

REASONS AGAINST BELIEF:  
A THEORY OF EPISTEMIC DEFEAT

by

Timothy D. Loughlin

A DISSERTATION

Presented to the Faculty of  
The Graduate College at the University of Nebraska  
In Partial Fulfillment of Requirements  
For the Degree of Doctor of Philosophy

Major: Philosophy

Under the Supervision of Professor Albert Casullo

Lincoln, Nebraska

May, 2015

REASONS AGAINST BELIEF:  
A THEORY OF EPISTEMIC DEFEAT

Timothy D. Loughlin, Ph.D.

University of Nebraska, 2015

Adviser: Albert Casullo

Despite its central role in our cognitive lives, rational belief revision has received relatively little attention from epistemologists. This dissertation begins to fill that absence. In particular, we explore the phenomenon of defeasible epistemic justification, i.e., justification that can be lost as well as gained by epistemic agents.

We begin by considering extant theories of defeat, according to which defeaters are whatever cause a loss of justification or things that somehow neutralize one's reasons for belief. Both of these theories are both extensionally and explanatorily inadequate and, so, are rejected.

We proceed to develop a novel theory of defeat according to which defeaters are reasons against belief. According to this theory, the dynamics of justification are explained by the competition between reasons for and reasons against belief. We find that this theory not only handles the counter-examples that felled the previous theories but also does a fair job in explaining the various aspects of the phenomenon of defeat. Furthermore, this theory accomplishes this without positing any novel entities or mechanisms; according to this theory, defeaters are epistemic reasons against belief, the mirror image of our epistemic reasons for belief, rather than *sui generis* entities.

To those writing dissertations,  
just swallow your pride and finish.

# ACKNOWLEDGEMENTS

I'm thankful to everyone who made me curious, rigorous, and bold. Special acknowledgment is due to Bill Edgar and Walt Soffer for awakening a love of philosophy in me, to Steve Swartzler, Cullen Gatten, Luke Elwonger, and Landon Hedrick whose company on the academic road was invaluable, to the faculty of the philosophy department at UNL who never made me feel like an imposter, to the CAPS staff who kept me sane, to Amy Green and the Cone Crew, my dearest friends, to those close to me who have had to deal with my gigantic melancholies and gigantic mirth, most of all Katie, without whom I'd have given up during the final leg of the race, and to Al, an exemplar.

# TABLE OF CONTENTS

0	TOPIC & METHODOLOGY	1
	0.1 Theoretical Adequacy	3
	0.2 The Plan of the Dissertation	7
	0.3 A Note on Terminology	8
1	THE NEED FOR A THEORY	10
	1.1 Defeat & Externalist Theories of Justification	11
	1.11 The First Horn of the Trilemma	12
	1.12 The Second Horn of the Trilemma	15
	1.13 The Third Horn of the Trilemma	19
	1.14 The Need for a Theory	26
	1.2 Higher-Order Evidence	27
	1.21 Identifying Higher-Order Evidence with Defeat	30
	1.22 Distinguishing Higher-Order Evidence from Defeat	35
	1.3 Conclusion	38
2	THREE THEORIES OF DEFEAT	39
	2.1 The Theories	40
	2.2 Easy Cases	41

2.3	The Causal Theory of Defeat	44
2.31	The Cases it Gets Right	45
2.32	The Cases it Gets Wrong	46
2.4	The Reason-Neutralizing Theory of Defeat	54
2.41	The Cases It Gets Right	55
2.42	The Case It Gets Wrong	61
2.5	Explanatory Inadequacy	66
2.6	The Quietist Theory of Defeat	71
2.7	Conclusion	73
3	THE REASONS AGAINST BELIEF THEORY OF DEFEAT	74
3.1	Perfunctory Chisholming	77
3.2	Extensional Adequacy	85
3.21	The Causal Theory	86
3.22	The Reason-Neutralizing Theory	89
3.3	Explanatory Adequacy	91
3.31	The Causal Theory	92
3.32	The Reason-Neutralizing Theory	98
3.33	Further Explanations	100
3.4	Epistemic Reasons For and Against Belief	102
3.5	Unjustified Belief and Justified Unbelief	106
3.6	Conclusion	110

4	SYMMETRY OF REASONS	112
4.1	Justification-Asymmetry	115
4.11	The First Argument	118
4.12	The Second Argument	126
4.13	Defeat & Proper Function	128
4.14	In Support of Justification-Symmetry	130
4.15	Coda	135
4.2	Other Asymmetries	137
4.21	Awareness-Asymmetry	139
4.22	Meta-Awareness-Asymmetry	146
4.23	Basing-Asymmetry	149
4.3	Conclusion	153
5	APPLICATIONS & IMPLICATIONS	155
5.1	Externalism & Defeat	156
5.2	Higher-Order Evidence	165
5.21	Supposed Differences	168
5.22	Dismissing Difference	176
5.23	Bracketing & Defeat	181
5.3	Conclusion	185
6	Works Cited	187

# PREFACE

## Topic & Methodology

You are looking for your keys and, because you remember leaving them on your desk, you look for them there. But, after shuffling your papers about and finding no keys, you decide to check in yesterday's laundry.

The above situation could be described in the following terms, though only philosophers would be likely to do so: (i) you form the belief that your keys are on the desk on the basis of your apparent memory, (ii) you acquire evidence that your keys are not on your desk and, (iii) on the basis of that evidence, you cease to believe that your keys are on your desk.

Nothing could be more common than this sort of belief revision. We are constantly updating our beliefs in light of new evidence. The formation of belief and the cessation of belief alike are attempts to accurately represent the world. Put this way, it seems strange to think of the formation of belief and the cessation of belief as anything other than two sides of the same coin.

Yet the conditions for appropriate, or justified, belief formation are often treated by epistemologists as wholly separate from the conditions for justified belief cessation. Consider this sampling of theories of epistemic justification:



In all of these theories, there is some set of positive conditions on justification, e.g., being formed on the basis of *prima facie* reasons, and then, almost as an afterthought, a negative condition, e.g., the absence of defeaters. This puzzling treatment of the conditions for justified cessation of belief is the topic of this dissertation.

To be clear, we are concerned with justification, not knowledge. We are exploring the relationship between one's being justified in believing, one's being unjustified in believing, and one's being justified in not believing. We try to remain as neutral as possible concerning different analyses of the concept of JUSTIFICATION itself, but there will be points at which some analysis will be necessary. For the most part, however, it will suffice to think of the claim that an agent is justified in believing a proposition as meaning no more and no less than that it is epistemically appropriate for that agent to form that belief.

It is convenient to refer to the topic of this dissertation as the phenomenon of epistemic defeaters for justification. As an approximation, in the case above the evidence you acquire when you search your desk would be an epistemic defeater for your justification for believing that your keys are on your desk. This title has the dubious virtue of possessing antecedent philosophical currency. It is a virtue because it more readily permits comparison between the claims of this dissertation and the extant literature. It is dubious because it carries with it inessential and potentially confusing baggage. On balance, we think the benefits outweigh the costs, but only if we are careful to distinguish between what is commonly believed of defeaters and what the lines of reasoning in this dissertation leads us to believe of defeaters. We will carefully explore

and critically evaluate theories of defeat that appear in the literature in order to draw out just what is fundamental about justified belief revision, which we will call defeat from here on, and what accretions are merely the result of the compounded mistakes of past theories. Since the goal of this dissertation is to explore and critically evaluate extant theories of defeat and, ultimately, develop and defend a novel theory of defeat, it will be helpful to spend some time here explaining the standards that will be employed when evaluating a theory.

### 1. *Theoretical Adequacy*

The adequacy of a philosophical theory is determined by two related standards: the standard of extensional adequacy and the standard of explanatory adequacy. The extension of a theory is the set of all instances in which the central concepts of that theory apply. For example, the extension of a theory of knowledge is the set of all and only cases of knowledge. Insofar as a theory of knowledge classifies some cases of non-knowledge as knowledge or cases of knowledge as non-knowledge, that theory is extensionally inadequate.

Extensional adequacy, however, is not sufficient for theoretical adequacy. After all, a theory of knowledge could, in principle, secure extensional adequacy by positing a disjunction, perhaps infinite, wherein each disjunct is a member of the set of cases of knowledge and all members of that set are included in that disjunction. In that way, this theory of knowledge would be extensionally adequate but it would not have illuminated the underlying nature of knowledge. We want more than an accurate account of a phenomenon in a philosophical theory; we also want an account that helps us to

understand that phenomenon. Just what it takes to achieve explanatory adequacy is difficult to identify, but the explanatory desiderata for a theory include the following.

Explaining otherwise unexplained or apparently *sui generis* phenomena in terms of concepts or mechanisms to which we are antecedently committed is desirable. For example, if one theory of knowledge posits that knowledge and justification can both be explained in terms of warrant and another theory posits that knowledge and justification are disparate phenomena, the latter theory is thereby at an explanatory disadvantage to the former theory since, other things being equal, it is more illuminating to unify rather than multiply phenomena.

A second explanatory desideratum illustrates the way in which explanatory adequacy is inextricably linked to extensional adequacy. It is desirable that a theory explain why the concepts central to that theory have the extension that they do. This is why the disjunctive theory of knowledge introduced above is explanatorily inadequate. Although it necessarily delivers the correct extension of the concept KNOWLEDGE, it fails to shed any light on why this belief, but not that belief, is an instance of knowledge. Similarly, a theory that has nothing to say about why this and not that falls within the extension of a concept would not help us to settle controversial cases. The extensions of many of the philosophically interesting concepts are controversial. What appears to one philosopher to be an extensionally adequate theory of knowledge might appear to another to be extensionally inadequate. Unless an explanation is provided for this proposed extension rather than that proposed extension, philosophical theorizing becomes nothing more than an uninteresting battle of intuitions.

This is not to say that intuitions about concepts have no place in philosophy, far from it. Our intuitions about concepts show us the core applications of the concept from which a theory can deviate only at some cost, i.e., a theory of knowledge that suggests that we have no knowledge of the external world is counter-intuitive and at an extensional disadvantage to a theory of knowledge that suggests that we do have knowledge of the external world, all else being equal.

All else is not always equal, however. This inadequacy can be outweighed by the explanatory power of the theory. If the theory that preserves our intuitions about core cases of the concept KNOWLEDGE fails to provide any explanation of knowledge, what it is, its purpose, or its apparent value, and the theory that sacrifices some of our intuitions is able to provide such explanations, that goes some way to making up the difference. Just how far it goes toward making up this difference depends on the strength of the explanations that the latter theory is able to provide and how strongly we're inclined to cling to our intuitions.

Just when the explanation gives way to the intuitions and when the intuitions give way to the explanation is a subject of study unto itself. Why is it that, when it was discovered that cetaceans are significantly dissimilar to other nektic creatures, it was the extension of the concept FISH that shifted rather than the criterion of biological similarity that defined that concept? Why is it that we are just as comfortable saying that witches were merely female outsiders who were scapegoated by their societies as we are saying that there never were any witches? Why is it that, when it was discovered that the particles that had been called “atoms” are divisible, we persisted in considering them to

be atoms rather than preserving the concept ATOM as an indivisible particle? Answering these questions goes well beyond the scope of the current project.<sup>1</sup> It should suffice for our purposes to note that we are committed to preserving neither our pre-theoretical intuitions about the extension of the concept of defeat nor our pre-theoretical notions about the conditions for defeat.<sup>2</sup>

Rather, we are committed to beginning with these and, through careful exploration of the concept and cases, making trade-offs between them when doing so makes for a better theory of defeat.<sup>3</sup> For example, some philosophers are loathe to identify any cases in which one loses no justification as cases of defeat. We will see that there are good theoretical reasons to reject this condition on defeat. The thing that such philosophers call “defeat” is actually a phenomenon that is explained by a mechanism that is present in some cases in which no justification is lost. However, it is this mechanism that we are interested in and not the term “defeat.” So, if one prefers to call the topic of the present work “potential defeat” and “potential defeater” or some such things, reserve the terms “defeat” and “defeater” for cases in which justification is lost, and agree that it is potential defeat that ultimately explains why that justification is lost, then any disagreement between us is merely terminological. That being said, and to reiterate our earlier point, we will use the terms “defeat” and “defeater” throughout as expedients.

---

1 See Thomas Kuhn (1962).

2 “Defeat” is a term of art and so we don't really have any pre-theoretical intuitions about it, anyway. The best we can do is consider our intuitions about when one is or is not justified in believing. As we will see in chapter 2, this is no replacement for a theory of defeat.

3 For more on this method, see Rudolf Carnap (1950) and, as it is applied specifically to epistemic concepts, Erik J. Olsson (2012, 2014).

## 2. *The Plan of the Dissertation*

In the first chapter of the dissertation we will explore two examples of the concept of DEFEAT being put to work in epistemology: the debate between internalists and externalists about justification and the controversies surrounding disagreement and higher-order evidence. We argue that in both cases, use of the concept of DEFEAT is premature. In the absence of a well-developed theory of defeat, such applications are bound to be inconclusive.

In the second chapter we present, develop, and ultimately reject three theories of defeat: the causal theory of defeat, the reason-neutralizing theory of defeat, and the quietist theory of defeat. We argue that the first two theories are extensionally inadequate and that all three theories are explanatorily inadequate.

Taking lessons from the failure of the three theories in the second chapter, we develop an improved theory of defeat in the third chapter: the eponymous reasons-against-belief theory. According to this theory, a defeat just is a reason to not believe. The apparent simplicity of this theories belies its power in handling problematic counter-examples and explaining the various aspects of the phenomenon of defeat.

In the fourth chapter we defend the reasons-against-belief theory from a family of objections. One of the strengths of the reasons-against-belief theory is that it treats defeat as the mirror image of justification, i.e., justification for believing proceeds in terms of reasons for believing and defeat proceeds in terms of reasons against believing. However, if there are significant asymmetries between justifiers and defeaters, then that casts doubt of this picture of defeat. A number of such apparent asymmetries will be considered and

rejected.

In the final chapter we return to the examples of applications of the concept of DEFEAT from the first chapter. We argue that the implications of the reasons-against-belief theory for those controversies are surprisingly modest. Since the theory asserts that defeaters are just further epistemic reasons, any trouble had by a theory of justification or of higher-order evidence in handling defeaters will not be independent of problems these theories have in handling justifiers.

### 3. *A Note on Terminology*

It is worth taking a moment here to explain some of the peculiar linguistic conventions used in this dissertation. For example, we write of “defeaters with respect to belief” rather than “defeater for a belief, justification, a person, etc.” The reason for this is that the former locution does not presuppose anything about the relationship between defeaters and oneself, one's beliefs, one's justification for belief, etc. The same cannot be said for the latter locution.

As a general rule, we will not assume that an agent has a particular belief. Instead, we will refer to the features of that belief were one to have it. For example, we might discuss John's being justified in believing some proposition instead of John's justified belief. It is sometimes particularly convenient to use the term “belief,” even when it needn't be supposed that one actually has the belief. In such cases, we might write of a defeater with respect to one's belief when what we mean is something that would be a defeater with respect to one's belief were one to have that belief. Where the distinction is particularly important, it will be noted.

Similarly, we endeavor to refer uniformly to mental states instead of slipping between references to mental states and references to propositions. For example, we try to never write of a defeater with respect to  $p$  when what we mean to refer to is a defeater with respect to the belief that  $p$ .

To help navigate some of the admittedly complicated epistemic relationships in this dissertation, we have tried to use the same the same shorthand throughout. The generic epistemic agent will always be referred to as “S,” e.g., S's believing. The mental state that is supposedly standing in the defeater role will always have the propositional content  $r$ , e.g., S's belief that  $r$  might be a defeater. The mental state with respect to which one might have a defeater will always have the propositional content  $q$ , e.g., S's belief that  $r$  might be a defeater with respect to S's belief that  $q$ . Lastly, the mental state that is supposedly standing in the justifying role with respect to the belief that  $q$  will always have the propositional content  $p$ , e.g., S's belief that  $r$  might be a defeater with respect to S's believing that  $q$  on the basis of S's belief that  $p$ .



# CHAPTER 1

## THE NEED FOR A THEORY

In this chapter it is argued that, although it is common to see the concept of defeat employed in epistemic debates, the absence of an explicit discussion of the mechanism of defeat vitiates those applications. If the concept of defeat is to be fruitfully employed in epistemology, a comprehensive theory of defeat is required. This need can be made apparent by looking at the use of the concept in two significant controversies: internalism/externalism about justification and the nature of disagreement and higher-order evidence.

First, the argument has been made that externalism about justification faces a problem in accommodating the phenomenon of defeat, that accommodating defeat can only be accomplished if a theory of justification includes some characteristically internalistic feature. However, no significant attention is paid in that debate to the nature of defeat. We will see that without a theory of defeat in place, the plausibility of the claim that a particular theory does or does not face problems in accommodating the phenomenon of defeat cannot be evaluated.

Second, for those who think that awareness of disagreement or evidence about one's evidence can decrease one's justification for holding a belief there is a temptation to think of such awareness as a defeater with respect to that belief. Others have argued that disagreement and other kinds of higher-order evidence differ in important ways from

defeaters. In both cases, drawing such a contrast is premature in the absence of a theory of how defeaters defeat; it does not obviously follow from the fact that higher-order evidence differs from paragons of defeat that higher-order evidence cannot be a defeater, nor does it obviously follow from the claim that instances of higher-order evidence are defeaters that they cause a loss of justification. It depends on just how defeat functions and just how higher-order evidence functions. Significant attention has been paid to the latter but hardly any has been paid to the former.

### 1. *Defeat & Externalist Theories of Justification*

That we can acquire defeaters with respect to beliefs we justifiably hold is a virtually uncontested feature of our cognitive lives. If this were not so, then once one became justified in believing that, say, it is raining outside, one would not cease to be justified even when faced with clear, sunny skies. Unless one holds that justification entails certainty, such a view is patently implausible. So, an adequate theory of epistemic justification must accommodate defeaters.

Laurence Bonjour (1985, 2003, 2006) and Thomas Grundmann (2009) have argued that externalist theories of justification face a particular difficulty in making this accommodation. At the core of each of their arguments is the claim that externalist theories of justification face a trilemma: either make no special accommodation for defeaters, make an accommodation but provide no motivation for it, or make an accommodation and provide a motivation for it.

In this section we will see that a theory does not face problems in resolving this trilemma merely in virtue of being externalist, nor does a theory escape these problems

merely in virtue of being internalist. Some internalist theories seem to face problems in resolving the trilemma while some externalist theories do not. We will see that the first two horns of the trilemma are equally problematic for the externalist and internalist. We will then see that the third horn is problematic for any theory of justification that is incongruous with the nature of defeat. An example of an internalist theory of justification that appears incongruous and an example of an externalist theory of justification that does not are developed. We will conclude that no full evaluation of theories of justification that take the third horn of the trilemma can be made without a theory of defeat in place.

### 1.1 *The First Horn of the Trilemma*

One of the goals of a theory of justification, perhaps the first goal, is to accurately identify the extension of the concept justification. Any theory that identifies one as justified in believing when one is not justified or as not justified in believing when one is justified misidentifies the extension of that concept and is inadequate in virtue of that.

Consider the following simple externalist theory of justification:

**Simple Externalism:** S is justified in believing that  $q$  if and only if S's belief that  $q$  is the result of a reliable process.

This is an externalist theory of justification because S might be completely unaware of the process by which the belief that  $q$  is formed, let alone the reliability of that process. S cannot tell from the inside whether or not S is justified in believing that  $q$ ; the condition on justification is external to S.

Setting aside whatever other problems this theory might face, this theory appears to be extensionally inadequate. Here's a counter-example from Alvin Goldman:

Suppose that Jones is told on fully reliable authority that a certain class of his memory beliefs are almost all mistaken. His parents fabricate a wholly false story that Jones suffered from amnesia when he was seven but later developed pseudo-memories of that period. Though Jones listens to what his parents have to say and has excellent reason to trust them, he persists in believing the ostensible memories from his seven-year-old past. Are these memory beliefs justified? Intuitively, they are not justified. But since these beliefs result from genuine memory and original perceptions, which are adequately reliable processes, our theory says that these beliefs are justified. (1979, 18)

This is a case of defeat; Jones' memory is unable to justify Jones in forming memorial beliefs because Jones has a defeater with respect to those beliefs in the form of the testimony of Jones' parents. As Goldman points out, our simple externalist theory appears to get this case wrong by classifying Jones' memorial beliefs as justified. Simple externalism takes the first horn of the trilemma introduced above and is extensionally inadequate for it.

Grundmann claims that internalist theories of justification do not face this problem. He writes:

From an internalist point of view defeaters do not pose any problem. According to internalism a belief is justified if the relevant psychological evidence rationally supports the belief. Now, by acquiring further evidence a formerly justified belief may no longer be rationally supported by the resulting total evidence. So, the problem would be solved by adopting some kind of epistemic internalism. (2009,

67)

The suggestion seems to be that the internalist can take the first horn of the trilemma without difficulty because internalist theories of justification essentially accommodate defeaters; no special accommodation is required. This suggestion is false. If an internalist theory classifies a proper subset of one's total evidence as one's relevant psychological evidence then that theory faces problematic cases of defeat. Consider the following simple internalist theory of justification:

**Simple Internalism:** S is justified in believing that  $q$  if and only if S is aware of some evidence that supports the belief that  $q$ .

This is an internalist theory of justification because S must be aware of the evidence that supports the belief that  $q$ . S can tell from the inside whether or not S is justified in believing that  $q$ ; the condition on justification is internal to S.

