

1. Is it possible for an agent to be justified in adopting one of a (non-singleton) set of doxastic positions, or is there a unique state that is justified given her evidence and situation?
2. Is it possible for an agent to be justified in believing p without being justified in believing that she's justified in believing p?

The first question comes up in considering a simple objection to the Equal Weight View developed by Kelly (2010). If we know that there are equally reasonable responses to evidence E, then learning that someone else made one of the judgments we didn't make seems clearly to be of no significance. If peer disagreement is of any significance, it is only because it threatens the reasonableness of our judgment, but getting from the existence of the disagreement to a conclusion about our judgment is hard without a uniqueness assumption.

The second question comes up in the discussion of ESJ, and in the discussion of defeaters. I'm going to argue that in a lot of the thought experiments used to motivate opposing views to mine, people are confusing what an agent is justified in believing with what she's justified in believing that she's justified in believing. That move presupposes that the two things can come apart, and that's rejected by some of the theorists I'm discussing.

My answers to these two questions are basically going to be the opposite to those endorsed by most proponents of the Equal Weight View. I'm going to argue for the view often called 'Epistemic Permissiveness' White (2005), and against the view that there is at most one rational state for an agent to be in. And I'm going to argue that we can have justified beliefs in cases where the agent would not be justified in believing that she has a justified belief. My main opponent on this second question will be Richard Feldman (2005), who has argues that this kind of possibility would imply paradox.

Chapter 7   Moral Uncertainty

A number of recent authors have argued that as well as (or perhaps even instead of) moral norms, there are norms governing what we should in some sense do when we are uncertain about morality. For instance, Ted Lockhart (2000) argues that we should do the action that is most probably (given our evidence) morally permissible. And Andrew Sepielli (2009) argues that we should maximise the expected goodness of our acts. This chapter is an argument against all such views.

Part of the chapter will be dedicated to raising concerns about the particular proposals for dealing with moral uncertainty that have been raised, and noting some of their odd consequences. But the larger part will be an argument that there isn't any interesting subject matter to do with moral uncertainty. What an agent should do just is what they morally should do; what they probably should do is not, I'm going to argue, normatively interesting.

One of the arguments for that will be a version of the 'moral fetish' argument developed by Michael Smith (1994). Another will be a version of the regress argument run in chapter 4 against JSE. And another will be to do with issues raised by agents who are not subject to moral uncertainty, but simply have false moral views.

I'm partially interested in this question because it is an interesting application of normative externalism. But it's also relevant to the arguments of the previous chapters. Some of my arguments will concern cases where an agent has misleading evidence about what to do, perhaps because she read the wrong philosophical books! If what I conclude in this chapter is wrong, some of those arguments will be weaker. So this chapter is needed to tie down some loose ends from those chapters.

Chapter 8   Evidential Quality

The bulk of this book will have been defending a version of evidentialism, the view that epistemic justification supervenes on evidence. But recently Alvin Goldman (2009) has presented an argument that is a challenge to evidentialism, or at least to the combination of evidentialism with the plausible thesis that all and only knowledge is evidence (Williamson, 2000). Goldman's argument, roughly, is that we can have inductive knowledge, but inductive knowledge isn't itself the basis for further inductive knowledge. So if evidence is knowledge, it seems that what we can infer doesn't just depend on what evidence we have, it depends on how we got that evidence, i.e., on whether that evidence was obtained inductively or otherwise.

My response is that we need to distinguish between the *quality* of different agents' evidence, as well as between the *quantity* of their evidence. Assume that S knows that p by induction, then conducts some observations that verify that p. S's knowledge that p is now more *sensitive*, since perception is more sensitive than induction. I'm going to argue that this extra sensitivity of S's knowledge makes p *better* evidence for S, and this explains why it can now ground more inductive conclusions, and this shows how to avoid the problem Goldman raised.

References

Bogardus, Tomas. 2009. "A Vindication of the Equal-Weight View." *Episteme* 6:324–335.

Christensen, David. 2007a. "Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals." *Oxford Studies in Epistemology* 2:3–31.

—. 2007b. "Epistemic Self-Respect." *Proceedings of the Aristotelian Society* 107:319–337.

—. 2007c. "Epistemology of disagreement: The good news." *Philosophical Review* 116:187–217.

—. 2010. "Higher-order evidence." *Philosophy and Phenomenological Research* 81:185–215.

Egan, Andy and Elga, Adam. 2005. "I Can't Believe I'm Stupid." *Philosophical Perspectives* 19:77–93.

Elga, Adam. 2007. "Reflection and disagreement." *Noûs* 41:478–502.

—. 2010. "How to disagree about how to disagree." In Warfield and Feldman (2010).

Enoch, David. 2010. "Not Just a Truthometer: Taking Oneself Seriously (but not too seriosuly) in Cases of Peer Disagreement." *Mind* Forthcoming.

Feldman, Richard. 2005. "Respecting the evidence." *Philosophical Perspectives* 19:95–119.

—. 2006. "Epistemological puzzles about disagreement." In *Epistemology Futures*, 216–226. Oxford University Press.

Frances, Bryan. forthcoming. "The reflective epistemic renegade." *Philosophy and Phenomenological Research* .

Goldman, Alvin. 2009. "Williamson on Knowledge and Evidence." In *Williamson on Knowledge*, 73–91. Oxford: Oxford University Press.

Guerrero, Alexander. 2007. "Don't Know, Don't Kill: Moral Ignorance, Culpability and Caution." *Philosophical Studies* 136:59–97.

Holton, Richard. 2009. *Willing, Wanting, Waiting*. Oxford: Oxford University Press.

Jackson, Frank. 1987. *Conditionals*. Blackwell: Oxford.

Jeffrey, Richard C. 1983. *The Logic of Decision*. Chicago: University of Chicago Press, 2nd edition.

Kelly, Thomas. 2005. "The epistemic significance of disagreement." In *Oxford Studies in Epistemology, Volume 1*, 167–196. Oup.

—. 2010. "Peer disagreement and higher order evidence." In Warfield and Feldman (2010).

Lackey, Jennifer. 2010. "What Should We Do When We Disagree." *Oxford Studies in Epistemology* Forthcoming.

Lockhart, Ted. 2000. *Moral Uncertainty and its Consequences*. Oxford University Press.

Matheson, Jonathan. 2009. "Conciliatory Views of Disagreement and Higher-Order Evidence." *Episteme: A Journal of Social Epistemology* 6:269–279.

Reichenbach, Hans. 1956. *The Direction of Time*. Berkeley: University of California Press.

Roush, Sherrilyn. 2009. "Second-Guessing: A Self-Help Manual." *Episteme* 6:251–268.

Sepielli, Andrew. 2009. "What to Do When You Don't Know What to Do." *Oxford Studies in Metaethics* 4:5–28.

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

Thune, Michael. 2010. "'Partial defeaters' and the epistemology of disagreement." *Philosophical Quarterly* 60:355–372.

Warfield, Ted and Feldman, Richard. 2010. *Disagreement*. OUP.

White, Roger. 2005. "Epistemic permissiveness." *Philosophical Perspectives* 19:445–459.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.

Wilson, Alastair. 2010. "Disagreement, equal weight and commutativity." *Philosophical Studies* 149.