This theory of justification requires that S be aware of evidence that supports the belief that  $q$  but does not require S's total evidence to support the belief that  $q$ . With slight modifications, Goldman's counter-example from above will apply to simple internalism:

Suppose that Jones is told on fully reliable authority that a certain class of his memory beliefs are almost all mistaken. His parents fabricate a wholly false story that Jones suffered from amnesia when he was seven but later developed pseudo-memories of that period. Though James listens to what his parents have to say and has excellent reason to trust them, he persists in believing the ostensible memories from his seven-year-old past. Are these memory beliefs justified? Intuitively, they are not justified. But since Jones is aware of evidence that supports Jones'

memorial beliefs, i.e., Jones' memories, simple internalism says that these beliefs are justified.

Jones' memories in isolation from the testimony of Jones' parents do support Jones' memorial beliefs and Jones is aware of these memories. So, Jones is aware of evidence that supports Jones' memorial beliefs. Simple internalism gets this case wrong since it classifies Jones' memorial beliefs as justified when, intuitively, they are not.

The simple internalist could, of course, supplement their theory with an account of support such that Jones' memories do not provide support when embedded in a doxastic system containing defeaters, e.g., Jones' belief about the testimony of his parents. This would get the case right but it does not undercut the current point. This addition is inessential to internalism so it remains the case that an internalist theory does not, merely in virtue of being internalist, accommodate defeat. Some further accommodation is necessary. Internalist and externalist theories alike are extensionally inadequate if they take the first horn of the trilemma.

## 1.2 *The Second Horn of the Trilemma*

We have just seen that both simple externalism and simple internalism are extensionally inadequate. As it stands, both theories are composed of only positive conditions, i.e., conditions that lay out what one must accomplish in order to be justified in believing. The theories can be changed to get the cases right by adding a negative condition on justification, i.e., a condition that lays out what one must avoid in order to be justified in believing. One such negative condition requires the absence of defeaters.<sup>4</sup>

---

<sup>4</sup> For an interestingly different approach to negative conditions on justification, see Maria Lasonen-Aarnio (2014).

Consider the following theories:

**Appended Simple Externalism:** S is justified in believing that  $q$  if and only if S's belief that  $q$  is the result of a reliable process and S has no defeaters with respect to the belief that  $q$ .

**Appended Simple Internalism:** S is justified in believing that  $q$  if and only if S is aware of some evidence that supports the belief that  $q$  and S has no defeaters with respect to the belief that  $q$ .

Since Jones has a defeater with respect to his memorial beliefs in the form of his beliefs about his parents' testimony, the appended theories both deliver the verdict that Jones is not justified in forming memorial beliefs. This is the intuitively correct result.

The appended theories of justification get the problematic cases right and so are extensionally adequate, at least as far as these cases are concerned. But accurately identifying the extension of the a concept is not the only goal of a theory of that concept. The theory should also explain why the concept has the extension that it has. It should make understandable the discrimination made by the concept. Failing to do so renders a theory explanatorily inadequate. So, while the appended theories are extensionally adequate, it remains to be seen whether or not, in the absence of a motivation for the inclusion of the negative condition apart from the goal of securing extensional adequacy, the theories are explanatorily adequate.

Michael Bergmann has suggested that no motivation beyond achieving the goal of extensional adequacy is necessary in order for a theory to be explanatorily adequate. The reason he provides for including the negative condition in his externalist theory of

justification is that:

[...] it just seems (perhaps after considering many examples) that a belief isn't justified if one has either a reason for thinking it false or a reason for doubting the reliability of its source - and that this is so whether or not the belief is in fact reliably formed. (2006b, 691)

This is a bald appeal to the intuitive extension of the concept of justification.

As Bonjour (2006, 748-749) points out, Bergmann's position is unsatisfactory. All Bergmann says in support of his position is that it is no worse than the internalist's position. He argues that the reason that Bonjour has given for the inclusion of a negative condition is no different from Bergmann's, that it consists only in an appeal to intuition about the extension of the concept of justification. But, at best, this would show that the theories are equivalent with respect to explanatory adequacy; it cannot show that both are explanatorily adequate. If Bergmann is right about this equivalence, then both theories are, in fact, explanatorily inadequate. An adequate theory of justification must say more about why there is this negative condition on justification than that it is necessary in order to secure extensional adequacy.

To see why, consider how a positive externalist condition on justification might be motivated. An externalist might insist that the concept of justification bears some relationship to the goal of forming true beliefs and that a belief's being formed by objectively reliable belief forming processes plays some important role in that relationship. What the argument for this claim is and what the nature of this relationship might be are irrelevant to our current point. What matters is that this externalist has



provided an explanation as to why justified beliefs must meet a particular condition.

Bergmann has told us that, intuitively, beliefs must satisfy a negative condition in order to be justified, but he has not explained why this might be so.

Perhaps there is a point at which no more fundamental explanations are possible. Perhaps we will find there something that is self-explained, something that is not in need of an explanation, or an unexplained and unexplainable brute fact. But even if this is so, even if we must give up on the search for explanations at some point, we must be careful to not give up on the explanatory goal too soon. There are features of the phenomenon of apparent defeat that seem to demand explanation. The phenomenon of defeat is not simple. Some defeaters are stronger than others. Some defeaters seem to function by supporting the negation of a proposition while others seem to function by removing support for a proposition. If we are to take the negative condition on justification as a brute fact, we must also regard as brute facts these features. But these features suggest an underlying mechanism that is both explicable and enlightening.

This suggestion might be misleading; it might be that defeat is inexplicable. This quietism about defeat should be taken seriously and we will address it at greater length in chapters 2. For now, however, it is enough to point out that quietism about defeat is itself a theory of defeat and that this theory is not warranted simply because it resolves the present trilemma. If quietism is warranted, it must be supported by deeper theoretical reasons. Thus, we can only know whether or not quietism is warranted by theorizing about defeat.

We see, then, that there must be some kind of motivation for the inclusion of a

negative condition in an externalist theory of justification or else that theory is explanatorily inadequate. But we also see that internalist theories that have no motivation for the inclusion of a negative condition are inadequate in the same way. Insofar as the goal of a theory of justification is not just to identify the justificatory status of a belief but also to explain why the belief has that justificatory status, any theory of justification that appeals only to brute intuition to support the inclusion of a condition on justification will be inadequate. It might be that all theories are doomed to be inadequate in this way, but that conclusion can be warranted only by theorizing about defeat and finding no plausible explanations for the phenomenon.

Taking the second horn of the trilemma, then, can only be vindicated by a theory of defeat. Still, perhaps no such theory is necessary; perhaps a motivation for the accommodation of defeat can be found without forming a thorough going theory of defeat, i.e., perhaps a theory of justification can accommodate defeat by taking the third horn of the trilemma.

### 1.3 *The Third Horn of the Trilemma*

In sub-sections 2.1 and 2.2 we saw that theories of justification must accommodate defeat on pain of extensional inadequacy and must motivate this accommodation on pain of explanatory inadequacy. Bonjour and Grundmann argue that the motivation for the negative condition in our appended theories is consonant with internalism but not with externalism. So, the third horn of the trilemma is supposed to be problematic for the externalist in a way that it is not for the internalist. To evaluate this claim, the motivation behind including the negative condition must be made clear.

Goldman's particular form of the negative condition requires that there be no reliable belief forming process available to S such that, had S used this process instead of the process S actually used, S would not have believed that *q*. Recalling the case of Jones above, since Jones has available a reliable belief forming process that would result in Jones not believing that *q*, i.e., listening to Jones' parents, Jones is not justified in forming beliefs on the basis of memory according to Goldman's appended theory of justification. Goldman motivates the inclusion of the condition as follows:

The justificational status of a belief is not only a function of the cognitive processes actually employed in producing it; it is also a function of processes that could and should be employed. (1979, 20)

It is not enough that one's cognitive process be reliable; one must pick the best of the available processes. Bonjour objects to this motivation:

The obvious problem here is how to interpret the suggestion that Jones should have used the alternative cognitive process in question. On the surface this seems to be an appeal to the idea of subjective rationality and as such would favor the alternative position [i.e., an internalist explanation of defeat]. (1985, 48)

Bonjour's objection seems to be that, although Goldman's positive conditions on justification are motivated by externalist considerations, Goldman seems to motivate the inclusion of the negative condition on justification by appeal to internalist considerations.

Bonjour spells out the objection in more detail elsewhere:

Is there any intelligible rationale for the requirement of negative internal justification, which the modified version of externalism accepts, that does not also

support the requirement of positive internal justification, which it rejects? My suggestion is that the only clear reason for the negative requirement is that accepting beliefs that are, as far as one can tell from one's own cognitive perspective, unlikely to be true is plainly irrational and irresponsible from an epistemic standpoint that aims at truth – even if those beliefs happen to be, unbeknownst to the person in question, reliably caused. But this reason plainly supports the positive requirement as well. (2003, 32-33)

BonJour's argument misses its mark. Even if the externalist motivates the inclusion of the negative condition by adverting to the claim that a belief that is unlikely to be true from one's own perspective is plainly irrational and irresponsible, this does not require the inclusion of a positive necessary condition on justification that a belief be likely to be true from one's own perspective. It may be that this claim concerning irrationality would support the inclusion of such a positive condition, but it does not necessitate the inclusion of that condition. As Bergmann (2006b, 691-692) points out, the motivation for the inclusion of the negative condition might also be a reason to include such a positive condition, but that reason might be outweighed by countervailing reasons. So, if the externalist faces a special problem motivating the inclusion of a negative condition, it must be for some other reason.

Grundmann agrees with BonJour that the problem with Goldman's motivation is that it is not in the spirit of externalism. The positive condition, that the belief forming process be reliable, is motivated by something like the connection between justification and truth. Reliability ensures that there is usually some connection between one's justified

beliefs and the truth. But Grundmann argues there is no obvious connection between the processes one could, but does not, employ and truth. He writes:

If we put aside the technical details, reliabilism explains all justificational relevant features as being objectively conducive to the goal of truth. But one does not see, how [Goldman's negative condition] fits into this general picture. [Goldman's negative condition] excludes cases in which someone does not adapt his beliefs to her internally available evidence. But [Goldman's negative condition] does not tell us why this internal adaptation is instrumentally good with respect to the goal of truth. (2009, 70-71)

Unlike Bonjour, who objects to Goldman's apparent use of internal rationality as a motivation for the inclusion of the negative condition without also including an internalist positive condition, Grundmann worries not that the motivation that Goldman employs for the inclusion of the negative condition is internalist nor that the inclusion of the negative condition is unmotivated, but rather that there is a bifurcation between the motivation for the positive condition and the motivation for the negative condition and that this bifurcation is unexplained. His worry is that the inclusion of the positive conditions of Goldman's theory are motivated by appeal to the goal of truth and that the motivation for the inclusion of the negative condition does not fit "into this general picture." Either the inclusion of a negative condition must be motivated by the same considerations that motivate the inclusion of the positive condition or else the difference in motivations must be explained; anything else will result in an explanatorily inadequate theory.<sup>5</sup> We can put

---

<sup>5</sup> Whether or not Grundmann's specific objection to Goldman is successful is beside our current point. Still, Goldman (2009) does argue that there is, indeed, a connection between methods one could but does not use and the truth.

the problem more carefully as follows:

**Bifurcation Problem:** If a theory of justification (i) motivates different conditions on justification in different ways and (ii) provides no explanation for this difference, then (iii) this theory of justification is explanatorily inadequate.

This is an important insight. The unexplained bifurcation in motivations is the true source of the problem that Bonjour and Grundmann have identified, not the fact that appended externalism includes a negative internalist condition but no positive internalist condition, *per se*. The fact that this bifurcation is particularly pronounced in many externalist theories of justification and apparently absent in many internalist theories of justification may be responsible for the mistaken belief that externalist theories of justification face a problem accommodating defeat in virtue of being externalist. But not all internalist theories have a unified motivation for the inclusion of their positive and negative conditions and it is not obvious that all externalist theories must have a bifurcation in these motivations.

Consider appended simple internalism again. The positive condition on justification, that S must be aware of some evidence that supports *q*, and the negative condition on justification, that S must have no defeaters with respect to *q*, might both be motivated by appeal to internal rationality. But different theorists have suggested significantly different standards for internal rationality. An internalist coherentist might tie the standard to the coherence of one's beliefs at a given time, whereas an internalist foundationalist might tie the standard to the validity or strength of the chain of inferences that produced one's beliefs. Were the positive conditions of an internalist theory of

justification motivated by the former standard of internal rationality and the negative conditions of that theory motivated by the latter standard of internal rationality, and were that bifurcation in motivations left unexplained, that theory of justification would have taken the third horn of the trilemma and would be explanatorily inadequate, despite being internalist. Thus, in principle, an internalist theory of justification might face the bifurcation problem.

The internalist theory suggested above might seem an unnatural gerrymander, but that is irrelevant. The point being demonstrated here is that an internalist theory cannot unproblematically take the third horn of the trilemma simply in virtue of being internalist and the theory suggested above demonstrates just that. Furthermore, cases of internalist theories suffering from a problematic bifurcation of motivations exist in the wild. In fact, the *locus classicus* of the literature on defeaters, John Pollock's *Knowledge and Justification* (1974) contains a proposal for just such an internalist theory.

On Pollock's theory, both justifiers and defeaters are reasons for belief. However, where the reasons for belief are themselves beliefs, reasons *qua* defeaters need not be justified in order to be efficacious while reasons *qua* justifiers must be justified in order to be efficacious. One reason that Pollock provides for the details of his negative condition, e.g., his no-defeaters condition, is that “One must proceed on the basis of whatever epistemic connections one sees – it is irrational to do anything else” (1974, 44). The suggestion seems to be that if S believes that S's justification for believing that *q* is defeated, then from S's own perspective it would be irrational for S to continue to believe that *q* and that this is enough to make it irrational *simpliciter* for S to continue to believe

that  $q$ .<sup>6</sup>

Notice, however, that were this motivation applied to the positive condition of Pollock's theory, it would follow that one's reasons *qua* justifiers need not be justified in order to provide justification, either. After all, if S believes that  $p$  entails  $q$  and S believes that  $p$ , albeit without justification, then from S's perspective there is a straightforward epistemic connection to  $q$ . If it is irrational to fail to proceed on the basis of the epistemic connections one sees, then it would be irrational for S to fail to believe that  $q$ , at least, if the question of  $q$ 's truth arose. But Pollock clearly does not accept this:

In order for a person to have a reason for believing something, it must be a good reason, and he must be justified in believing that it is true. So let us define "S has the logical reason P for believing that Q" to mean "(1) S justifiably believes that P, (2) P is a logical reason<sup>7</sup> for S to believe that Q, and (3) P is a good reason for S to believe that Q." (1974, 35)

So, whatever the standard of internal rationality that Pollock employs to motivate the positive condition of his theory, it is not the same as the standard of internal rationality that he employs to motivate the negative condition of his theory, i.e., Pollock's theory of justification satisfies (i) of the bifurcation problem. Furthermore, this bifurcation of motivations goes unexplained and, so, Pollock's internalist theory of justification also satisfies (ii) of the bifurcation problem and, thus, (iii) is explanatorily inadequate.

---

6 There is some reason to think that Pollock might not mean for the principle above to be used to justify the claim that a defeater need not be justified in order to defeat. However, without that principle, it is even less clear why one would accept the kind of asymmetry that Pollock suggests. In any case, it is the asymmetry itself that matters here. Pollock's position is explored in much greater detail in chapter 4.

7 By "logical reason" Pollock means something on the basis of which it is logically possible for one to be justified in believing.



Internalist theories, then, cannot take the third horn of the trilemma without difficulty simply in virtue of being internalist. Conversely, it may be possible for an externalist theory of justification to take the third horn without facing the bifurcation problem. Consider a theory of justification that asserts that S is justified in believing that  $q$  if and only if S's belief that  $q$  objectively coheres to S's other beliefs. Objective coherence is not an internally accessible feature of beliefs and so this would be a form of externalism. On this theory, there seems to be one explanation for both why some beliefs increase the justification for believing something and why some beliefs decrease the justification for believing something. Since, on this theory, a belief is justified to the extent that it objectively coheres with one's other beliefs, beliefs that increase that coherence increase the justification for the belief and beliefs that decrease that coherence decrease the justification for the belief. Such a theory would not satisfy condition (i) of the bifurcation problem and thus, while this theory may face insuperable problems elsewhere, the possibility of such a theory demonstrates that an externalist theory need not obviously include problematically bifurcated motivations for its positive and negative conditions merely because it is externalist. If there is some feature of externalism that guarantees a bifurcation in motivations for the positive and negative conditions, it is not obvious what this feature is.

#### 1.4 *The Need for a Theory*

We have shown that if externalist theories cannot adequately accommodate defeat, it is not because they are externalist, *per se*, but rather because no adequate explanation for the inclusion of both the positive and negative condition on justification can be

provided. We have also shown that internalist theories can fail in precisely the same way, though we have not attempted to show that they must. Accommodating the phenomenon of defeat is a necessary condition on forming an adequate theory of justification, but this accommodation cannot be made unless the phenomenon of defeat can be explained within that theory of justification.

The lesson here is not that internalist or externalist theories do or do not face special problems in accommodating defeat. Rather, the lesson is that we cannot evaluate whether or not a theory of justification adequately accommodates defeat until we have a better idea of the nature of defeat. If, after exploring the phenomenon of defeat, we find that there are good reasons for thinking that different standards of rationality apply to acting on defeaters than apply to acting on justifiers, Pollock's bifurcated theory of justification might be adequate after all. Furthermore, if we find that some features of defeat cannot be explained by appeal to objective coherence alone, the objective coherentist theory suggested above might not be adequate, even though it motivates its positive and negative conditions by appeal to a common underlying principle. These theories simply cannot be evaluated in the absence of a theory of defeat.

## 2. Higher-Order Evidence

The study of the epistemic consequences of disagreement and the effects of evidence about our evidence, i.e., higher-order evidence, on the justification for our beliefs has increased dramatically in the last decade.<sup>8</sup> These issues are particularly pressing for philosophers since there is hardly a philosophical claim one could make that

---

8 See Thomas Kelly (2005) for the *locus classicus* of the recent literature on disagreement.

is not denied by someone else who has thought just as long and hard about the topic. As

Peter van Inwagen writes:

How can I believe (as I do) that free will is incompatible with determinism or that unrealized possibilities are not physical objects or that human beings are not four-dimensional things extended in time as well as in space, when David Lewis - a philosopher of truly formidable intelligence and insight and ability - rejects these things I believe and is already aware of and understands perfectly every argument that I could produce in their defense? (1996, 138)

If disagreement reduces one's justification, then philosophical beliefs are rarely, if ever, justified. Furthermore, although worries about the epistemic status of our bases for belief do not often occur to most, they are never far from the thoughts of philosophers (or at least epistemologists).

The way in which defeat might be relevant to this area of study is fairly obvious. Defeat is closely connected to the loss of justification and, the worry is, peer disagreement and higher-order evidence might cause a loss of justification. Naturally, one might speculate on what relationship, if any, defeaters have to disagreement and higher-order evidence.

Defeat enters into the conversation on higher-order evidence and disagreement<sup>9</sup> in at least two ways. First, it has been argued that instances of higher-order evidence are defeaters and, thus, that they cause a loss of justification or, perhaps, that higher-order evidence causes a loss of justification and, thus, instances of higher-order evidence are defeaters; the direction of the support is not altogether obvious. Second, it has been

---

9 Which we will refer to as simply "higher-order evidence" after this point.

argued that the effects of higher-order evidence differ significantly from those of defeat and, thus, cannot be accommodated without new theories of justification or, at least, modifications to extant theories of justification.

In both of these cases the appeal to the concept of defeat is premature. We will see that in the first case, the argument fails because, although it is difficult to see just what the argument is supposed to be, once the structure of the argument is made clear and we see what sort of relation is claimed to hold between higher-order evidence and defeat, it becomes plain that without a theory of defeat in place we are unwarranted in asserting that this connection exists.

We will see that in the second case the argument fails because the accommodation of ordinary defeat by extant theories of justification is, as demonstrated in section 2 above, already problematic. We saw that we need a theory of defeat in order to determine whether or not such accommodation remains problematic. Prior to such a theory, it is simply impossible to conclude that higher-order evidence necessitates special accommodation above and beyond that necessary for standard cases of defeat. After a theory of defeat has been constructed, there are three possibilities. First, it might turn out that both defeat and higher-order evidence can be accommodated by extant theories of justification. Second, it might be that extant theories of justification cannot handle ordinary defeat and, thus, traditional theories of justification are overturned by the phenomenon of ordinary defeat as much as they are by higher-order evidence. Third, it might be that extant theories of justification can handle ordinary defeat but cannot handle higher-order evidence and, thus, those who argue that higher-order evidence requires

special accommodation are vindicated.<sup>10</sup> Which of these outcomes will be realized depends on the theory of defeat and cannot be determined prior to it.

### 2.1 *Identifying Higher-Order Evidence with Defeat*

Richard Feldman (2005) defends the claim that when it comes to higher-order evidence, we should “respect our evidence.” What that means is that when one comes to have evidence that one's evidence for believing that  $q$  is not good one should revise one's beliefs because such evidence causes a loss of justification for one to believe that  $q$ .

Imagine, for example, the following case developed by David Christensen:

**Deductive Drug:** I’m asked to be a subject in an experiment. Subjects are given a drug, and then asked to draw conclusions about simple logical puzzles. The drug has been shown to degrade people’s performance in just this type of task quite sharply. In fact, the 80% of people who are susceptible to the drug can understand the parameters of the puzzles clearly, but their logic-puzzle reasoning is so impaired that they almost invariably come up with the wrong answers.

Interestingly, the drug leaves people feeling quite normal, and they don’t notice any impairment. In fact, I’m shown videos of subjects expressing extreme confidence in the patently absurd claims they’re making about puzzle questions. This sounds like fun, so I accept the offer, and, after sipping a coffee while reading the consent form, I tell them I’m ready to begin. Before giving me any pills, they give me a practice question:

Suppose that all bulls are fierce and Ferdinand is not a fierce bull.

---

<sup>10</sup> There is also the possibility that extant theories can handle higher-order evidence but cannot accommodate standard defeaters. This possibility is not explored.

Which of the following must be true? (a) Ferdinand is fierce; (b) Ferdinand is not fierce; (c) Ferdinand is a bull; (d) Ferdinand is not a bull.

I become extremely confident that the answer is that only (d) must be true. But then I'm told that the coffee they gave me actually was laced with the drug. My confidence that the answer is "only (d)" drops dramatically. (2010, 187)

Feldman argues that this is the correct response; one's confidence in their answer should diminish in the face of this higher-order evidence. This is so, Feldman contends, because being told that the coffee was laced with the drug defeats one's justification for one's answer.

Feldman's position can be broken into two parts: first, (i) if S is antecedently justified in believing that  $q$ , then S's higher-order evidence to the effect that her evidence for believing that  $q$  is no good renders S unjustified in believing that  $q$  and, second, (ii) S's higher-order evidence in this case constitutes a defeater with respect to S's belief that  $q$ .

The relationship that Feldman takes to hold between (i) and (ii) is not transparent. The most obvious way to interpret it is as (i) being reason to accept (ii). However, Feldman's topic is higher-order evidence, not defeaters. It is unclear, then, why he would spend any time on (ii). The interpretation on which (i) is supposed to support (ii) becomes still less plausible in light of Feldman's response to the objection that higher-order evidence has effects that differ from the effects of standard cases of defeat. According to this objection, ordinary cases of defeat challenge the connection between the evidence and the belief by raising doubts about the trustworthiness in a specific circumstance of a

generally reliable evidential connection. For example, having reason to believe that there is a hidden red light shining on the desk in front of you is reason to not rely on the fact that this desk looks red in forming your beliefs about the color of the desk; it is not a reason to not rely on color appearances in forming beliefs about the colors of things generally. Higher-order evidence on the other hand provides reason to mistrust the general connection rather than the application of that connection in a specific instance (2005, 112-113).

If (i) is supposed to provide reason to believe that (ii) is true, but what Feldman is really concerned about is the truth of (i), then the obvious response to this objection would be to concede that instances of higher-order evidence might not be defeaters, i.e., abandon (ii), but insist that they still cause a loss of justification, i.e., retain (i). In other words, Feldman could agree that instances of higher-order evidence are not defeaters but deny that defeaters are the only things that can result in a loss of justification. Feldman does not do this; instead, he maintains that, although the effects of higher-order evidence might differ from the effects of standard cases of defeat, there is no reason to deny that instances of higher-order evidence are defeaters.

It is worth noting that already it is apparent that Feldman's argument is in need of a theory of defeat. To assert that, despite this apparent difference, instances of higher-order evidence remain defeaters without providing any supporting reasons is to simply assert one's intuitions about what is or is not a defeater. This gives the objector no reason to change their mind since, obviously, they don't share that intuition. What is needed here is a theory of how defeat works that makes it clear that, although there might be

superficial differences between the effects of higher-order evidence and the effects of standard defeaters, there nonetheless remains fundamental commonalities in virtue of which instances of higher-order evidence are best understood as defeaters.

Given the above considerations, the most plausible interpretation of Feldman's argument is one on which (ii) is supposed to support (i). If this were true, then Feldman would have to defend (ii) on pain of losing support for (i). This interpretation, then, makes Feldman's response to his objector consonant with the stated topic of Feldman's paper, the epistemic effects of higher-order evidence.

Furthermore, Feldman asserts (2005, 103-104) that the following two claims are different ways of making the same point:

- (1) (E&D) does not provide good support for T.
- (2) D defeats one's justification, E, for T.

Applied to the irrationality-inducing drug case, E is the evidence provided by ratiocination about the logical puzzle, T is the belief in the answer, and D is one's awareness that one has been told that the coffee was laced with the irrationality-inducing drug. Perhaps the reasoning, then, is this:

- (P1) (E&D) does not provide good support for T
- (P2) If (E&D) does not provide good support for T, then D defeats one's justification, E, for T
- (P3) If D defeats one's justification, E, for T, then D causes one to cease to be justified in believing T
- (C1) So, D causes one to cease to be justified in believing T



Whether or not we have good reason to accept (P1) is not at issue here. What is at issue is what Feldman says, or rather what he doesn't say, in defense of (P2) and (P3). He, like most epistemologists, seems to simply assume the truth of substantive claims about the nature of defeat, writing:

One use for the word “defeater” in epistemology applies here. Roughly, a proposition X is a justification defeater for proposition P for a person provided the person was justified in believing P prior to becoming justified in believing X but as a result of becoming justified in believing X, the person is no longer justified in believing P. Where Y is the person’s original evidence for P, and Y adequately supports P, the conjunction of X and Y fails to support P. (2005, 104)

Note that Feldman actually runs together two different accounts of defeat. He first suggests that a defeater is any proposition that causes a loss of justification. This is a causal theory of defeat. This would explain why he might think that (P3) is true. He then suggests, with no defense, that a defeater is a proposition that somehow negates one's evidence. This is a reason-neutralizing theory of defeat. This would explain why he might think that (P2) is true.

It may be that the causal theory of defeat and the reason-neutralizing theory of defeat are coextensive, but that is not obviously the case. We will explore both of these theories in chapter 2 and ultimately reject them, but note here just that Feldman seems to have made substantive assumptions about the nature of defeat in support of his theory of disagreement. These assumptions are unwarranted. If there is a path that leads from (ii) to (i), a theory of defeat is needed to find it.

A natural worry to have at this point is that (ii) might be a merely terminological point. After all, Feldman asserts only that this is one use for the word “defeater.” He, presumably, is more concerned with the phenomena of the loss of justification and the neutralizing of reasons than he is about what those phenomena should be called. Perhaps we are making a mountain of a mole-hill.

This objection is misguided. Recall that our goal is to develop a theory of defeat that sheds light on the underlying mechanism of defeat. If it turns out that the loss of justification and the neutralizing of reasons are phenomena that arise from the nature of defeat but are distinct from it, and that some beliefs can cause a loss of justification without sharing that nature or that some beliefs can share in that nature without neutralizing reasons, as we will see in chapter 3, then Feldman is tying his theory of higher-order evidence to phenomena without a single underlying epistemic nature. Just as it is a mistake to think of the Gettier debate as merely terminological,<sup>11</sup> so too is it a mistake to think of the current objection as such. We want theories that are unified and illuminating; tying one's theory of higher-order evidence to phenomena that are neither is problematic.

## 2.2 *Distinguishing Higher-Order Evidence from Defeat*

*Contra* Feldman, Christensen (2010) argues that, whatever similarities higher-order evidence might bear to standard defeaters, the effects of higher-order evidence differ in some significant respects from those of defeaters. Christensen maintains that a standard defeater with respect to S's belief that  $q$  changes the weight of

---

11 See William Lycan (2006) and Timothy Williamson (2000, 30)

S's evidence with respect to  $q$ . If S comes to have reason to believe that not- $q$  or that special circumstances apply that make unreliable an otherwise reliable indicator of  $q$ , the support S's evidence provides for believing that  $q$  is thereby lessened. Recall the example of hidden red light and the desk used above. In that case, coming to have reason to believe that a red light is shining on the desk decreases the evidential support that the desk appearing red provides for one's belief about the color of the desk, Christensen asserts.

Higher-order evidence with respect to S's belief that  $q$ , on the other hand, does not affect the weight of S's evidence with respect to  $q$ . Rather, higher-order evidence gives S reason to “*put aside or bracket* [her] original reasons for [believing that  $q$ ]” (2010, 195). S's evidence still supports her believing that  $q$ , but her higher-order evidence makes her use of that support somehow illicit. Recall the irrationality-inducing drug case from above. In that case Christensen maintains that:

[...] there is a clear sense in which the facts which are not in doubt - the parameters of the puzzle - leave no room for anything other than my original answer. Or, to put it another way, the undoubted facts support my answer in the strongest possible way - they entail my answer - and this kind of connection cannot be affected by adding more evidence. Moreover, I even correctly see the entailment, and initially believe my answer in virtue of seeing the entailment.

(2010, 195)

Being told that one's coffee was laced with the drug, Christensen suggests, cannot affect the connection between one's evidence and one's answer in this case. So, it must be that

defeaters and higher-order evidence function in fundamentally different ways and we need a new epistemic mechanism, bracketing, to explain this difference.

This conclusion is not warranted. It is true that nothing can sever the entailment between one's evidence and one's answer in the irrationality-inducing drug case above, but it does not follow from this that defeaters and higher-order evidence differ significantly. Christensen, following Pollock, accepts the claim that some standard defeaters function by attacking the connection between one's evidence and one's belief (2010, 193-194). First, the details of this view are sketchy. As we will see in chapter 2, it is not at all clear what this attack on evidential connections amounts to. If Christensen's claim is to be warranted, Pollock's theory of defeat must first be made clear.

Furthermore, even if Christensen were right about higher-order evidence differing from some standard defeaters in this respect, that still wouldn't warrant the conclusion that higher-order evidence differs significantly from standard defeaters *simpliciter*. Some defeaters with respect to the belief that  $q$  seem to function by being reasons to believe that not- $q$ . Defeaters of this sort would raise no immediate doubts about the connection between one's evidence and the belief that  $q$ . So, there are already clear examples of standard defeaters that do not challenge evidential connections. Even if higher-order evidence differs from those standard defeaters that do raise such challenges, that does not give us reason to believe that they differ from those that do not. Christensen would have to explain how bracketing differs from whatever it is we do with our evidence for believing that  $q$  when we acquire reason to believe that not- $q$ . In order to do that, he would need to reference an account of just what we do in such cases, and that is a part of

the project of providing a theory of defeat.

None of this is to say that Christensen is wrong. It may well turn out that the best theory of defeat we can find will support his contention. We are suggesting, as before, that without careful theorizing about the fundamental nature of defeat, his conclusion is premature. We need a theory of defeat before we try to use the concept to settle epistemic controversies.

### 3. *Conclusion*

In this chapter we've seen that putting the concept of defeat to work in evaluating theories of justification or settling philosophical controversies, at least in the case of the proxies that we've picked, is bound to be fruitless prior to the development of a thoroughgoing theory of defeat. A careful exploration of the theories of defeat, such as they are, is required to lay the groundwork and set the parameters for such applications.

In the next chapter, we introduce three theories of defeat that can be found in the literature: the causal theory of defeat, the reason-neutralizing theory of defeat, and quietism about defeat. We will see that all three of these theories are inadequate. However, although they all fail, their failure is instructive. In the remainder of the dissertation a new theory of defeat is developed (chapter 3), defended (chapter 4), and applied to the controversies considered in this chapter (chapter 5).

# CHAPTER 2

## THREE THEORIES OF DEFEAT

Surprisingly, relatively little has been written concerning just what defeaters are or just how they function. Defeat is alluded to, mentioned, and noted extensively in contemporary epistemology, but rarely is it discussed as a topic in its own right. As we demonstrated in the previous chapter, even when the nature of defeat is central to their arguments, philosophers have shied away from any attempt provide an explanation of how defeat works. Instead, their claims rest on intuitions concerning whether or not a subject ceases to be justified in holding a given belief in this or that thought-experiment. These intuitions are simply not robust enough to support the work they're being put to, and when intuitions conflict with one another an underlying theoretical framework is needed to settle the dispute.

However, even if our intuitions about these thought-experiments were probative, it would still be unsatisfactory to leave the issue at that intuitive level. If we are to understand a feature so central to our cognitive lives we must ask not just whether but also why and how one's justification for believing some proposition is defeasible. A philosophical theory of defeat is needed.

In this chapter, we will develop three theories about the mechanism of defeat that are suggested, if sometimes only vaguely, by the literature. All three of these theories are rejected on the grounds that they are theoretically inadequate. However, the failure of

these theories is instructive, and in the next chapter we will develop a theory of defeat that avoids their mistakes.

### 1. *The Theories*

The first of the three theories that we will explore, which we can call the causal theory of defeat, asserts that the defining characteristic of a defeater is that it causes a loss of justification:

**The Causal Theory:** S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if S's believing that  $r$  causes S to become unjustified in believing that  $q$ .

The second theory, which we can call the reason-neutralizing theory of defeat, asserts that the defining characteristic of a defeater is that it is, somehow, able to neutralize the epistemic effect of one's reasons for belief:

**The Reason-Neutralizing Theory:** S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if S's belief that  $p$  is a reason for S to believe that  $q$  and S's belief that  $(p \& r)$  would not be a reason for S to believe that  $q$ .

The third theory, which we can call the quietist theory of defeat, asserts that there is no underlying explanation for the function of defeaters:

**The Quietist Theory:** Whether or not S's belief that  $r$  is a defeater with respect to S's belief that  $q$  is a fundamental epistemic fact and, thus, inexplicable.

We will see that there are counter-examples to both the causal theory and the reason-neutralizing theory and that they are thus extensionally inadequate. Neither theory is sufficiently sensitive to the importance of the relationship between the defeater and the

source of justification.

We will see further that all three theories are unable to explain why S's belief that  $r$  is a defeater with respect to S's belief that  $q$  and, as such, are explanatorily inadequate. None of the theories provide satisfactory answers to the question, "Why and how does a defeater defeat?"

## 2. *Easy Cases*

The cases that will best illustrate the differences between the three target theories of defeat are ones in which there is a difference between the epistemic agent's propositional justification and her doxastic justification. A helpful way of thinking about the distinction between propositional and doxastic justification is in terms of epistemic support provided as contrasted with epistemic support being utilized. When a belief is propositionally justified, one has epistemic support available for that belief, whether or not one is utilizing that support. When a belief is doxastically justified, one has epistemic support available for that belief and one is utilizing that support.

Let's begin by looking at a case that illustrates the difference between propositional and doxastic justification:

- (T<sub>1</sub>) S believes that S's atlas shows that Corsica is north of Sardinia. S also believes that S's friend asserted that Corsica is north of Sardinia. S believes that Corsica is north of Sardinia solely on the basis of her belief that her atlas shows Corsica being north of Sardinia.

The standard account of justification, which we will accept uncritically at this point, asserts that both S's atlas and her friend's testimony provide support for S's belief that



Corsica is north of Sardinia, i.e., that they both propositionally justify this belief, but since S has formed her belief that Corsica is north of Sardinia on the basis of her atlas alone and ignored her friend's testimony, only the support provided by S's atlas is being utilized in S's belief that Corsica is north of Sardinia, i.e., that only S's atlas is doxastically justifying S's belief.<sup>12,13</sup>

In other words, in (T<sub>1</sub>) S is doxastically justified in believing that Corsica is north of Sardinia. S would be doxastically justified in believing that Corsica is north of Sardinia were S to believe that Corsica is north of Sardinia on the basis of her friend's testimony instead. Regardless of her basis for believing that Corsica is north of Sardinia, S is propositionally justified in believing that Corsica is north of Sardinia.

So, the above will be a basic case of justification that will be referred to throughout. We will use the cases in which that justification is or is not lost to distinguish between two different kinds of defeat and to illustrate features and failings of the three theories introduced above. However, we are not thereby implying that the loss of justification is necessary, sufficient, or even central to defeat. In fact, we expressly deny that it is any of those things and demonstrate why below. Loss of justification is just the most conspicuous consequence of defeat. For now, take the following cases merely as useful illustrations of defeat.

In our first case of defeat, S loses doxastic justification for believing that Corsica is north of Sardinia but retains propositional justification:

---

12 The relationship between epistemic reasons, epistemic support, propositional justification, and doxastic justification is more complicated than this and it will be given some attention in the next chapter.

13 For ease of presentation, we will sometimes refer to something other than a mental state as a reason for belief, e.g., "S's atlas" instead of "S's belief that S's atlas shows that Corsica is north of Sardinia." This is merely a convenience and is not meant to imply a commitment to any particular view about what can or cannot count as an epistemic reason.

(T<sub>2a</sub>) S comes to believe that S's atlas is unreliable.

When S comes to believe that her atlas is unreliable, this removes the justification provided by her atlas but leaves intact the justification provided by her friend's testimony. However, since S's belief that Corsica is north of Sardinia is based on S's atlas alone, her belief that Corsica is north of Sardinia rests only on the justification provided by the atlas. So, this belief is left unjustified.

In other words, in (T<sub>2a</sub>) S is not doxastically justified in believing that Corsica is north of Sardinia. In (T<sub>2a</sub>) S would be doxastically justified in believing that Corsica is north of Sardinia on the basis of her friend's testimony. Regardless of S's basis for believing that Corsica is north of Sardinia, S is propositionally justified in believing that Corsica is north of Sardinia in (T<sub>2a</sub>) in virtue of the fact that S believes that her friend asserted that Corsica is north of Sardinia.

In our second case of defeat, S loses both propositional justification and doxastic justification for believing that Corsica is north of Sardinia:

(T<sub>2b</sub>) S comes to believe that S's teacher asserted that Corsica is not north of Sardinia.

Imagine that, instead of coming to believe that the atlas is unreliable, S comes to believe that her teacher contradicted both what the atlas indicates and what her friend asserts. In this situation, it doesn't matter whether S believes that Corsica is north of Sardinia on the basis of her atlas or on the basis of her friend's testimony; her belief would be rendered unjustified either way.

Note that S's teacher has not really weakened the support the atlas and the

testimony of S's friend provide, but rather has put pressure on S's belief that Corsica is north of Sardinia in such a way that the atlas and the testimony of her friend no longer provide enough support to act as a basis for that belief.<sup>14</sup> So, neither her atlas nor her friend's testimony provide adequate<sup>15</sup> support for her belief that Corsica is north of Sardinia in light of her teacher's assertion.

In other words, in (T<sub>2b</sub>) S would not be doxastically justified in believing that Corsica is north of Sardinia regardless of the basis for that belief. Furthermore, regardless of S's basis for believing, S is not propositionally justified in believing that Corsica is north of Sardinia in (T<sub>2b</sub>).

Usually, the difference between the kinds of defeat illustrated above would typically be characterized in terms of type II and type I defeat, undercutting and rebutting defeat, or undermining and rebutting defeat respectively.<sup>16</sup> We do not use these terms because they presuppose distinctions between types of defeat prior to a robust theory of how defeat functions. Such assumptions are unwarranted. For now, simply note that the distinction above is not intended to be the typical distinction, even if it coincides with the way that distinction.

### 3. *The Causal Theory of Defeat*

The causal theory of defeat asserts that S's belief that *r* is a defeater with respect to S's belief that *q* if and only if S's believing that *r* causes S to become unjustified in

---

14 These vague metaphors of support and pressure will be made explicit and precise in the next chapter.

15 The term "adequate" here is significant. Our considered view is that defeaters do not change how much support one's reasons provide for one to believe but rather the standard for what counts as adequate support. More is made of this in chapter 3.

16 See Pollock (1974), Pollock and Cruz (1999), and Casullo (2003) for examples of each.

believing that  $q$ . The clearest example of this position can be found in Michael Bergmann's account of justification and defeat. Bergmann writes :

What are defeaters? The basic idea is that they are mental states of a person, S, that cause a justified belief of S to become unjustified. (2006a, 155)

A similar account can be found in Richard Feldman's discussion of disagreement.

Feldman writes:

Roughly, a proposition X is a justification defeater for proposition P for a person provided the person was justified in believing P prior to becoming justified in believing X but as a result of becoming justified in believing X, the person is no longer justified in believing P. (2005, 104)

Thomas Grundmann's entry on defeasibility theory in *The Routledge Companion to Epistemology* contains a causal characterization of defeat as well:

Defeaters have an effect on the epistemic status of a belief. They wholly or partly remove the rationality (i.e. the justification or warrant) that individual beliefs formerly did possess. Defeaters are thus local “removers of rationality.” (2011, 156)

Beside the explicit endorsement of the causal theory by these authors, it is implicitly endorsed at many points in the literature when authors move from the claim that a belief causes a loss of justification to the claim that the belief is a defeater.

### 3.1 *The Cases It Gets Right*

The causal theory has some intuitive appeal. After all, loss seems to be central to the folk notion of defeat *simpliciter*, e.g., if team A defeats team B, then team B loses the

match, and so it makes sense to try to explain defeaters for epistemic justification in terms of that which causes a loss of justification. Furthermore, the causal theory gets the cases of (T<sub>2a</sub>) and (T<sub>2b</sub>) right:

(T<sub>2a</sub>) S comes to believe that S's atlas is unreliable.

Recall that S's belief that Corsica is north of Sardinia is based exclusively on S's atlas. So, S's coming to believe that her atlas is unreliable causes S to cease to be doxastically justified in believing that Corsica is north of Sardinia. So, the causal theory classifies S's belief that her atlas is unreliable as a defeater, which is the intuitively correct result.

Similarly:

(T<sub>2b</sub>) S comes to believe that S's teacher asserted that Corsica is not north of Sardinia.

Since S's coming to believe that her teacher contradicted her atlas and her friend's testimony causes S to both cease to be doxastically justified in believing that Corsica is north of Sardinia and to cease to be propositionally justified in believing that Corsica is north of Sardinia, the causal theory classifies the teacher's assertion as a defeater, which is also the intuitively correct result.

### 3.2 *The Cases It Gets Wrong*

Though it may be able to handle these standard cases, it is appealingly simple, and it has connections to the folk notion of defeat *simpliciter*, the causal theory will not work. Some defeaters remove neither S's propositional justification nor S's doxastic justification for believing that Corsica is north of Sardinia. Imagine that we're back in (T<sub>1</sub>), before S came to doubt her atlas's reliability or hear her teacher's contradictory assertion. At this

point, S believes that Corsica is north of Sardinia on the basis of her atlas alone. Now, imagine that:

(T<sub>3</sub>) S comes to believe that her friend is unreliable with respect geography.

In (T<sub>3</sub>) S is doxastically justified in believing that Corsica is north of Sardinia since S believes this not on the basis of her friend's testimony but rather on the basis of her atlas. Her new belief that her friend is unreliable is only relevant to her friend's testimony, the support provided by which is not being utilized by S's actual belief that Corsica is north of Sardinia.

In (T<sub>3</sub>) S would not be justified in believing that Corsica is north of Sardinia on the basis of her friend's testimony, but since S was not believing that Corsica is north of Sardinia on the basis of her friend's testimony, this does not represent a loss of actual justification between (T<sub>1</sub>) and (T<sub>3</sub>). Regardless of S's basis for believing that Corsica is north of Sardinia, S remains propositionally justified in believing that Corsica is north of Sardinia in virtue of her atlas.

(T<sub>1</sub>) is a case of epistemic overdetermination,<sup>17</sup> i.e., S's propositional justification for believing that Corsica is north of Sardinia is assured by S's atlas and also by S's friend's testimony, such that were S to lose the support provided by her atlas or her friend's testimony, but not both, as she does in (T<sub>2a</sub>) and (T<sub>3</sub>) respectively, S would remain propositionally justified in believing that Corsica is north of Sardinia. S's belief that her friend is unreliable with respect to geography in (T<sub>3</sub>) is clearly similar in a significant way to S's belief that her atlas is unreliable in (T<sub>2a</sub>). In both cases, S loses a source of

---

17 See Albert Casullo (2005) for a discussion of the relevance of epistemic overdetermination to defeasibility conditions on *a priori* justification.

support. Any account of defeaters that counts S's belief that her atlas is unreliable in (T<sub>2a</sub>) as a defeater with respect to S's belief that Corsica is north of Sardinia but does not count S's belief that her friend's testimony is unreliable in (T<sub>3</sub>) as a defeater with respect to S's belief that Corsica is north of Sardinia, as the causal theory does, is inaccurate.

Because of cases of epistemic overdetermination, it is misleading to speak of one's justification for a belief being defeated. Cases of epistemic overdetermination demonstrate that the loss of justification in a case of defeat, when it occurs, is the result of a more fundamental loss, the loss of adequate support from this or that source. To make this point clearer, consider the following:

(T<sub>4</sub>) S comes to believe that things will go badly for her if she believes that her atlas shows that Corsica is north of Sardinia or that her friend asserted that Corsica is north of Sardinia. Because of S's belief that about things going badly for her, S ceases to believe that her atlas shows that Corsica is north of Sardinia and that her friend asserted that Corsica is north of Sardinia.

In (T<sub>4</sub>) S is not doxastically justified in believing that Corsica is north of Sardinia since S has ceased to believe that her atlas shows that Corsica is north of Sardinia. Furthermore, in (T<sub>4</sub>) S could not be justified in believing that Corsica is north of Sardinia on the basis of S's belief that her friend asserted that Corsica is north of Sardinia since S has ceased to believe that, too. Regardless of S's basis for believing that Corsica is north of Sardinia, S is not propositionally justified in believing this in (T<sub>4</sub>) since S has ceased to hold those beliefs that were providing S with support for believing that Corsica is north of Sardinia.<sup>18</sup>

---

18 Assuming that S's belief that  $p_1$  and S's belief that  $p_2$  were the only beliefs that made S propositionally

This is a case of an epistemically deviant causal chain running from S's belief about how things will go for her depending on how she manages her doxastic life to the loss of S's justification for believing that Corsica is north of Sardinia. If an account takes the central notion of defeat to be that of loss of justification, then S's belief that her atlas is unreliable in (T<sub>2a</sub>) would be more similar in the relevant respects to S's belief about how things will go for her depending on how she manages her doxastic life in (T<sub>4</sub>) than it would be to S's belief that her friend's testimony is unreliable in (T<sub>3</sub>). But this is false; S's belief about how her life will go is not a defeater with respect to S's belief that Corsica is north of Sardinia despite the fact that it causes S to cease to be justified in believing this.

After all, were S to cease to believe that her atlas indicates that Corsica is north of Sardinia in (T<sub>3</sub>), S's belief that her friend's testimony is unreliable would indeed cause S to cease to be propositionally justified in believing that Corsica is north of Sardinia because it would render inadequate the support that S's friend's testimony provides for S to believe that Corsica is north of Sardinia. In other words, at (T<sub>3</sub>), S's belief that her friend's testimony is unreliable does render inadequate the support that her friend's testimony provides even if, because S was not utilizing that support, it doesn't remove any actual justification. If rendering the support provided by a given source inadequate is what makes that belief a defeater in the counter-factual version of (T<sub>3</sub>) where S does not believe that her atlas indicates that Corsica is north of Sardinia, then it also makes it a defeater in the actual version of (T<sub>3</sub>) since it still renders inadequate the unutilized support that S's friend's testimony provides for S to believe that Corsica is north of Sardinia.

---

justified in believing that *q*. We'll stipulate that they are.



At this point, the proponent of the causal theory might object that in (T<sub>3</sub>), S's belief that her friend's testimony is unreliable does cause S to lose justification for believing that Corsica is north of Sardinia even if it does not cause S to become unjustified in believing that Corsica is north of Sardinia. Perhaps the problem is that the causal theory as stated is a theory of what Bergmann calls "full defeaters" (Bergmann 2006a, 155ff6), whereas S's belief that her friend's testimony is unreliable is only a partial defeater; it causes a loss of some of S's justification for believing that Corsica is north of Sardinia but does not cause S to become unjustified *simpliciter* in believing this. So, the causal theory of defeat would, this objection asserts, classify S's belief that her friend's testimony is unreliable as a partial defeater with respect to S's belief that Corsica is north of Sardinia but would not classify it as a full defeater with respect to this belief. This is meant to be an intuitively plausible result.

This objection fails, if for no other reason then because it does not correct the causal theory's misidentification of S's belief about how her life will go depending on how she orders her doxastic life as a defeater with respect to S's belief that Corsica is north of Sardinia in (T<sub>4</sub>). So, even if it does correct the misidentification in (T<sub>3</sub>), (T<sub>4</sub>) remains a counter-example. However, the objection does not even correct the misidentification in cases like (T<sub>3</sub>) because it depends on the false assumption that the justification provided by all sources of epistemic support is cumulative.

If the justification provided by distinct sources of epistemic support is not always cumulative, then (T<sub>3</sub>) need not be a case where S loses any justification at all to believe that Corsica is north of Sardinia. In order for the objection to work, it must be that if S's

belief that  $p_n$  provides support for S to believe that  $q$  and S's belief that  $p_m$  provides support for S to believe that  $q$ , then, if S's belief that  $p_n$  and S's belief that  $p_m$  are not the same source of support, S has more justification for believing that  $q$  than S would have were S to believe either that  $p_n$  or that  $p_m$  but not both that  $p_n$  and that  $p_m$ . If this principle of cumulative justification is not true, then it will be possible to construct cases, of which perhaps (T<sub>3</sub>) is one, in which S's belief that  $r$  is a defeater with respect to S's belief that  $q$  but it does not cause S to lose any amount of justification for believing that  $q$ .

If you believe that one hundred equally reliable people have told you that  $q$ , then coming to believe that another person has told you that  $q$  would not increase your justification for believing that  $q$  at all. Alternatively, if you've checked your stove one hundred times to make sure it is off, then checking it another time will not increase your justification for believing that the stove is off. At least, that's how it seems. But, the defender of the causal theory might continue, our intuitions in these cases are not probative since they are not able to distinguish between no increase in justification and a marginal increase in justification. These might be cases in which your justification is increased, but only to such a small degree that it might as well not have been increased at all.

Even if you don't share the suggested intuitions about these cases or you take our intuitions about these cases to be non-probative, there are counter-examples to the principle of cumulative justification. Imagine that:

Sigurd believes that her friend, Urik, asserted that Corsica is north of Sardinia.

Sigurd also believes that her other friend, Egil, said, "Urik is not a reliable source

with respect to whether or not Corsica is north of Sardinia but, as a matter of fact, Corsica is north of Sardinia.” Egil is, and is believed to be, exactly as reliable as Urik is, or was believed by S to be,<sup>19</sup> prior to Egil's testimony.

Egil's testimony simultaneously renders inadequate whatever support Urik's testimony might have provided for Sigurd to believe that Corsica is north of Sardinia while providing the exact same amount of support for Sigurd to believe that Corsica is north of Sardinia. So, despite the fact that Sigurd has two distinct sources of support for her belief that Corsica is north of Sardinia, she has no more justification for believing that Corsica is north of Sardinia than she would were she to have only one of those sources.

This counter-example exploits the fact that the principle of cumulative justification ignores the possibility that a single belief might be both a source of support for one to believe that  $q$  and a defeater with respect to one's belief that  $q$ .<sup>20</sup> It's worth noting, and much more will be made of this in the next section, that S's belief that Egil asserted that [Urik is not a reliable source with respect to whether or not (Corsica is north of Sardinia) but, as a matter of fact, (Corsica is north of Sardinia)]<sup>21</sup> is also a defeater with respect to S's belief that Corsica is north of Sardinia, despite the fact that it does not decrease Sigurd's justification for believing that Corsica is north of Sardinia. It renders inadequate the support that is provided by Sigurd's belief that Urik asserted that Corsica is north of Sardinia, and thus functions in the same way that Sigurd's belief that Urik is unreliable with respect to the relative locations of Corsica and Sardinia would function.

Cases like this one are undoubtedly complicated and, perhaps, abnormal. They

---

19 Whichever you think matters for justification.

20 Thanks to Landon Hedrick for pushing this issue.

21 We will use square brackets to clarify scope where necessary.

raise questions about how finely our belief states are carved and to what belief in a conjunction really amounts. Still, there is nothing in them that seems impossible and we maintain that they stand as counter-examples to both the principle of cumulative justification and the causal theory of defeat.

The problems for the causal theory do not end there, however. Jonathan Kvanvig (2007) points to another type of case that causes problems for the causal account of defeat. Imagine the following sequence of events:

(T<sub>1</sub>) S believes that S's atlas shows Corsica being north of Sardinia. S also believes that S's friend asserted that Corsica is north of Sardinia. S believes that Corsica is north of Sardinia solely on the basis of her belief that her atlas shows Corsica being north of Sardinia.

(T<sub>2b</sub>) S comes to believe that S's teacher asserted that Corsica is not north of Sardinia.

(T<sub>5</sub>) S comes to believe that S's geography teacher is unreliable.

In (T<sub>5</sub>) S is doxastically justified in believing that Corsica is north of Sardinia and S would be doxastically justified in believing that Corsica is north of Sardinia if that belief were based on S's friend's testimony instead. Regardless of S's basis for believing that Corsica is north of Sardinia, S is propositionally justified in believing this at (T<sub>5</sub>). In this case, the causal theory of defeat does not classify the testimony of S's teacher as a defeater since S's belief that her teacher is unreliable prevents it from causing S to become unjustified in believing that Corsica is north of Sardinia.

In this case, S's belief that her teacher is unreliable is a defeater-defeater. Despite

the fact that S does not lose any justification for believing that Corsica is north of Sardinia, S's teacher's testimony is a defeater with respect to S's belief that Corsica is north of Sardinia. Consider an analogy to a theory of justification. If a theory of justification took there to be no significant similarity between a reason for S to believe that  $q$  when there is no relevant defeater and a reason for S to believe that  $q$  when there is a relevant defeater, then that theory would be missing an important, underlying feature of S's reason to believe that  $q$  that is shared in both the case with and the case without a defeater. It would be missing something fundamental about reasons for belief. Similarly, a theory of defeat that does not recognize an important, underlying similarity between an undefeated defeater and a defeated defeater is missing something fundamental about defeaters.

So, (T<sub>3</sub>)<sup>22</sup>, (T<sub>4</sub>), and (T<sub>5</sub>) are all counter-examples to the causal theory of defeat. While causing the loss of justification for an otherwise justified belief might be an important property commonly associated with defeaters, if we are interested in developing an accurate account of how defeaters function, and we are, we should reject any account that takes it to be the central characteristic of defeaters. We will return to discuss the explanatory inadequacy of the causal theory in section 5 below.

#### 4. *The Reason-Neutralizing Theory of Defeat*

We demonstrated in the previous section that the causal theory of defeat cannot accommodate cases of epistemic overdetermination, epistemically deviant causal chains, or defeated defeaters. The problem seems to be that the causal theory of defeat is

---

<sup>22</sup> Or (T<sub>3</sub>)'s more complicated conjunctive cousin.

insensitive to the relationship between different defeaters with respect to S's belief that  $q$  and different sources of support for S to believe that  $q$ . The popular<sup>23</sup> reason-neutralizing theory of defeat resolves this problem by isolating the relationship between a source of support and a defeater and asserting that the latter renders the former somehow inert.

#### 4.1 *The Cases It Gets Right*

According to the reason-neutralizing theory, S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if S's belief that  $p$  is a reason for S to believe that  $q$  and S's belief that  $(p \& r)$  would not be a reason for S to believe that  $q$ . This account clearly gets (T<sub>2a</sub>) and (T<sub>2b</sub>) right:

- (T<sub>1</sub>) S believes that S's atlas shows Corsica being north of Sardinia. S also believes that S's friend asserted that Corsica is north of Sardinia. S believes that Corsica is north of Sardinia solely on the basis of her belief that her atlas shows Corsica being north of Sardinia.
- (T<sub>2a</sub>) S comes to believe that S's atlas is unreliable.
- (T<sub>2b</sub>) S comes to believe that S's teacher asserted that Corsica is not north of Sardinia.

Since S's belief that her atlas shows that Corsica is north of Sardinia is a reason to believe that Corsica is north of Sardinia but S's belief that [her atlas shows that Corsica is north of Sardinia but S's atlas is unreliable] would not be a reason for S to believe that Corsica

---

23 This view has been suggested in slightly different forms by many writers: see John Pollock (1974, 42 and 1986, 38), Roderick Chisholm (1977, 72-3), Paul Moser (1985, 66), Richard Foley (1987, 18), William Alston (1989, 238), Matthias Steup (1998, 13), John Pollock and Joseph Cruz (1999, 195), Richard Feldman (2005, 104), Jonathan Kvanvig (2007, 119), and David Christensen (2010). People love this theory.

is north of Sardinia, the reason-neutralizing theory correctly classifies S's belief that the atlas is unreliable as a defeater.

Similarly, since S's belief that her atlas shows that Corsica is north of Sardinia and S's belief that her friend asserted that Corsica is north of Sardinia are reasons for S to believe that Corsica is north of Sardinia but S's beliefs that [her atlas shows that Corsica is north of Sardinia but her teacher asserted that Corsica is not north of Sardinia] and that [her friend asserted that Corsica is north of Sardinia but her teacher asserted that Corsica is not north of Sardinia] would not be reasons for S to believe that Corsica is north of Sardinia, the reason-neutralizing theory correctly classifies S's belief that her teacher asserted that Corsica is not north of Sardinia as a defeater.

It's interesting to note that Bergmann (2006a, 159-60) and Feldman (2005, 104)<sup>24</sup> suggest that the difference between the causal theory of defeat and the reason-neutralizing theory of defeat is merely terminological. Proponents of the reason-neutralizing theory usually write as though S's belief that *r* defeats S's belief that *p* *qua* reason for S to believe that *q*, whereas proponents of the causal theory tend to write as though S's belief that *r* defeats S's belief that *q*.<sup>25</sup> With respect to these differences, Bergmann writes:

There isn't much at stake in this difference of opinion about how to think of defeatees. Consider Pollock's example where I believe that the widgets moving

---

24 Bergmann does this explicitly. Feldman, on the other hand, merely does this implicitly by presenting a causal theory and a reason-neutralizing theory one after the other as though they are two different, but equivalent, ways of characterizing defeat. See the discussion of Feldman in the previous chapter for more on this.

25 We have tried to be neutral about just what gets defeated in a case of defeasible justification. This is why we write of S's belief that *r* being a defeater *with respect to* S's belief that *q*, rather than *for* S's belief that *q*.

along the conveyor belt are red and I believe this on the basis of my visual experience. Such a belief is, it seems, a justified belief. But suppose a superintendent comes along and tells me the widgets are being irradiated by a red light so that they will look red even if they aren't red. My new realization that the widgets will look red even if they aren't red gives me a defeater. But what is it a defeater for? If you think defeatees are beliefs, you will say it is a defeater of my vision-based belief that the widgets are red: you will think that belief ceases to be justified. If you think defeatees are reasons, you will say it is a defeater for my reason for that belief, namely, my visual experience: you will think that visual experience loses its power to confer justification on my belief that the widgets are red. But notice that those who think defeatees are beliefs will agree that the visual experience loses its power to confer justification on my belief that the widgets are red; they'll say that that is why the belief is defeated. And those who think defeatees are reason will agree that the belief that the widgets are red ceases to be justified; they'll say that that is the result of the visual experience being defeated. So there isn't really any substantive disagreement here as far as I can tell. It's just a difference in terminology. (2006a, 160)

While Bergmann is correct that both the causal theory of defeat and the reason-neutralizing theory of defeat deliver the same result in the case he considers, he's wrong that the difference between the two theories is merely terminological. This can be seen by considering how the reason-neutralizing theory handles the cases that were problematic for the causal theory:



(T<sub>3</sub>) S comes to believe that her friend is unreliable with respect to geography. The causal theory incorrectly classifies S's belief that her friend is unreliable with respect to geography as a non-defeater with respect to S's belief that Corsica is north of Sardinia because S's propositional justification for believing that Corsica is north of Sardinia is overdetermined. This is not a problem for the reason-neutralizing theory since it targets specific sources of support. Since S's belief that her friend asserted that Corsica is north of Sardinia is a reason for S to believe that Corsica is north of Sardinia and S's belief that [her friend asserted that Corsica is north of Sardinia and her friend is unreliable with respect to geography] would not be a reason for S to believe that Corsica is north of Sardinia, the reason-neutralizing theory correctly classifies S's belief that her friend is unreliable with respect to geography as a defeater with respect to S's belief that Corsica is north of Sardinia.

Similarly, the reasons-reason-neutralizing theory can handle cases of deviant causal chains where the causal-theory cannot. Recall our case:

(T<sub>4</sub>) S comes to believe that things will go badly for S if S believes that her atlas shows that Corsica is north of Sardinia or that her friend asserted that Corsica is north of Sardinia. Because of S's belief about things going badly for her, S ceases to believe that her atlas shows that Corsica is north of Sardinia and that her friend asserted that Corsica is north of Sardinia.

The causal theory incorrectly classifies S's belief about how things will go for her as a defeater with respect to S's belief that Corsica is north of Sardinia since it causes S to cease to be justified in believing that Corsica is north of Sardinia, albeit circuitously. This

is not a problem for the reason-neutralizing theory since it takes no explicit note of causal relations. Since S's beliefs that her atlas shows that Corsica is north of Sardinia and that her friend asserted that Corsica is north of Sardinia are reasons for S to believe that Corsica is north of Sardinia and S's beliefs that [her atlas shows that (Corsica is north of Sardinia) but things will go badly for S if she believes (her atlas shows that Corsica is north of Sardinia)] and that [her friend asserted that (Corsica is north of Sardinia) but things will go badly for S if she believes the her (friend asserted that Corsica is north of Sardinia)] would still be reasons for S to believe that Corsica is north of Sardinia, the reason-neutralizing theory correctly classifies S's belief about how things will go for her as a non-defeater with respect to S's belief that Corsica is north of Sardinia. Since the reason-neutralizing theory is stated in terms of what would or would not be a reason, rather than what is or is not a reason, it places no significance on loss of justification due to actual loss of beliefs.

Lastly, consider the following sequence:

- (T<sub>1</sub>) S believes that S's atlas shows Corsica being north of Sardinia. S also believes that S's friend asserted that Corsica is north of Sardinia. S believes that Corsica is north of Sardinia solely on the basis of her belief that her atlas shows Corsica being north of Sardinia.
- (T<sub>2b</sub>) S comes to believe that S's geography teacher asserted that Corsica is not north of Sardinia.
- (T<sub>5</sub>) S comes to believe that S's teacher is unreliable.

The causal theory incorrectly classifies S's belief in (T<sub>5</sub>) that her teacher asserted that

Corsica is not north of Sardinia as a non-defeater with respect to S's belief that Corsica is north of Sardinia since it does not cause S to become unjustified in believing that Corsica is north of Sardinia. This is because S's belief that her teacher is unreliable is a defeater-defeater, which prevents her teacher's testimony from causing a loss of justification. This, too, is no problem for the reason-neutralizing theory.

Since S's believes that her atlas shows that Corsica is north of Sardinia and that her friend asserted that Corsica is north of Sardinia are reasons for S to believe that Corsica is north of Sardinia but her beliefs that [her atlas shows that (Corsica is north of Sardinia) but her teacher asserted that (Corsica is not north of Sardinia)] and that [her friend asserted that (Corsica is north of Sardinia) but her teacher asserted that (Corsica is not north of Sardinia)] would not be reasons for S to believe that  $q$ , the reason-neutralizing theory correctly classifies S's belief that her teacher asserted that Corsica is not north of Sardinia as a defeater with respect to S's belief that Corsica is north of Sardinia.

Interestingly, since S's beliefs that [her atlas shows that (Corsica is north of Sardinia) and her teacher asserted that (Corsica is not north of Sardinia) but her teacher is unreliable] and that [her friend asserted that (Corsica is north of Sardinia) and her teacher asserted that (Corsica is not north of Sardinia) but her teacher is unreliable] would be reasons for S to believe that Corsica is north of Sardinia, the reason-neutralizing theory would correctly classify S's belief that [her teacher said that (Corsica is not north of Sardinia) but her teacher is unreliable] as a non-defeater with respect to S's belief that Corsica is north of Sardinia. This result will be relevant when considering the counter-

example to the reason-neutralizing theory below.

Since the causal theory and the reason-neutralizing theory deliver different results for the same cases, they are, *contra* Bergmann and Feldman, not equivalent. Since the reason-neutralizing theory delivers the correct result in all the cases in which the causal theory delivers the correct result as well as several cases where the causal theory delivers the incorrect result, the reason-neutralizing theory of defeat is, in fact, superior to the causal theory of defeat.

#### 4.2 *The Case It Gets Wrong*

Despite its advantages over the causal theory, the reason-neutralizing theory is still inadequate. The biconditional fails read left to right; there might be an  $r$  such that S's belief that  $r$  is a defeater with respect to S's belief that  $q$  even though there is no  $p$  such that S's belief that  $p$  is a reason for S to believe that  $q$  but S's belief that  $(p \& r)$  would not be a reason for S to believe that  $q$ . Recall the case that was used above to refute the claim that justification is cumulative. In that case, S has a belief that plays a dual-role: it is both a reason for S to believe that  $q$  and also a defeater with respect to S's belief that  $q$ . The reason-neutralizing theory, not to mention the causal theory, incorrectly classifies such a belief as a non-defeater with respect to S's belief that  $q$ .<sup>26</sup> Consider the following case:

Glancing at the clock, S forms the belief that it says that it is 2 o'clock and, on this basis, forms the belief that it is 2 o'clock. Noticing her looking at the clock, her friend says, "Oh, that clock is broken, although, as it happens it is 2 o'clock right now." S believes that her friend asserted that the clock is broken but it is 2 o'clock.

---

<sup>26</sup> Jake Chandler (2013) develops a similar counter-example.

S's belief that the clock indicates that it is 2 o'clock is a reason for S to believe that it is 2 o'clock and S's belief that [the clock indicates that (it is 2 o'clock) but her friend asserted that (the clock is broken)] would not be a reason for S to believe that it is 2 o'clock, and so the reason-neutralizing theory would correctly classify S's belief that her friend asserted that the clock is broken as a defeater with respect to S's belief that it is 2 o'clock.

The same cannot be said for S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)]. Notice that in this case S comes to form only one belief, that [her friend asserted that (the clock is broken but it is 2 o'clock)], rather than the belief that her friend asserted that the clock is broken and the belief that her friend asserted that it is 2 o'clock. S's belief that the clock indicates that it is 2 o'clock is a reason for S to believe that it is 2 o'clock but so, too, would S's belief that [the clock indicates that (it is 2 o'clock) and (her friend asserted that the clock is broken but it is 2 o'clock)] be, albeit not in virtue of the clock-content of the belief but rather in virtue of time-content. So, the reason-neutralizing theory of defeat would classify S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] as a non-defeater with respect to S's belief that it is 2 o'clock.

This is the incorrect result. Notice that the advantage that the reason-neutralizing account has over the causal theory is that the reason-neutralizing account is sensitive to the sources of support. The reason-neutralizing theory is better than the causal theory because it, whether by design or not, accommodates the important fact that it is not a change in the amount of justification that S has to believe that *q* that matters to defeat but rather the adequacy of the support that a given source provides for S to believe that *q* that

is important. As Bergmann noted in the passage above:

[...] those who think defeaters are beliefs will agree that the visual experience loses its power to confer justification on my belief that the widgets are red; they'll say that that is why the belief is defeated. (2006a, 160)

So, even a proponent of the causal theory of defeat admits that it is the (in)adequacy of support provided by a source (or “the power to confer justification” of that source) that matters for defeat. S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] prevents the clock from providing adequate support for S to believe that it is 2 o'clock, so given the fact that it is a loss of adequate support from a source and not a loss of justification *simpliciter* to believe that matters, S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] is a defeater with respect to S's belief that it is 2 o'clock, regardless of the fact that S's belief that [the clock indicates that (it is 2 o'clock) and her friend asserted that the (clock is broken but it is 2 o'clock)] would still be a reason for S to believe that it is 2 o'clock.

Recall the correct result that the reason-neutralizing theory delivers in the case of defeated-defeaters: S's belief that her teacher asserted that Corsica is not north of Sardinia is a defeater with respect to S's belief that Corsica is north of Sardinia but S's belief that [S's teacher asserted that (Corsica is north of Sardinia) but S's teacher is unreliable] would not be. One might wonder why we should think that S's belief that [S's teacher asserted that (Corsica is north of Sardinia) but S's teacher is unreliable] would not be a defeater but S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] is; after all, both S's belief that [S's atlas shows that (Corsica is north of

Sardinia) and S's teacher asserted that (Corsica is north of Sardinia) but S's teacher is unreliable] and S's belief that [the clock indicates that (it is 2 o'clock) and her friend asserted that (the clock is broken but it is 2 o'clock)] would be reasons for S to believe the relevant propositions, that Corsica is north of Sardinia or that it is 2 o'clock, respectively.

This worry highlights an important point. Whether or not S is left with a reason to believe that  $q$  is no more relevant to defeat than whether or not S is left justified in believing that  $q$ , so the fact that both S's belief that [S's atlas shows that (Corsica is north of Sardinia) and S's teacher asserted that (Corsica is north of Sardinia) but S's teacher is unreliable] and S's belief that [the clock indicates that (it is 2 o'clock) and her friend asserted that (the clock is broken but it is 2 o'clock)] would be reasons for S to believe the respective propositions is irrelevant to whether or not S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] is a defeater with respect to S's belief that it is 2 o'clock.

What is important is whether or not S has reason to not rely on a source of support for S to believe that  $q$ . The thing to notice here is that the clock case is similar to (T<sub>2a</sub>) in that both S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] and S's belief that her atlas is unreliable bear the same relationship to the relevant beliefs, i.e., that the clock indicates that it is 2 o'clock and that her atlas shows that Corsica is north of Sardinia. With respect to S's belief that the clock indicates that it is 2 o'clock, the 2 o'clock-content of S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] is an idle cog. S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] does the same thing to S's belief that the clock indicates that it is 2 o'clock

as S's belief that her atlas unreliable does to her belief that her atlas indicates that Corsica is north of Sardinia, i.e., prevents it from providing adequate support for S to form the relevant belief.

As the causal theory did with respect to S's belief in (T<sub>2a</sub>) and (T<sub>3</sub>), the reason-neutralizing theory misses an important similarity between S's belief in (T<sub>2a</sub>) that S's atlas is unreliable and S's belief in the clock case that [her friend asserted that (the clock is broken but it is 2 o'clock)]. If one wants to understand how S's belief that her atlas is unreliable functions with respect to S's believing that Corsica is north of Sardinia in (T<sub>2a</sub>), one's explanation of that phenomenon must also explain the function of S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] with respect to S's believing that it is 2 o'clock in the clock case. Thus, it cannot be that S's belief that her atlas is unreliable in (T<sub>2a</sub>) functions as a defeater in virtue of the fact that S's belief that [S's atlas shows that (Corsica is north of Sardinia) but S's atlas is unreliable] would not be a reason for S to believe that Corsica is north of Sardinia because S's belief that [the clock indicates that (it is 2 o'clock) and her friend asserted that (the clock is broken but it is 2 o'clock)] would be a reason for S to believe that it is 2 o'clock. This would be to treat unified phenomena as though they were disparate.

So, the reason-neutralizing theory of defeat incorrectly classifies S's belief that [her friend asserted that (the clock is broken but it is 2 o'clock)] as a non-defeater with respect to S's belief that it is 2 o'clock, and so is extensionally inadequate. It fails despite the fact that it is more source specific than the causal theory; it fails because it is still not source specific enough. More will be made of this in the next chapter.



## 5. *Explanatory Inadequacy*

In the previous sections we argued that the causal theory and the reason-neutralizing theory of defeat are extensionally inadequate. But let's assume that they are not; let's assume that they correctly distinguish defeaters from non-defeaters in all cases. If we're interested in understanding how defeaters function, that won't be enough for us. An adequate theory of defeat needs to help us understand not just that, but why and how, S's belief that  $r$  is a defeater. Neither the reason-neutralizing theory nor the causal theory can do this.

Consider the causal theory first. It asserts that S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if S's believing that  $r$  causes S to become unjustified in believing that  $q$ . Clearly this will not help us to understand why S's belief that  $r$  is a defeater since it has nothing to tell us about how it causes S to become unjustified in believing that  $q$ . The causal theory of defeat explains defeat in the same way that saying, "When water falls from the sky, it's raining," explains rain. Even if it were accurate, it would not be particularly informative. We want to know why the water falls from the sky and we want to know why defeaters cause a loss of justification.

The reason-neutralizing theory of defeat appears to be more informative than the causal theory. It asserts that S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if S's belief that  $p$  is a reason for S to believe that  $q$  and S's belief that  $(p \& r)$  would not be a reason for S to believe that  $q$ . This indicates that defeat has something to do with a source of support for S to believe that  $q$  and that source ceasing to provide adequate support when conjoined with a defeater.

However, the explanation that the reason-neutralizing theory offers is shallow. It isn't at all obvious why the fact that S's belief that  $p$  is a reason for S to believe that  $q$  but S's belief that  $(p \& r)$  would not be a reason for S to believe that  $q$  makes S's belief that  $r$  a defeater with respect to S's belief that  $q$ . S, presumably, has plenty of beliefs that are not reasons for S to believe that  $q$ , let alone the infinite number of beliefs that, were S to have them, would not be reasons for S to believe that  $q$ . The fact that S has these beliefs or could have these beliefs has no obvious bearing on S's justification for believing that  $q$ . It isn't clear what is special about the fact that S's belief that  $(p \& r)$  would not be a reason for S to believe that  $q$  in virtue of which S's belief that  $r$  can affect S's justification for believing that  $q$ . More needs to be said.

If S's belief that  $r$  were to cause S to cease to believe that  $q$  on the basis of  $p$  and begin believing that  $q$  on the basis of  $(p \& r)$  instead, that would make it clear why S's justification for believing that  $q$  might be affected. Since S's belief that  $(p \& r)$  is not a reason for S to believe that  $q$ , it doesn't provide any support for S's belief that  $q$ . But this explanation is implausible for several reasons.

First, as we saw in section 3, loss of justification for believing that  $q$  is not a necessary condition on defeat. Second, S might believe that  $r$  but fail to recognize that  $r$  is related in any way to S's beliefs that  $p$  or S's belief that  $q$ . In such a case there is no reason to think that S's basis for believing that  $q$  would change simply because S came to believe that  $r$ .<sup>27</sup> In fact, though S both believes that  $p$  and believes that  $r$ , S might not believe that  $(p \& r)$  and so could not base any belief on the belief that  $(p \& r)$ , anyway. It

---

<sup>27</sup> We do not mean to take a stand here on whether or not a belief must be recognized as a defeater in order to function as a defeater. We will attend to this issue in chapter 4. We only mean to point out that this picture of defeat settles that question by fiat.

might be suggested that each of S's beliefs is based on S's total evidence, but this is implausible if the basing relation is taken to be a causal relation since the relevant sustaining cause for any given belief will not be one's total evidence but rather some proper subset of one's total evidence. Furthermore, this would make doxastic justification identical to propositional justification since it would make one's total evidence for believing that  $q$  and the evidence on which one bases one's belief that  $q$  one and the same.

Third, S might recognize the relevance of  $r$  to S's belief that  $q$  but stubbornly refuse to let that change her basis for believing that  $q$ . It is not hard to imagine a student, say, believing that God created the world on the basis of the belief that the Bible asserts that God created the world even when that student comes to believe that there is good reason to believe that the Bible is not a reliable source of information. When asked why she believes that God created the world given the evidence against the Bible she might blithely, but sincerely, respond, "Because the Bible says that God created the world." So, recognizing the relevance of a defeater to one's beliefs does not obviously force one to change one's basis for belief.

Fourth, even if S does change her basis for believing that  $q$  from her belief that  $p$  to her belief that  $(p \& r)$  upon coming to believe that  $r$ , this would not explain the distinctive force of defeaters. Assume that S's belief that  $p$  is a reason for S to believe that  $q$  but S's belief that  $t$  is not a reason for S to believe that  $q$ . Assume further that S's belief that  $(p \& t)$  would be a reason for S to believe that  $q$ , i.e., S's belief that  $t$  is not a defeater with respect to S's belief that  $q$  according to the reason-neutralizing theory of defeat. Were S to change her basis for believing that  $q$  from S's belief that  $p$  to S's belief that  $t$

she would cease to be justified in believing that  $q$ . Assuming that it is S's coming to believe that  $q$  on the basis of a belief that is not a reason for S to believe that  $q$ , e.g., S's belief that  $(p \& r)$ , instead of a reason for S to believe that  $q$ , e.g., S's belief that  $p$ , the loss of justification that S suffers in coming to believe that  $q$  on the basis of the belief that  $t$  has the same kind of explanation as the loss of justification in a case of defeat.

This case, similar in some respects to the problematic case of  $(T_4)$  above, is counter-intuitive in and of itself, but has even less plausible consequences. If S came to base her belief that  $q$  on S's belief that  $p$  and also on S's belief that  $t$ , i.e., if S were to come to have multiple bases for her belief that  $q$ , she would still be justified in believing that  $q$ . However, intuitively, if S were to base her belief that  $q$  on her belief that  $p$  and also on her belief that  $(p \& r)$  she would cease to be justified, even though her belief that  $p$  is, by hypothesis, a reason for her to believe that  $q$ . But if the cause of loss of justification in the case of defeat were the fact that S bases her belief that  $q$  on a belief that is not a reason for her to believe that  $q$ , we should have the same result in both these cases. So, it cannot be that defeaters function by changing the basis of one's belief from a basis that is a reason for that belief to another basis that is not a reason for that belief. This explanation, too, treats as unified phenomena that are actually disparate.

Perhaps, instead, S's belief that  $r$  is a defeater with respect to S's belief that  $q$  because it causes S's belief that  $p$  to cease to be a reason for S to believe that  $q$ . This picture solves some of the problems facing the previous picture. If S's belief that  $r$  causes S's belief that  $p$  to cease to be a reason for S to believe that  $q$ , then it does not matter whether or not S recognizes the relevance of her belief that  $r$  to her beliefs that  $p$  or that

$q$ . It also does not matter that  $S$  might not believe that  $(p \& r)$ . On this picture, defeat does not require one to change their basis for belief at all, so the stubborn student case above would pose no problem.

Furthermore, this interpretation is supported by remarks made by Pollock in reference to what he calls type II or undercutting defeaters. He said of such a defeater that it “attacks the connection” between one's evidence and one's belief (1974, 42). Assuming that this attack metaphor is meant to convey some destructive power, it seems that Pollock conceived of at least some instances of defeat as severing evidential connections.

However, this picture faces problems of its own. First, it makes the formulation of the reason-neutralizing theory of defeat inexplicable. If this is the correct picture of how defeaters work, the account should be put as follows:  $S$ 's belief that  $r$  is a defeater with respect to  $S$ 's belief that  $q$  if and only if  $S$ 's belief that  $p$  is a reason for  $S$  to believe that  $q$  given that  $S$  has no other beliefs and  $S$ 's belief that  $p$  is not a reason for  $S$  to believe that  $q$  given that  $S$  believes that  $p$  and that  $S$  believes that  $r$  and has no other beliefs. Since this is not the account that we are given, we should be suspicious of this interpretation.

Second, this view takes a controversial stance on the nature of reasons to believe. On Pollock's own account of reasons, for example,  $S$ 's belief that  $p$  is a reason for  $S$  to believe that  $q$  if and only if it is logically possible for  $S$  to be justified in believing that  $q$  on the basis of her belief that  $p$  (Pollock and Cruz 1986, 36).  $S$ 's belief that  $r$  cannot change whether or not it is possible for  $S$  to be justified in believing that  $q$  on the basis of her belief that  $p$ , so this picture of defeat is inconsistent with Pollock's own view of

reasons.

Third, this picture still would not explain why a defeater is a defeater since it does not tell us why or how S's belief that *r* causes S's belief that *p* to cease to be a reason for S to believe that *q*. Indeed, it is puzzling how something could be a reason to believe in one case and cease completely to be a reason to believe in another.<sup>28</sup>

What's more, since it is only some defeaters that are supposed to attack evidential connections, this account would not explain how defeat functions, but rather only how some instances of defeat function. Positing multiple mechanisms of defeat is, *ceteris paribus*, a cost for a theory. A theory that could explain defeat with a single mechanism would be superior.

## 6. *The Quietist Theory of Defeat*

In the face of these difficulties, it has been suggested that the mechanism of defeat is inexplicable, that it is a brute fact.<sup>29</sup> We can call this “the quietist theory of defeat.” On this theory, the best we can do is consider various cases and deliver intuitive judgments about whether or not a given belief is a defeater with respect to another given beliefs. We might be able to produce useful and accurate generalizations from these cases, but these generalizations would not be informative with respect to the underlying mechanism of defeat, perhaps because there is none. It is worth distinguishing between two kinds of

---

28 Some interesting work on *prima facie* moral reasons and silencing by Jonathan Dancy (1993, 47-53) notwithstanding. In any case, my point here is not that something can lose its status as a reason; this might be the fundamental mechanism of defeat. What has not been shown is why some thing's satisfying the reason-neutralizing theory of defeat gives us any reason to believe that it changes the reason status of the one-time reason to believe that *q*.

29 Stewart Cohen (1983, 290-1) argues against this view, but no one explicitly argues for this view in the literature anywhere, but it has been suggested in conversation between Cohen and Pollock and between by several other philosophers, including Michael Bergmann.

quietism, if only to set one aside. Epistemic quietism with respect to the nature of defeat is only committed to our inability to acquire an illuminating picture of the mechanism of defeat. Metaphysical quietism is committed to the further claim that there is no underlying mechanism of defeat. We are only concerned with epistemic quietism.<sup>30</sup>

This situation is undesirable in and of itself. It's doubly undesirable given that there is disagreement between philosophers concerning just what counts as a defeater. The quietist theory could not help us settle these disputes. But beyond these, admittedly practical, reasons for rejecting the quietist theory of defeat, there are good theoretical reasons as well. For example, the defeater in (T<sub>2a</sub>) functions differently from the defeater in (T<sub>2b</sub>). The former is source specific, but the latter does not seem to be. Given the importance of taking the source of justification into account in developing a theory of defeat, as noted in section 4, this is a puzzling fact. Surely there is a reason why one is source specific and the other is not. Or consider the fact that defeaters come in different strengths. Surely there is an explanation for this variety; it points to some underlying mechanism. Or consider the possibility that some sources of justification, say *a priori* reasoning, might be immune to some sources of defeat, say empirical observation. Surely this is not just a brute fact.

What is required for explanatory adequacy is contextually determined. In contexts where an explanation is required for each and every feature of our epistemic lives, no theory will be adequate. But the fact that no theory can be explanatorily adequate in all contexts is not a reason to hold theories to no standard of explanatory adequacy whatsoever. We may be forced to accept quietism, but not without a fight.

---

30 Thanks to Brian Fiala and Christopher Gibilisco for pushing this distinction.

## 7. *Conclusion*

The causal theory of defeat and the reason-neutralizing theory of defeat are both explanatorily inadequate and extensionally inadequate. The quietist theory of defeat is trivially extensionally adequate but obviously, and by design, explanatorily inadequate. If there were no better theories of defeat available, then we would, perhaps, be forced to accept the quietist theory.

But there is a better theory available, one suggested by the short-comings of the three theories considered in this chapter. This theory, which we can call “the reasons-against-belief theory of defeat,” asserts, roughly, that S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if S's belief that  $p$  would be a reason for S to believe that  $q$  and S's belief that  $r$  is a reason for S to not believe that  $q$  on the basis of S's belief that  $p$ . We will defend this theory and present it in greater detail in the next two chapters.



# CHAPTER 3

## THE REASONS AGAINST BELIEF THEORY OF DEFEAT

In the previous chapter we argued that most of the existing theories of defeat are extensionally inadequate and all are explanatorily inadequate. An adequate theory of defeat must get the cases right and must also explain how it is that defeat occurs. In this chapter we will present, explain, and defend a theory that does just that.

Recall the two main theories that were considered in the last chapter, the causal theory and the reason-neutralizing theory:

**The Causal Theory:** S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if S's believing that  $r$  causes S to become unjustified in believing that  $q$ .

**The Reason-Neutralizing Theory:** S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if S's belief that  $p$  is a reason for S to believe that  $q$  and S's belief that  $(p \& r)$  would not be a reason for S to believe that  $q$ .

Both of these theories can handle a standard case of defeat, such as the following:

S justifiably believes that  $q$  on the basis of S's belief that  $p$ . Subsequently, S comes to believe that,  $r$ ,  $p$  is not a reliable indicator with respect to  $q$ .

Since S's belief that  $r$  would cause S to cease to be justified in believing that  $q$ , the causal theory correctly classifies it as a defeater and since S's belief that  $(p \& r)$  would not be a

reason for S to believe that  $q$  but S's belief that  $p$  would be a reason for S to believe that  $q$ , the reason-neutralizing theory also correctly classifies it as a defeater.

Both of the theories fail, however, when faced with non-standard cases of defeat like the following:

S justifiably believes that  $q$  on the basis of S's belief that  $p$ . Subsequently, S comes to believe that,  $r_1$ , [ $p$  is not a reliable indicator with respect to  $q$  but ( $p_1$  is a reliable indicator with respect to  $q$ ) and  $p_1$ ], and S also comes to base S's belief that  $q$  on S's belief that  $r_1$ .

We can call cases like this “double-agent cases” since S's belief that [ $p$  is not a reliable indicator with respect to  $q$  but ( $p_1$  is a reliable indicator with respect to  $q$ ) and  $p_1$ ] is both a defeater with respect to S's belief that  $q$ , in virtue of the first conjunct, and a reason for S to believe that  $q$ , in virtue of the second and third conjuncts. Since S's belief that  $r_1$  is itself sufficient reason for S to believe that  $q$ , it would not cause S to cease to be justified in believing that  $q$  and S's belief that ( $p \& r_1$ ) would be a reason for S to believe that  $q$ . So, both the causal theory and the reason-neutralizing theory would classify S's belief that  $r_1$  as a non-defeater with respect to S's belief that  $q$ . But this is the wrong result.

In the previous case, S's belief that  $r$  is a defeater with respect to S's belief that  $q$  because it is a reason for S to believe that  $p$  is not a reliable indicator of the truth-value of  $q$ . If we are to understand how defeaters function, we must investigate this relation between being a reason for S to believe that  $p$  is not a reliable indicator and being a defeater with respect to S's belief that  $q$ . But notice that S's belief that  $r_1$  in the present case also indicates that  $p$  is not a reliable indicator of the truth-value of  $q$ . It plays the

same function as S's belief that  $r$  did in the previous case. Indeed, S's belief that  $r_1$  has two functions, one of which is to act as a reason for S to believe that  $q$  but the other of which is to do precisely what S's belief that  $r$  did in the previous case. Any theory of defeat that classifies S's belief that  $r$  as a defeater but does not classify S's belief that  $r_1$  as a defeater is extensionally inadequate and, what's more, doomed to explanatory inadequacy.

It is a mistake, then, to think that the features picked out by the causal theory and the reason-neutralizing theory are essential to defeat. They may, indeed, be important and common consequences of defeat, but they neither amount to, nor illuminate, defeat. Let us look, instead, at what is fundamentally going on in cases of defeat. In standard cases of defeat S would not be justified in believing that  $q$ , as is often noted. However, what gets less attention is the fact that S would be justified in not believing that  $q$ .

The scope of the negation here matters. If S is justified in believing something, then S must have most epistemic reason to believe it, whatever epistemic reasons are. Importantly, the analog to this when S is justified in not believing something is not an absence of most epistemic reason to believe that thing but rather the presence of most epistemic reason to not believe that thing. It may well be that an absence of reason to believe something is, or gives one, a reason to not believe that thing, but it is that reason *qua* reason to not believe that makes S justified in not believing it. S being justified in not believing that  $q$  does entail S not being justified in believing that  $q$ ; this is why a reason for S to not believe that  $q$  might decrease S's justification for believing that  $q$ . That being said, there are important differences between a theory that asserts that defeat occurs

through the presence of reasons against belief and a theory that asserts that defeat occurs through the absence of reasons for belief. So, if S's belief that  $r$  is a defeater with respect to S's belief that  $q$ , then S has epistemic reason to not believe that  $q$ . We can make most sense of the phenomenon of defeat if we identify that reason with the defeater, S's belief that  $r$ . This suggests the following theory of defeat:

**The Reasons-Against-Belief Theory:** S's belief that  $r$  would be a defeater with respect to S's belief that  $q$  if and only if S's belief that  $r$  would be a reason for S to not believe that  $q$ .

### 1. *Perfunctory Chisholming*

Before proceeding, a modification must be made to the reasons-against-belief theory. The problem with the causal theory of defeat and the reason-neutralizing theory of defeat is that they both fail to focus on the status of the relationship between S's belief that  $p$  and S's belief that  $q$ . This is obvious in the case of the causal theory since it makes no mention of S's belief that  $p$  at all. It's less obvious in the case of the reason-neutralizing theory, but still present. Although S's belief that  $p$  plays a significant role in the reason-neutralizing theory of defeat, the status of S's belief that  $r$  as a defeater or non-defeater is, according to that theory, ultimately determined by the status of the relationship between S's belief that  $(p \& r)$  and S's belief that  $q$ , not by the status of relationship between S's belief that  $p$  and S's belief that  $q$ .

As it stands, the reasons-against-belief theory also fails to focus adequately on the relationship between S's belief that  $p$  and S's belief that  $q$ . Jake Chandler puts the point in the following way. Consider a case in which E is evidence for one to believe H and D is a

double-agent defeater with respect to H:

One suggestion that would not be of any help here would be to recast [the reason-neutralizing theory of defeat] with a different negational scope, requiring, not that E&D not be a reason to believe H, but that E&D be a reason to not believe H.

Indeed, in the switch case [a double-agent case], E&D is clearly not a reason to not believe H. The upshot of the modified principle would be that, contrary to intuitions, D is not a defeater for E's support for H. (2013, 50)

Chandler's point is that the simple reasons-against-belief theory of defeat doesn't handle double-agent cases any better than the reason-neutralizing theory does. The simple version of the theory fails because it ignores John Pollock's insight that some defeaters are source-specific. Chandler suggests a modification that takes this insight into theory:

Where D and E are jointly consistent propositions, D is a defeater for E's support for H if and only if D is a reason to not believe that E is a reason to believe H.

(2013, 50)

Since S's belief that  $r_1$  in the double-agent case is indeed a reason for S to believe that [her belief that  $p$  is not a reason for her to believe that  $q$ ], Chandler's reasons-against-belief theory gets this problem case right.

However, Chandler's theory faces three problems, one spurious and two serious.

First, the spurious objection. Gregory Wheeler argues that Chandler's theory is too broad:

The problem with [Chandler's theory] is that it fabricates phantom support for a defeater to defeat: D may be a reason to not believe that E is a reason to believe H - which thereby suffices for D to defeat E's support for H - without E being a

reason for H in the first place. For example, inspecting the first 10 light bulbs from a production line and finding all 10 defective (E) does not provide a reason to believe that the next bulb off the line is faultless (H). (2014, 15)

There's just no positive connection between the first 10 light bulbs from the line being defective and the next one being faultless. There may well be a negative connection, but for Wheeler's case he only needs there to be no positive connection, i.e., for E to not be evidence for H.

Now, imagine that you know that all of the light bulbs coming off the line today use the most recent delivery of filaments and that an oxidized filament will result in a defective bulb. Wheeler continues:

[...] learning that the last delivery of filaments to the factory are all oxidized (D) is a reason to believe that the 10 defective light bulbs do not provide a reason to believe that the next bulb is faultless. By [Chandler's theory], D is a defeater for E's support for H even though E does not support H. (2014, 15)

Wheeler's reasoning seems to be this: since D is a reason to believe that not-H, it is also a reason to believe that [the E is not a reason to believe H] and, thus, a reason to not believe that [E is a reason to believe H]. Wheeler's worry is that this would make D a defeater for E's support for H even though E never supported H in the first place. That is supposed to be intuitively implausible and, thus, an indication that Chandler's theory of defeat is inadequate.

Wheeler's argument is flawed. First, it is typically taken to be the case that one can have a reason to believe that not-q without calling into question one's reasons for

believing that  $q$ . Wheeler's move from having reason to believe that not- $H$  to having reason to believe that  $[E$  is not a reason to believe that  $H]$  requires defense but none is provided.

Second, at least part of the problem is an illusion cast by the terminology that Chandler and Wheeler use when discussing defeat. It is a mistake to assume, prior to exploring the nature of defeat, that what is defeated in a case of defeat is the support provided for a belief. This is why we prefer the phrase “a defeater with respect the belief that  $q$ ” to “a defeater for the support for the belief that  $q$ .” Indeed, if we rephrase Chandler's theory and Wheeler's objection in that more neutral language, the problem seems to disappear:

**Chandler's Neutral Reasons-Against-Belief theory:**  $S$ 's belief that  $r$  is a defeater with respect to  $S$ 's belief that  $q$  if and only if  $S$ 's belief that  $r$  is a reason for  $S$  to not believe that  $[S$ 's belief that  $p$  is a reason for  $S$  to believe that  $q]$ .

And:

$S$ 's belief that  $r$  may be a reason for  $S$  to not believe that  $[S$ 's belief that  $p$  is a reason for  $S$  to believe that  $q]$  - which thereby suffices for  $S$ 's belief that  $r$  to be a defeater with respect to  $S$ 's belief that  $q$  - without  $S$ 's belief that  $p$  being a reason for  $S$ 's to believe that  $q$  in the first place. For example, inspecting the first 10 light bulbs from a production line and finding all 10 defective ( $p$ ) does not provide a reason to believe that the next bulb off the line is faultless ( $q$ ). Even so, learning that the last delivery of filaments to the factory are all oxidized ( $r$ ) is a reason to believe that the 10 defective light bulbs do not provide a reason to believe that the

next bulb is faultless. By Chandler's theory, S's belief that  $r$  is a defeater with respect to S's belief that  $q$  even though S's belief that  $p$  does not support S's belief that  $q$ .

On a neutral reading of Wheeler's objection, all of his claims seem to be true, but there is no problem. S's belief that the last delivery of filaments to the factory were oxidized is indeed a defeater with respect to S's belief that the next bulb off the line will be faultless.

Still, it may seem like there is a genuine objection lurking nearby. Imagine that, instead of S believing that the filaments in the last shipment were oxidized, S believes that [her epistemology instructor said that observing that the last ten bulbs were defective is no reason to believe that the next bulb will be faultless]. This is a reason for S to believe that [S's belief that  $p$  is not a reason for S to believe that  $q$ ]. On Chandler's account, this would make S's belief about her epistemology instructor a defeater with respect to S's belief that  $q$  even though S never had reason to believe that  $q$  in the first place. If Chandler were right about this, then defeaters are more plentiful than might have been supposed. For any proposition,  $p$ , if S believes that  $p$  and S's belief that  $p$  is not a reason for her to believe that  $q$ , S can have a defeater with respect to her belief that  $q$  in the form of a reason to believe that [her belief that  $p$  is not a reason for her to believe that  $q$ ].

There really is no problem with this. Recall from the previous chapter that it was argued that the causal theory of defeat is extensionally inadequate, in part, because it fails to classify defeaters that don't cause a loss of justification as defeaters. The underlying mechanism of defeat can be prevented from removing justification: perhaps that defeater



is defeated or perhaps it is source-specific to a source of support that is unused. Defeaters can be inert in this respect. Given that, there is no reason not to let a thousand defeaters bloom.

Given what's been said so far, when S's belief that *r* does cause a loss of justification, this could well be in virtue of the fact that it is a reason for S to not believe that [S's belief that *p* would be a reason for S to believe that *q*]. When it doesn't, this might be in virtue of the fact that S doesn't believe that *p*, or that S doesn't base her belief that *q* on her belief that *p*, or that S does base her belief that *q* on her belief that *p* but S's belief that *p* never justified S in believing that *q* in the first place. It needn't be because S's belief that *r* is not a defeater. However, if you are still uncomfortable with calling these things defeaters, you are urged to look over section 1 of chapter 1; there you'll find what you need to translate from our talk of inefficacious defeaters to terminology you might find more suiting.

However, Chandler's theory of defeat does face two serious problems. While taking Pollock's insight that some defeaters are source-specific into account, Chandler neglects the fact that some defeaters are source-neutral. As Pollock points out:

[...] if P is a prima facie reason for S to believe that Q, then any reason for S to believe that Q is false (even though P is true) is a defeater [...] The second kind of defeater attacks the connection between P and Q rather than attacking Q directly.

(1974, 42)

What differentiates these two kinds of defeaters is the fact that the latter, but not the former, challenges the connection between S's belief that *p* and S's belief that *q*. The

former kind of defeater leaves that connection unchallenged, i.e., it gives S no reason to not believe that [S's belief that  $p$  is a reason for S to believe that  $q$ ]. In short, Chandler's theory is extensionally inadequate because it fails to classify a paradigm example of a defeater as a defeater.

Furthermore, Chandler's theory suffers from the same problem that all of the theories from the last chapter faced: it is explanatorily inadequate. Recall that the reason-neutralizing theory of defeat provides no explanation for how it is supposed to be that the fact that S's belief that  $(p \& r)$  is not a reason for S to believe that  $q$  is relevant to S's belief that  $q$  if S does not base her belief that  $q$  on her belief that  $(p \& r)$  or, indeed, if S does not even believe that  $(p \& r)$ . It may be true that in many cases, where S's belief that  $r$  is a defeater with respect to S's belief that  $q$  and S's belief that  $p$  is a reason for S to believe that  $q$ , that S's belief that  $(p \& r)$  would not be a reason for S to believe that  $q$ . Even so, this fact gives us no clear picture of the mechanism of defeat.

Similarly, it may be true that in many cases, where S's belief that  $r$  is a defeater with respect to S's belief that  $q$  and S's belief that  $p$  is a reason for S to believe that  $q$ , that S's belief that  $r$  is a reason for S to not believe that [S's belief that  $p$  is a reason for S to believe that  $q$ ]. This, too, would not mean that we have a clear picture of the mechanism of defeat. It is not at all clear why having reason to believe that [S's belief that  $p$  is not a reason for S to believe that  $q$ ] would have any effect whatsoever on the evidential value of S's belief that  $p$ . Indeed, this again raises the question that initiated this project: if one needn't believe that one's reasons for belief are good in order to be justified in believing on the basis of them, then why should having reason to believe that they are not good, or

reason to not believe that they are good, reduce one's justification? It is not that we deny that this occurs, but rather that Chandler's theory gets us no closer to an explanation for its occurrence.

So, in shaping the reasons-against-belief theory of defeat, both the Scylla of ignoring source-specific defeaters and the Charybdis of ignoring source-neutral defeaters must be avoided, all while steering toward an explanation of defeat. In pursuit of this, let's return to what's fundamental about standard cases of defeat. First, when S has a defeater with respect to her belief that  $q$ , it is not just that S would not be justified in believing that  $q$ ; S would also be justified in not believing that  $q$ . However, where  $r$  is the claim that S's belief that  $p$  is not a reliable indicator with respect to  $q$ , S's belief that  $r$  is not a reason for S to not believe that  $q$  *simpliciter* but rather a reason for S to not believe that  $q$  on the basis of S's belief that  $p$ . Chandler had the right idea in shifting the scope of the negation, but he misidentifies the content of the reason that S's belief that  $r$  constitutes. It is not that S's belief that  $r$  is a reason for S to not believe that [S's belief that  $p$  is a reason for S to believe that  $q$ ] but rather that S's belief that  $r$  is a reason for S to not believe that  $q$  on the basis of S's belief that  $p$ . So, the reasons-against-belief theory should read:

**The Revised Reasons-Against-Belief Theory:** S's belief that  $r$  would be a defeater with respect to S's belief that  $q$  if and only if S's belief that  $r$  would be a reason for S to not believe that  $q$  on some particular basis.

This adjustment is not unproblematic, and the problems that it generates will be discussed below, but let this be the working formulation of the theory. Until the next revision,

“reasons-against-belief theory” will refer to the revised reasons-against-belief theory.

## 2. *Extensional Adequacy*

The reasons-against-belief theory handles the standard cases of defeat just as well as the causal theory and the reason-neutralizing theory. Recall our cases from the previous chapter:

- (T<sub>1</sub>) S believes that S's atlas shows that Corsica is north of Sardinia. S also believes that S's friend asserted that Corsica is north of Sardinia. S believes that Corsica is north of Sardinia solely on the basis of her belief that her atlas shows Corsica being north of Sardinia.
- (T<sub>2a</sub>) S comes to believe that S's atlas is unreliable.
- (T<sub>2b</sub>) S comes to believe that S's teacher asserted that Corsica is not north of Sardinia.

Since S's belief that her atlas is unreliable is a reason for S to not believe that Corsica is north of Sardinia on the basis of S's belief that her atlas shows Corsica being north of Sardinia, the reasons-against-belief theory correctly classifies S's belief that her atlas is unreliable as a defeater. So far, so good.

But (T<sub>2b</sub>) seems like it might pose a problem for the reasons-against-belief theory; S's belief that her teacher asserted that Corsica is not north of Sardinia doesn't seem to be relative to any particular basis for belief. S's belief that her teacher asserted that Corsica is not north of Sardinia might be a reason for S to not believe that Corsica is north of Sardinia *simpliciter*, but it does not seem to be a reason for S to not believe that Corsica is north of Sardinia on any particular basis. The adjustment offered in the previous

section now seems to be causing trouble for the theory.

However, despite appearances, S's belief that her teacher asserted that Corsica is not north of Sardinia is a reason for S to not believe that Corsica is north of Sardinia on some particular basis. This is so because it is a reason for S to not believe that Corsica is north of Sardinia on any basis whatsoever. If it is a reason for S to not believe that Corsica is north of Sardinia on any basis whatsoever, then it is also trivially a reason for S to not believe that Corsica is north of Sardinia on the basis of S's atlas. This is an interesting feature of this view that we will revisit below.

### 2.1 *The Causal Theory*

Next, consider the cases that the causal theory was not able to handle. First is the case of epistemic overdetermination:

(T<sub>3</sub>) S comes to believe that her friend is unreliable with respect geography.

The causal theory incorrectly classifies S's belief that her friend is unreliable with respect geography as a non-defeater with respect to S's belief that Corsica is north of Sardinia because the justification for S to believe that Corsica is north of Sardinia is overdetermined.

The reasons-against-belief theory handles this case because S's belief that her friend is unreliable with respect geography is indeed a reason for S to not believe that Corsica is north of Sardinia on the basis of S's belief that her friend asserted that Corsica is north of Sardinia. Here we see an indication of the surprising, but unproblematic, propagation of defeaters under the reasons-against-belief theory of defeat alluded to above.

Consider next the second problem case, this one involving a deviant causal chain from one's belief to the loss of one's justification:

(T<sub>4</sub>) S comes to believe that things will go badly for S if S believes that her atlas shows that Corsica is north of Sardinia or that her friend asserted that Corsica is north of Sardinia. Because of S's belief that about things going badly for her, S ceases to believe that her atlas shows that Corsica is north of Sardinia and that her friend asserted that Corsica is north of Sardinia.

In this case, the causal theory incorrectly classifies S's belief that things will go badly for S if S believes that her atlas shows that Corsica is north of Sardinia or that her friend asserted that Corsica is north of Sardinia as a defeater with respect to S's belief that Corsica is north of Sardinia because that belief, through a deviant causal chain,<sup>31</sup> causes S to cease to be justified in believing that Corsica is north of Sardinia.

The reasons-against-belief theory too seems to fail with this case. A reason to disbelieve one's basis for belief is, presumably, also a reason to not form beliefs on that basis. Thus, since S's belief about how things will go for her is a reason for S to not believe that her atlas indicates that Corsica is north of Sardinia, it is also a reason for S to not believe that Corsica is north of Sardinia on the basis of S's belief that her atlas indicates that Corsica is north of Sardinia.

S's belief about things going badly for her might be a reason for S to not believe that Corsica is north of Sardinia on the basis of her atlas, but it is not a reason of the right sort. S's belief about how things will go for her is a practical reason for S to not believe

---

31 The belief about how things will go for her causes her to cease to hold the beliefs upon which she might have based her belief that Corsica is north of Sardinia, thus rendering her unjustified in believing that Corsica is north of Sardinia.

that Corsica is north of Sardinia on the basis of her atlas; it is not an epistemic reason. So, we should read the reasons-against-belief theory as:

**The Revised Epistemic Reasons-Against-Belief Theory:** S's belief that  $r$  would be a defeater with respect to S's belief that  $q$  if and only if S's belief that  $r$  would be an epistemic reason for S to not believe that  $q$  on some basis.

On that reading of the theory it delivers the correct result in (T<sub>4</sub>). Such a revision is not available to the causal theory since it makes no mention of reasons. This modification, too, is not without its problems; they will be addressed below. This is the final form of the reasons-against-belief theory of defeat; from here on, "reasons-against-belief theory" will refer to the revised epistemic reasons-against-belief theory.

Consider next a sequence of events in which one comes to have a defeater and then a defeater for that defeater:

(T<sub>1</sub>) S believes that S's atlas shows Corsica being north of Sardinia. S also believes that S's friend asserted that Corsica is north of Sardinia. S believes that Corsica is north of Sardinia solely on the basis of her belief that her atlas shows Corsica being north of Sardinia.

(T<sub>2b</sub>) S comes to believe that S's teacher asserted that Corsica is not north of Sardinia.

(T<sub>5</sub>) S comes to believe that S's geography teacher is unreliable.

At (T<sub>5</sub>), the causal theory incorrectly classifies S's belief that S's teacher asserted that Corsica is not north of Sardinia as a non-defeater because S's belief that her teacher is unreliable prevents S's belief that S's teacher asserted that Corsica is not north of Sardinia

from being efficacious.

The reasons-against-belief theory correctly classifies S's belief that S's teacher asserted that Corsica is not north of Sardinia as a defeater since it is indeed a reason for S to not believe that Corsica is north of Sardinia on the basis of S's belief that her teacher asserted that Corsica is north of Sardinia, albeit a reason that does not justify S in not believing that Corsica is north of Sardinia on the basis of S's belief that her teacher asserted that Corsica is north of Sardinia. Just as S's belief that  $p$  can be a reason for S to believe that  $q$  but fail to justify S in believing that  $q$  because of a defeater, S's belief that  $r$  can be a reason for S to not believe that  $q$  on the basis of S's belief that  $p$  even though it does not justify S in not believing that  $q$  on the basis of S's belief that  $p$  because of a defeater. Here we again see an indication of the unproblematic propagation of defeaters under the reasons-against-belief theory of defeat.

## 2.2 *The Reason-Neutralizing Theory*

Turn now to the double-agent case that the reason-neutralizing theory was unable to handle:

- (T<sub>1</sub>) S believes that S's atlas shows Corsica being north of Sardinia. S also believes that S's friend asserted that Corsica is north of Sardinia. S believes that Corsica is north of Sardinia solely on the basis of her belief that her atlas shows Corsica being north of Sardinia.
- (T<sub>6</sub>) S comes to believe that [S's atlas is unreliable but S's geography teacher says that Corsica is north of Sardinia].

The reason-neutralizing theory misidentifies S's belief that [S's atlas is unreliable but S's



geography teacher says that Corsica is north of Sardinia] as a non-defeater with respect to S's belief that Corsica is north of Sardinia since S's belief that [her atlas indicates that Corsica is north of Sardinia & (S's atlas is unreliable but S's geography teacher says that Corsica is north of Sardinia)] would be a reason for S to believe that Corsica is north of Sardinia.

The reasons-against-belief theory handles this case since S's double-agent belief is a reason for S to not believe that Corsica is north of Sardinia on the basis of her atlas. This is so even though S's double-agent belief, first, is not a reason for S to not believe that Corsica is north of Sardinia *simpliciter* and, second, is a reason for S to believe that Corsica is north of Sardinia *simpliciter*.

It might be thought that a counter-example to the reasons-against-belief theory might be developed along the same lines as the double-agent cases. Consider:

(T<sub>7</sub>) S comes to believe that [S's geography teacher asserted that Corsica is not north of Sardinia and S's geography teacher is unreliable].

Since S's belief that [S's geography teacher asserted that Corsica is not north of Sardinia and S's geography teacher is unreliable] is not a reason for S to not believe that Corsica is north of Sardinia on the basis of her atlas, the reasons-against-belief theory does not classify S's belief that [S's geography teacher asserted that Corsica is not north of Sardinia and S's geography teacher is unreliable] as a defeater. But since S's belief that [S's geography teacher asserted that Corsica is not north of Sardinia and S's geography teacher is unreliable] contains the conjunct [S's geography teacher asserted that Corsica is not north of Sardinia] and S's belief that S's geography teacher asserted that Corsica is not

north of Sardinia is a defeater with respect to S's belief that Corsica is north of Sardinia, then it might be thought that S's belief that [S's geography teacher asserted that Corsica is not north of Sardinia and S's geography teacher is unreliable] must also be a defeater with respect to S's belief that Corsica is north of Sardinia. After all, this seems to be the reasoning that was employed at (T<sub>6</sub>) to reject the reason-neutralizing theory of defeat.

This objection is mistaken. The relevant difference between (T<sub>6</sub>) and (T<sub>7</sub>) is that in (T<sub>6</sub>), [S's geography teacher asserted that Corsica is not north of Sardinia ] is not conjoined to a defeater with respect to S's belief that S's geography teacher asserted that Corsica is not north of Sardinia, whereas in (T<sub>7</sub>) it is. Pollock's insight, that even if S's belief that *p* is a reason for S to believe that *q*, it might not be the case that S's belief that (*p*&*r*) is a reason for S to believe that *q*, is correct and at play here. Even though S's belief that S's geography teacher asserted that Corsica is not north of Sardinia would be a reason for S to not believe that Corsica is north of Sardinia on the basis of S's atlas, S's belief that [S's geography teacher asserted that Corsica is not north of Sardinia and S's geography teacher is unreliable] would not be a reason for S to not believe that Corsica is north of Sardinia on the basis of her atlas. The mistake that Pollock made was in thinking that this feature of reasons for belief explains or is identical to defeat; it does not and is not because it is not always present in cases of defeat, as (T<sub>6</sub>) illustrates. Instead, as we will argue below, when this feature is present it does not explain, but rather is explained by, defeat.

### 3. *Explanatory Adequacy*

Our version of the reasons-against-belief theory of defeat is extensionally

adequate, which is enough to put it ahead of the causal theory, the reason-neutralizing theory, and Chandler's theory of defeat. However, the quietist theory of defeat is also extensionally adequate, albeit trivially. The true importance of the reasons-against-belief theory lies not in the fact that it is extensionally adequate; as noted in chapter 1, extensional adequacy is only one component of theoretical adequacy. The true importance of the reason-against-belief theory lies in the fact that it is extensionally adequate because it is explanatorily adequate, whereas the quietist theory is, obviously, silent on just what the mechanism of defeat is.

### 3.1 *The Causal Theory*

Even if it were extensionally adequate, the causal theory would be explanatorily inadequate because it does not tell us what causes the loss of justification, when it occurs. According to the reasons-against-belief theory, when S loses justification to believe that  $q$  due to defeat this is because, first, S was justified in believing that  $q$  because S believes that  $q$  on some basis and, second, S has acquired reason to not believe that  $q$  on that basis.

This suggests that S's justification for believing that  $q$  is best thought of as a function of S's reasons for believing that  $q$  and S's reasons for not believing that  $q$ . When S has a defeater with respect to S's belief that  $q$  but does not lose justification, this is because the defeater is a reason for S to not believe that  $q$  on some basis but S possesses another basis for belief to which the defeater is not relevant.

Notice that we have here introduced a theoretical framework for justification in which the reasons-against-belief theory is situated. Many have provided accounts of justification according to which one is justified in believing to whatever degree their

reasons support the belief, less whatever defeaters one might have with respect to that belief. These accounts then go on to focus on what it takes to have a reason for belief to the exclusion of a thorough discussion of what “less whatever defeaters one might have” amounts to. The framework we're suggesting is a way of filling in this neglected clause. To be clear, what we're suggesting is that:

**Justification:** S is justified in believing that  $q$  on a given basis if and only if S has more reason to believe that  $q$  on that basis than S has to not believe that  $q$  on that basis.

This is intended to be an ecumenical claim. We're trying to presume as little as possible about what a reason to believe (or a reason to not believe, for that matter) amounts to. We will address objections to this framework in chapter 4.

There is a puzzle that needs to be addressed here, however. There seems to be a mismatch between S's reasons for and against believing that  $q$ . S's belief that  $p$  seems to be a reason for S to believe that  $q$  *simpliciter*, but S's belief that  $r$  might well be a reason for S to not believe that  $q$ , not *simpliciter*, but rather only on the basis of S's belief that  $p$ . So, this does not seem to be a simple case of reasons for S to do something, i.e., believe that  $q$ , versus reasons for S to not do that same thing. Obviously, there is no straightforward function from S's reasons to take some action,  $\phi$ , and S's reasons to not take some other action,  $\psi$ , to whether or not S should  $\phi$ . This is problematic for the reasons-against-belief theory of defeat because an important aspect of that theory is its ability to explain the phenomena surrounding defeat and this competition of reasons for and against belief is necessary for that explanation.

There are two ways forward here. The first would be to establish that a reason for S to not believe that  $q$  on the basis of S's belief that  $p$  just is, in the context where the relationship between S's belief that  $p$  and S's belief that  $q$  is being considered, a reason for S to not believe that  $q$  *simpliciter*. The second is to establish that S's belief that  $p$  is never a reason for S to believe that  $p$  *simpliciter* but only a reason for S to believe that  $q$  on the basis of S's belief that  $p$ . Either way would result in a match between S's reasons for belief and S's reasons against belief.

The first approach will not work. Imagine a case in which S has multiple sources of justification, her beliefs that  $p_1$  and that  $p_2$ , for her belief that  $q$ . Were S to acquire a defeater that is relevant to S's belief that  $p_1$  but not to S's belief that  $p_2$ , then if S's belief that  $r$  were merely a reason for S to not believe that  $q$  *simpliciter*, then S should lose justification to believe that  $q$ . But this is not what happens, assuming that S's belief that  $p_2$  provides just as much justification for S's belief that  $q$  as S's belief that  $p_1$  does.<sup>32</sup> S's degree of justification for believing that  $q$  is unchanged. So, it cannot be that S's belief that  $r$  is a reason for S to not believe that  $q$  *simpliciter*. But then we are left with the same problem as before: S's belief that  $p_1$  is a reason for S to believe that  $q$  *simpliciter* but S's belief that  $r$  is not a reason for S to not believe that  $q$  *simpliciter*, it is a reason for S to not believe that  $q$  on the basis of S's belief that  $p_2$ .

The second approach is more promising. The suggestion is that all reasons for belief are self-promoting, i.e., they are not reasons to believe *simpliciter* but rather reasons to believe on the basis of themselves. S's belief that  $p$ , if a reason for S to believe

---

<sup>32</sup> See the discussion of the principle of cumulative justification in chapter 2 for the argument for the possibility of such a case.

that  $q$ , is a reason for S to believe that  $q$  on the basis of  $p$ . This would solve the problem facing the reasons-against-belief theory since it re-establishes the symmetry between the reasons for and against belief; they are not reasons for and against belief *simpliciter*, they are reasons for and against belief on a certain basis. What's more, the case for this treatment of reasons is quite strong. If S's belief that  $p_1$  were a reason for S to believe that  $q$  *simpliciter*, then were S to believe that  $q$  on the basis of some non-reason, say S's belief that  $p_2$ , S would still be justified in believing that  $q$ . But this is not right. Reasons only justify beliefs if those beliefs rest on those reasons.

Imagine, for example, that S reads in a reliable newspaper that it will rain today. Finishing her tea, S notices that the leaves at the bottom of the mug form a cloud-like shape and on that basis believes that it will rain today. Certainly, S is not justified in so believing, this despite the fact that she has a perfectly good reason to believe that it will rain. What explains this? We contend that what explains it is that she doesn't have a good reason to believe *simpliciter* that it will rain, but rather that she has a good reason to believe on the basis of the newspaper that it will rain. Note here that there is no difference in the content of the beliefs being considered. Rather, the difference between the beliefs is a difference in their basis. So, it isn't that two sources cannot both provide reason for one to believe that  $q$ ; they certainly can. It is just there is no such thing as a belief that  $p$  *simpliciter*; there is only a belief based on this source that  $p$  and a belief based on that source that  $p$ .

Consider the same case cast in a different light. Let's consider two possible beliefs, both with the same propositional content, that are differentiated by their bases.

Belief x is a belief that it will rain today formed on the basis of the newspaper while belief y is a belief that it will rain today formed on the basis of the tea leaves. Obviously, what S reads in the newspaper can only be a reason for S to form belief x, not y. Since we can differentiate beliefs with the same content by their bases for belief, a reason for belief can never be a reason to believe a given content *simpliciter* since it can never be a reason to form those beliefs that have that content but are based on some other reason. But this is equivalent to saying that what S reads in the newspaper is a reason for S to believe that it will rain today on the basis of the newspaper and, generally, all reasons for belief are reasons for belief on the basis of that reason.

It might be thought that the suggestion that reasons for belief are self-promoting faces clear counter-examples. Some reasons for belief seem to require an auxiliary belief to promote them. For example:

Suen, a chemistry student, notices that his litmus paper has turned blue in the solution. He knows that this means something, but he does not know what.

Checking the lab manual, he realizes that the litmus paper turns blue only in the presence of a base. Suen believes on the basis of his litmus paper that the solution is basic.

Although Suen's belief that the litmus paper is blue is a reason for him to believe that the solution is basic, he seems to only have a reason to believe that the solution is basic on the basis of the litmus paper after he forms the auxiliary belief that litmus paper turns blue only in the presence of a base. Thus, it seems that not all reasons for belief are self-promoting.

This example is misleading. There are two possible interpretations of what's going on before Suen looks at the lab manual: either he has a reason to believe that the solution is basic or he does not. If he does, then one can have a reason to believe some proposition without realizing that one has a reason to believe that proposition. If one can have a reason to believe some proposition without realizing that one has such a reason, then there is no reason why one couldn't have a reason to believe a proposition for that reason without realizing that one has a reason to believe that proposition for that reason. By the first option, the example above is no counter-example.

On the other hand, if the litmus paper is not a reason for Suen to believe that the solution is basic prior to looking at the lab manual, then it is the conjunction of two beliefs, i.e, the belief that the litmus paper is blue and the belief that litmus paper turns blue only in the presence of a base, that is the reason for Suen to believe that the solution is basic, not merely the belief that the litmus paper is blue. So, Suen does not have a reason to believe that the solution is basic on the basis of his belief that the litmus paper is blue; he has a reason to believe that the solution is basic on the basis of the conjunction of his beliefs that the litmus paper is blue and that litmus paper turns blue only in the presence of a base. This, too, poses no counter-example to the claim that reasons are self-promoting.

This way of thinking about epistemic reasons is clearly related to the distinction between doxastic and propositional justification. It may well be that these two ways of thinking of epistemic reasons, i.e., first, as providing justification to believe on this or that basis and, second, as providing propositional justification regardless of the basis of



one's belief and doxastic justification only when one's belief is based on that reason, are equivalent.<sup>33</sup> It doesn't matter for the reasons-against-belief theory whether these two are equivalent or not, so long as the self-promoting account of epistemic reasons is plausible.

That being said, it does seem that the idea that epistemic reasons are self-promoting casts a new light on the concepts of propositional and doxastic justification closer together. On this account, one never has propositional justification for believing something *simpliciter* but rather only propositional justification for believing on a particular basis. Doxastic justification, then, would be acquired when one did believe on that basis. This does not resolve the many issues surrounding the basing relationship and this is not the place to pursue this issue further, but the conceptual bridge that the claim that reasons are self-promoting forges between the two kinds of justification suggestive of new directions that research in that field might take.

### 3.2 *The Reason-Neutralizing Theory*

We turn now to the reason-neutralizing theory of defeat. Even if it were extensionally adequate, the reason-neutralizing theory would be explanatorily inadequate, as noted above, because there is no obvious reason why it should be that the fact that S's belief that  $(p \& r)$  would not be a reason for S to believe that  $q$  when S's belief that  $p$  is a reason to believe that  $q$  makes S's belief that  $r$  a defeater with respect to S's belief that  $q$ . This is particularly clear in cases where S believes that  $r$  and believes that  $p$  but does not believe that  $(p \& r)$ . There seems to be no reason in such a case that the status of S's belief that  $(p \& r)$  would be relevant to the status of S's belief that  $r$  or S's belief that  $q$ .

---

<sup>33</sup> Thanks to Landon Hedrick for pushing this issue.

It may well be that many who find the reason-neutralizing theory plausible are tacitly assuming that S's belief that *r* somehow renders S's belief that *p* no longer a reason for S to believe that *q*. However, as noted in the previous chapter, the prime proponent of the reason-neutralizing theory, Pollock himself, maintains an account of reasons according to which this cannot be the case.

On Pollock's account of reasons, S's belief that *p* is a reason for S to believe that *q* if and only if it is logically possible for S to be justified in believing that *q* on the basis of her belief that *p* (1986, 36). But S's belief that *r* cannot change whether or not it is logically possible for S to be justified in believing that *q* on the basis of her belief that *p*. Any view according to which a defeater changes a beliefs status from reason to non-reason will be inconsistent with any view that is committed to the reason-for relation holding necessarily. What is more, even if one doesn't go in for views of reasons that hold this commitment, one is still left with the burden of explaining just how S's belief that *r* renders S's belief that *p* a non-reason for S to believe that *q*. This might not be an impossible task, but it has yet to be done.<sup>34</sup>

The reasons-against-belief theory can explain why, when S's belief that *r* would be a defeater with respect to S's belief that *q*, S's belief that (*p*&*r*) might not be a reason for S to believe that *q* even though S's belief that *p* is a reason for S to believe that *q*. According to this theory, when S's belief that *r* is a defeater with respect to S's belief that *q* it is a reason for S to not believe that *q* on some basis, say, S's belief that *p*. Whatever else must be true of an epistemic reason for S to believe that *q*, it must auger in favor of S's believing that *q*. But, while S's belief that *p* might auger in favor of S's believing that

---

34 Or even attempted, as far as is evident.

$q$ , that might be neutralized when  $p$  is conjoined with  $r$ . Since S's belief that  $r$  augers in favor of S's not believing that  $q$  on the basis of S's belief that  $p$ , the net effect of S's belief that  $(p \& r)$  might auger neither in favor nor against S's believing that  $q$ . Thus, S's belief that  $(p \& r)$  might not be a reason for S to believe that  $q$  even though S's belief that  $p$  would be.

### 3.3 *Further Explanations*

Better still, the reasons-against-belief theory of defeat can answer the question that motivated this project. Recall that the question arose out of an apparent asymmetry:

S need not believe or be justified in believing that her belief that  $p$  is a reliable indicator of whether or not  $q$  in order for her to be justified in believing that  $q$  on the basis of her belief that  $p$ , but were she to come to have good reason to believe that her belief that  $p$  is not a reliable indicator of whether or not  $q$ , then she would lose that justification.

There seems to be no positive meta-justification requirement on justification but there does seem to be a negative meta-justification requirement. That asymmetry is puzzling.

According to the reasons-against-belief theory, S's belief that  $p$  is a reason for S to believe that  $q$  only if it is a reason for S to believe that  $q$  on the basis of her belief that  $p$ . If S has reason to believe that [her belief that  $p$  is an unreliable indicator of whether or not  $q$ ], then S has reason to not believe that  $q$  on the basis of  $p$ . The apparent asymmetry in meta-justification requirements turns out to be a red herring; the puzzling phenomenon is explained simply in terms of competition between reasons for and against belief.

Recall, also, some of the reasons given in the previous chapter for rejecting the

quietist theory of defeat: variation in source-specificity between defeaters and variation in the strength of defeaters. The reasons-against-belief theory is able to explain these phenomena.

First, some defeaters are source-specific because they are reasons to not believe on the basis of those specific sources while others are source-neutral because they are reasons to not believe on any basis whatsoever. The reasons-against-belief theory is able to make sense of the way that defeaters were characterized in the beginning of chapter 2. Recall that source-specific defeaters were said to render inadequate the support provided by a specific source while source-neutral defeaters were said to render inadequate the support provided by any source. It was suggested that source-neutral defeaters do not weaken the support provided by a source, but rather they put pressure on one's belief in such a way that one's sources of support for that belief no longer provide enough support for that belief. We can now see that this is really what's going on in both cases of source-neutral defeat and cases of source-specific defeat. In both cases, the degree of support provided by one's sources remains unchanged; we do not need to introduce a mechanism by which reasons or support is rendered inert, attacked, or destroyed. Rather, defeaters render support inadequate by raising the bar for what degree of support is adequate.<sup>35</sup> When one gains reasons to not believe that *q*, one doesn't lose one's reasons to believe that *q*; rather, one simply needs more reason to believe that *q* in order to outweigh the

---

35 Note that the loss of some thing's reason status is not incompatible with the reasons-against-belief theory of defeat. It might be that in some or all cases of efficacious defeat, the one-time reasons to believe that *q* lose their reason status. However, this phenomenon would be explained by the reason-against-belief theory rather than explaining it since one does not always lose justification in cases of defeat and, thus, the one-time reasons to believe that *q* cannot always lose their reason status in cases of defeat. We do insist, however, that the metaphysical status of reasons is unclear if something can be turned from a reason to believe into a non-reason to believe.

reasons against believing that  $q$ .

Second, the strength of defeaters on the reasons-against-belief theory is no more mysterious than the strength of evidence, which is to say that since evidence and defeaters are taken to be mirror images of one another on this theory, we should expect that any explanation of the strength of evidence to apply to defeaters as well.

Furthermore, if there is such a thing as partial defeat, i.e., defeat that reduces but does not eliminate one's justification, that can also be explained by this theory by amending the justification framework in the following way:

**Degree of Justification:** The degree of justification  $S$  has for believing that  $q$  on a given basis is determined by how much more reason  $S$  has to believe that  $q$  on that basis than  $S$  has to not believe that  $q$  on that basis.

If a defeater gives one more reason to not believe that  $q$  on the basis of one's belief that  $p$  than one previously had, but not so much more that one comes to have less reason to believe that  $q$  on the basis of the belief that  $p$  than one has to not believe that  $q$  on the basis of the belief that  $p$ , then, according to the amended framework above, one's degree of justification for believing that  $q$  on the basis of the belief that  $p$  will decrease but persist.

#### 4. *Epistemic Reasons For and Against Belief*

A problem still remains for thinking of defeaters as reasons to not believe on a particular basis. The difference between epistemic reasons and non-epistemic reasons is often characterized in the following way: epistemic reasons make the truth of a proposition believed more or less likely whereas non-epistemic do not. However, a

reason to not believe that  $q$  *qua* reason to not believe that  $q$ , as compared to, say, reason to believe that not- $q$ , would not make  $q$  any more or less likely to be true. Thus, the reasons-against-belief theory seems to be at odds with this common way of characterizing epistemic reasons.<sup>36</sup>

This problem evaporates when one considers the following dilemma: either all defeaters are epistemic reasons or some are not. If some epistemic defeaters are not epistemic reasons, then the reasons-against-belief theory is well positioned to handle that by dropping the epistemic qualifier in “epistemic reason to not believe.” However, there is no obvious reason to think that there are any epistemic defeaters that are non-epistemic reasons.

Taking the other horn of the dilemma, then, we assume that all epistemic defeaters are epistemic reasons. Consider the paradigm examples of defeaters with respect to S's belief that  $q$ : evidence that supports not- $q$  and evidence that supports the belief that [S's belief that  $p$  is not a reason for S to believe that  $q$ ]. In both of these cases, the defeaters bear on the likelihood of the truth of some proposition:  $q$  in the first case and [S's belief that  $p$  is a reason for S to believe that  $q$ ] in the second. Thus, by this way of distinguishing epistemic from non-epistemic reasons, defeaters are epistemic reasons. What's more, they are reasons for S to not believe that [ $q$  on the basis of S's belief that  $p$ ] in virtue of the fact that they are epistemic reasons for these beliefs. Let us stipulate:

**Epistemic Reasons Against Belief:** An epistemic reason to not believe that  $q$  is something that is a reason to not believe that  $q$  in virtue of being an epistemic reason for some belief.

---

36 Thanks to Andrew Spaid for this objection.

Recall the case that necessitated the inclusion of the epistemic modifier to the theory:

- (T<sub>4</sub>) S comes to believe that things will go badly for S if S believes that her atlas shows that Corsica is north of Sardinia or that her friend asserted that Corsica is north of Sardinia. Because of S's belief that about things going badly for her, S ceases to believe that her atlas shows that Corsica is north of Sardinia and that her friend asserted that Corsica is north of Sardinia.

The stipulation introduced above allows the reasons-against-belief theory to correctly classify S's belief about how things will go for her as a non-defeater since, although it is a reason for S to not believe that Corsica is north of Sardinia and it is an epistemic reason to believe something, e.g., that she had better stop believing that her atlas shows that Corsica is north of Sardinia, it is not a reason of the former kind in virtue of being an epistemic reason. Rather, it is a reason of the former kind in virtue of being a practical reason.

Consider, on the other hand, a paradigm case of defeat:

- (T<sub>1</sub>) S believes that S's atlas shows that Corsica is north of Sardinia. S also believes that S's friend asserted that Corsica is north of Sardinia. S believes that Corsica is north of Sardinia solely on the basis of her belief that her atlas shows Corsica being north of Sardinia.
- (T<sub>2a</sub>) S comes to believe that S's atlas is unreliable.

In this case, S's belief that her atlas is unreliable is a reason for S to not believe that Corsica is north of Sardinia on the basis of her atlas in virtue of the fact that it is a reason for S to believe that the fact that her atlas indicates that Corsica is north of Sardinia does

not make it more likely that Corsica is north of Sardinia. So, the stipulation resolves the problem.

It is worth noting that this connection between epistemic reasons to believe and epistemic reasons to not believe is a piece of the puzzle of understanding defeat. Pollock characterized defeat as follows:

To say that  $P$  is a prima facie reason for  $S$  to believe-that- $Q$  is to say that in the absence of any other information  $S$  is justified in believing that it would not be true that  $P$  unless it were true that  $Q$ . Let us symbolize this subjunctive conditional as " $P \Rightarrow Q$ ". Then  $P$  is a prima facie reason for  $S$  to believe-that- $Q$  iff  $S$  is prima facie justified<sup>37</sup> in believing-that- $(P \Rightarrow Q)$ . A defeater must be a reason for thinking that this conditional is false, i.e., a reason for  $S$  to believe that  $\sim(P \Rightarrow Q)$ . (1974, 42)

The reasons-against-belief theory of defeat, with the above stipulation, allows us to explain why this might be so. If  $S$  has a reason to believe that it is not true that [ $p$  would not be the case unless  $q$  were also the case], it is eminently plausible that this is a reason for  $S$  to not believe that [ $q$  on the basis of her belief that  $p$ ]. Furthermore, it is in virtue of  $S$ 's belief that  $r$  being an epistemic reason for  $S$  to believe that [it is not true that  $p$  would not be the case unless  $q$  were also the case] that  $S$ 's belief that  $r$  is a reason for  $S$  to not believe that [ $q$  on the basis of  $S$ 's belief that  $p$ ]. The reasons-against-belief theory of defeat, then, explains why being an epistemic reason of this sort would make a belief a

---

37 According to Pollock:

$P$  is prima facie justified for  $S$ " means "It is necessarily true that if  $S$  believes (or were to believe) that  $P$ , and  $S$  has no reason for thinking that it is false that  $P$ , then  $S$  is (or would be) justified in believing that  $P$ . (1974, 30)



defeater.

Reasons to believe that not- $q$  or to believe that one's evidence does not support  $q$  are both reasons to not believe that  $q$  on the basis of your evidence. There may well be other kinds of reasons to believe that also qualify as such. Below we will briefly discuss the possibility that a reason to believe that one has no reasons to believe that  $q$  is a reason to not believe that  $q$ . In chapter 5 we will discuss the possibility that higher-order evidence constitutes a distinct kind of reason to not believe that  $q$ . The reasons-against-belief theory has no essential commitment to what kinds of epistemic reasons to believe constitute epistemic reasons to not believe that  $q$ ; what it is committed to is the claim that the only reasons that can be epistemic reasons to not believe that  $q$  are reasons such that they are reasons to not believe that  $q$  in virtue of being epistemic reasons to believe something.

### 5. *Unjustified Belief and Justified Unbelief*

One of the inspirations for the reasons-against-belief theory of defeat was the suggestion that in standard cases of defeat, one does not just cease to be justified in believing but also becomes justified in not believing. This lead to the suggestion that:

**Justification:** S is justified in believing that  $q$  on a given basis if and only if S has more reason to believe that  $q$  on that basis than S has to not believe that  $q$  on that basis.

There's a pleasing symmetry to this thesis, a symmetry that is upset by the observation that, although one can only be justified in believing that  $q$  if one has more reason to believe that  $q$  than one has to not believe that  $q$ , one is justified in not believing that  $q$

even in the absence of defeaters. This would mean that, if the reasons-against-belief theory is correct, one needs more reason to believe than to not believe in order to be justified in believing but need not have more reason to not believe than to believe in order to be justified in not believing.

Imagine, for example, that you're considering whether or not it snowed today. You haven't yet left bed and you're listening to the radio, but it hasn't broadcast the weather yet. You have no reason to believe that it has snowed, but you also don't appear to have any defeaters with respect to the the belief that it has snowed. When you're in that situation, the result that would seem to sit most comfortably with the theoretical framework we've suggested is that you are neither justified in believing nor in not believing that it snowed. But, while it is true that you are not justified in believing that it did not snow, you do seem to be justified in not believing that it did snow, i.e., you do seem to be justified in withholding belief. What can we make of this?

First, note that this poses no damning problem for the reasons-against-belief theory of defeat. It is an explanatory burden for the theory, but it is not a burden peculiar to the reasons-against-belief theory. Even a theory of justification that had no room for defeaters at all would need to explain why we are justified in withholding belief in the apparent absence of all reasons. It is not inconsistent with the reasons-against-belief theory that there be an epistemic principle that being justified in not believing is a default, that one need no reason to not believe in order to be justified in not believing.

Second, note also that not everyone sees a problem here. Some have reported to me that they cannot imagine how not having reason to believe that  $q$  could fail to be a

reason to not believe that  $q$ . It is simply obvious to them that there is this connection between not having a reason to believe and having a reason to not believe.<sup>38</sup> If you're in that camp, then you can skip the rest of this section. If, however, you are even mildly discomforted by the reasoning above, then there are two routes by which we might try to mitigate this mild discomfort.

It might be suggested that when one has no reason to believe that  $q$  and no reason to not believe that  $q$ , then one is justified in believing and justified in not believing. The idea would be that when one is in a state of equipoise, one may permissibly believe that  $q$  or withhold judgment with respect to whether or not  $q$ . This view has more currency than epistemologists might wish to believe. At least among students, the view that when there are no reasons for or against believing something, or when those reasons are balanced, belief or withholding of belief are both acceptable is not uncommon. Furthermore, this situation might change in different epistemic contexts. It might be unacceptable to accept a scientific hypothesis when the reasons for and against believing that hypothesis are balanced, but perfectly acceptable to do the same in a non-scientific context.

However, if that response strikes you as unacceptable, there is another more palatable response. A lack of reasons to believe that  $q$  is itself a reason to believe that not- $q$ , and thus, also, to not believe that  $q$ . It is an oft repeated but simply false aphorism that an absence of evidence is not evidence of absence. At least in some cases, failure to observe evidence for some claim is a reason to disbelieve that claim. Think again about the snow case. The probability that it has snowed is much greater given that the radio

---

38 This might be related to Bergmann's no-reason defeaters (2006a, 159f12).

announces that it has snowed than it is given that the radio has not announced it.

Similarly, it is much more likely that, at any given point, the radio will announce that it has snowed given that it has snowed than it is given that it hasn't snowed, i.e., it's very unlikely that the radio will announce it has snowed unless it has snowed. So, while the radio announcing it has snowed is evidence that it has snowed, it is also true that it having snowed is evidence that at any given point the radio will announce that it has snowed. Conversely, it not having snowed is evidence that at any given point the radio will not announce that it has snowed and, what matters for the argument here, that the radio not announcing that it has snowed at any given point is some reason, perhaps not much at all, to believe that it has not snowed and, thus, some reason to not believe that it has snowed. An absence of reason to believe is itself a reason to not believe.

This second route is particularly nice since it preserves the connection between epistemic reasons to believe and epistemic reasons to not believe that was stipulated above in section 4. It is, however, a bit confusing as to how one can have a reason to not believe that  $q$  on some basis when one has no bases for believing that  $q$ . This might be a limiting case of defeat; one has a reason to not believe on some basis because one has a reason to believe that one has no basis for believing, just as one has a reason to not go swimming in a pool out back if one believes that there is no pool out back. In any case, it is clear that the apparent asymmetry between justification for believing and justification for not believing when one is in a state of equipoise is no problem for the reasons-against-belief theory of defeat.

## 6. *Conclusion*

The reasons-against-belief theory of defeat developed in this chapter may not seem particularly bold. There is a temptation to shrug at the difference between one's not having a reason to believe and one's having a reason to not believe. But as we've demonstrated, that difference provides a surprisingly robust explanation of the phenomenon of defeat. It explains why a defeater might cause one to lose justification and why it might not. It explains why the conjunction of a defeater with respect to one's belief that  $q$  and a reason for one to believe that  $q$  is sometimes a reason to believe that  $q$  and sometimes not a reason to believe that  $q$ . It explains why having a reason to believe that one's evidence for  $q$  is not a reliable indicator of  $q$  might reduce one's justification for believing that  $q$  even though one needn't have believed that one's evidence for  $q$  was a reliable indicator in the first place. It can explain what occurs when one has a defeater for a defeater and why some defeaters might be stronger than others. It can explain why some defeaters are source-specific and others are source-neutral. Most importantly, it explains why a defeater is a defeater and what relationship a defeater bears to one's reasons for belief. It does all this while remaining extensionally adequate, handling the cases that proved problematic for other theories of defeat.

Obviously, the reasons-against-belief theory does not provide a complete explanation of our justificatory practices. One might reasonably ask what makes one reason to not believe stronger than another or what makes something a reason to not believe in the first place. While the reasons-against-belief theory does not answer these questions, by putting defeaters on the same level as reasons for belief, it reduces the

number of questions that need answering. We should expect that the answer we get to the question of what makes something a reason to believe that  $q$  will also illuminate what makes something a reason to not believe that  $q$ . Contrast this with the situation we'd be in if we accepted the reason-neutralizing theory of defeat: we might answer the question of what makes something a reason to believe that  $q$  without shedding any light on the question of what makes a proposition such that when it is conjoined with a reason to believe that  $q$ , the resulting conjunction is not a reason to believe that  $q$ .

These results speak strongly in favor of the reasons-against-belief theory of defeat, but we should not be surprised if it turns out not to be the correct theory of defeat. Given the superficial treatment that defeat has received thus far, it is better to see the reasons-against-belief theory as a spur to further research into defeat. It us to be hoped that it is superseded by another theory that is yet more illuminating, yet more powerful. The reasons-against-belief theory of defeat raises the bar for adequacy when it comes to a theory of defeat; it provides a model of the sort of theory that we should be trying to construct.

In the remaining chapters, we take on two tasks. The first, in chapter 4, is to defend that reasons-against-belief theory from the argument that defeaters cannot be understood as the mirror image of reasons because there are significant asymmetries between reasons to believe and defeaters. The second, in chapter 5, is to see whether or not reasons-against-belief theory has any implications for the applications of defeat highlighted in chapter 1.

# CHAPTER 4

## SYMMETRY OF REASONS

In the previous chapter we argued that a defeater with respect to some belief is a reason to not hold that belief on a particular basis. One of the attractions of this theory is that it unifies the apparently distinct phenomena of gain and loss of justification as different aspects of a single phenomenon: the interplay of reasons for and against believing on a particular basis.

However, there are significant problems for this attractive view already present in the literature on defeat. These problems take the form of apparent asymmetries between reasons to believe and defeaters that, if more than merely apparent, cast doubt on the claim that reasons to believe and defeaters are mirror images of each other. If these asymmetries are real, then some more complicated mechanism will be necessary to explain the loss of justification, when it occurs.

The first of these asymmetries, suggested by the work of John Pollock (1974) and Michael Bergmann (1997, 2005, 2006a, 2006b), is as follows:

**Justification-Asymmetry:** While (a) S must be justified in believing that  $p$  in order for S's belief that  $p$  to be efficacious as a reason for S to believe that  $q$ , it is not the case that (b) S must be justified in believing that  $r$  in order for the S's belief that  $r$  to be efficacious as a defeater with respect to S's belief that  $q$ .

If this asymmetry holds, then there is a significant dissimilarity between reasons

to believe and defeaters.

Second is a family of asymmetries concerning the awareness conditions on the efficacy of reasons to believe and defeaters:

**Awareness-Asymmetry:** While (1a) S must believe/realize that  $p$  in order for the fact that  $p$  to be efficacious as a reason for S to believe that  $q$ , it is not the case that (1b) S must believe/realize that  $r$  in order for the fact that  $r$  to be efficacious as a defeater with respect to the belief that  $q$ .

**Meta-Awareness-Asymmetry:** While (2a) S need not believe/realize that  $p$  is a reason for S to believe that  $q$  in order for the belief that  $p$  to be efficacious as a reason for S to believe that  $q$ , it is not the case that (2b) S need not believe/realize that S's belief that  $r$  is a defeater with respect to S's belief that  $q$  in order for S's belief that  $r$  to be efficacious as a defeater with respect to S's belief that  $q$ .

**Basing-Asymmetry:** While (3a) S must base S's belief that  $q$  on S's belief that  $p$  in order for S's belief that  $p$  to be efficacious as a reason for S to believe that  $q$ , it is not the case that (3b) S must base S's failure to believe/ceasing to believe that  $q$  on S's belief that  $r$  in order for S's belief that  $r$  to be efficacious as a defeater with respect to the belief that  $q$ .

If any of these asymmetries hold, then defeaters and reasons to believe cannot be of a kind since they function in significantly different ways.

Given the success of the reasons-against-belief theory demonstrated in the previous chapter, the case for the proposed asymmetries would have to be strong indeed to pose a true challenge to that theory. But the case for most of the asymmetries seems to



be nothing more than intuitions about cases, and unclear intuitions at that. We argue that these intuitions are not robust enough to support the proposed asymmetries and that, where more substantial arguments are provide, they are all fatally flawed. Furthermore, not only are there no strong reasons to accept the asymmetries, there are some significant reasons to reject them, too.

Before proceeding to the discussion of the apparent asymmetries, we should remark on the term “efficacious.” The effect most closely related to a reason to believe is providing justification for that belief and the effect most closely related to a defeater with respect to that belief is the loss of justification for that belief. But the justification in question might be propositional or doxastic. Furthermore, a reason to believe might be efficacious but fail to provide any justification because one has stronger reason to disbelieve and a defeater might be efficacious but remove no justification because of epistemic overdetermination.<sup>39</sup> Contrast the state such reasons and defeaters would be in to the state a belief *qua* reason for S believe that *q* is thought to be in when it is unjustified. In such a case, the reason is thought to not count in favor of S believing that *q* at all. We can call reasons to believe and defeaters of the former kind efficacious and those of the latter kind inefficacious.

This can all be simplified by stipulating that throughout this chapter, except where explicitly noted otherwise, the only kinds of reasons to believe we will be considering are beliefs *qua* reasons to believe and, similarly, the only kinds of defeaters we will be considering are beliefs *qua* defeaters.<sup>40</sup> Furthermore, in the case of reasons to believe, and

---

<sup>39</sup> See chapter 2 for more on epistemic overdetermination.

<sup>40</sup> As compared to, say, perceptual experiences *qua* reasons to believe/defeaters. This distinction will be discussed in section 1 below.

except where explicitly stated otherwise, “efficacious” will refer to that power in virtue of which a reason to believe provides one with doxastic justification for a belief when one has no defeaters with respect to that belief. Similarly, in the case of defeaters and except where explicitly stated otherwise, “efficacious” will refer to that in virtue of which a defeater renders a belief doxastically unjustified when one has but a single source of support for that belief.

### 1. *Justification-Asymmetry*

The first asymmetry to consider is what we can call justification-asymmetry. The thought is that although reasons to believe need to be justified in order to be efficacious, defeaters need not be:

**Justification-Asymmetry:** While (a) S must be justified in believing that  $p$  in order for S's belief that  $p$  to be efficacious as a reason for S to believe that  $q$ , it is not the case that (b) S must be justified in believing that  $r$  in order for the S's belief that  $r$  to be efficacious as a defeater with respect to S's belief that  $q$ .

Consider the following case:

You justifiably believe on the basis of your justified belief that you left your pen on the coffee table that there is where it remains. Later, you come to unjustifiably believe that your pen is on the end table. You realize that your pen cannot be both on the coffee table and the end table.

We argued in chapter 3 that the unjustified belief that the pen is on the end table, were it justified at least, would render your belief that your pen is on the coffee table unjustified because it is a reason for you to not believe that the pen is on the coffee table on the basis

of your belief that you left your pen on the coffee table (or any other basis, for that matter).

But, according to the current objection, the defeater would also accomplish this in its current, unjustified state. If, as many epistemologists maintain, reasons to believe are only efficacious when they are justified, then reasons to not believe should only be efficacious when they are justified, too. Otherwise, there would be an asymmetry between reasons for and against belief that would require an explanation. So, at least one of the following four claims is true: (i) it cannot be that a defeater is essentially a reason for one to not believe, (ii) the reasons-against-belief theory of defeat has a significant explanatory burden to discharge, (iii) reasons to believe needn't be justified in order to be efficacious after all, or (iv) an unjustified defeater is not efficacious. We want to avoid the first two of these claims, so we will argue that one of the latter two must be true.

We've introduced the apparent asymmetry that is the topic of this section as follows: a reason to believe must be justified in order to be efficacious but a defeater needn't be justified in order to be efficacious. Though helpful to get our bearings, this introduction is technically misleading. To be clear, the topic of this chapter is not whether or not justification is a necessary condition on the efficacy of defeaters, *per se*. Many believe that non-belief states, e.g., perceptual experiences, can provide justification, even though they themselves need no, or indeed can have no, justification. Bergmann supports this claim with the following case:

Suppose I'm hiking on a trail in the mountains and that I read in a guide book, which I reasonably believe to be reliable, that upon reaching the peak at the end of

the trail, one will not be able to see any lakes even when the weather is clear. I thus form the presumably justified belief that I won't be able to see a lake from there. But when I get to the peak at the end of the trail, I look down and see clearly what is obviously the lake from which the hiking trail began. That newly acquired visual experience makes my belief that one can't see any lake from the peak cease to be a justified one. (2006a, 155-156)

There are some doubts about Bergmann's description of the case. It is not so obvious that it is the visual experience of the lake, rather than the justified belief he (presumably) forms on the basis of that experience that does the defeating. If, for example, he reached the peak, saw the lake, but was lost in thought and so did not form the belief that there is a lake visible from the peak, would his belief that there is no lake visible from the peak be justified or not? The answer to this question is not obvious; we address a related point at length in section 2.1 below. But for the sake of the current argument, we'll grant Bergmann's claim: visual experience can defeat beliefs.

Visual experience is generally taken to not be the sort of thing that can be justified. It follows, then, that some defeaters are not justified. This highlights an important distinction: we should not be any more concerned with the question, "can an efficacious defeater be unjustified?" than we should be with the question, "can an efficacious justifier be unjustified?" Our question is, rather, "Are defeaters relevantly similar to justifiers with respect to the degree of justification necessary for efficacy?" If defeaters are relevantly similar to justifiers, then the answer to the second question above will also answer the first. Take whatever justification conditions are necessary for a

reason to believe to be efficacious; our topic is whether or not there are more, fewer, or otherwise different justification conditions that a defeater must satisfy in order to be efficacious. Thus, the topic of this section concerns a justification-asymmetry between reasons to believe and defeaters.<sup>41</sup>

There are two arguments given for justification-asymmetry. The first argument employs cases in which an agent has an unjustified defeater with respect to an antecedently justified belief. The defender of justification-asymmetry contends that we evaluate the cognitive state of this agent negatively and that this negative evaluation is explained by the unjustified defeater removing the justification from the antecedently justified belief.

According to the second argument, a belief is justified only if it is an appropriate response to one's evidence. But, the argument continues, even unjustified beliefs are part of one's evidence. Thus, holding a belief is an inappropriate response to one's evidence when one's evidence contains an unjustified belief that conflicts with it. Thus, unjustified defeaters are still efficacious.

### 1.1 *The First Argument*

The first argument depends exclusively on the purported intuitive reactions to a thought-experiment. Still, if the intuitions are strong and clear, then the reasons-against-belief theory denies them only at a theoretical cost. It is worth, then, spending time scrutinizing the thought-experiment and our purported intuitive reactions to them.

---

41 We do deviate from this model in section 1.4 and 1.5 by demonstrating how, if reasons to believe must be justified in order to be efficacious, justification-symmetry would permit explanations of various phenomena. For most of this chapter, however, we try to remain neutral on just what are the efficacy conditions for reason to believe.

Consider the following case presented by Pollock:

[...] suppose that S is trying to predict the color of marbles drawn from an urn.

Let us suppose that fifteen marbles have been drawn so far, and they have all been red. S might conclude inductively that the next marble will also be red. If he has no reason for thinking either that the next marble will not be red or that there is something peculiar about the urn so that the inductive generalization does not provide him with a way of getting to know whether the next marble will be red, then it seems we would agree that he would be justified in believing that the next marble will be red. In other words, he has a *prima facie* reason for thinking that the next marble will be red. But in fact, this is not enough. Let us suppose that S believed, without justification, that he has ESP and that on that basis he can tell that the next marble will be black. Although his belief in this defeater is not justified, the mere fact that he does believe it would make it irrational for him to believe that the next marble will be red and so would prevent him from being justified in believing that. (1974, 43-44)

Pollock does not tell us why the mere fact that S believes that he has ESP and that on that basis he can tell that the next marble will be black makes it irrational for him to believe that the next marble will be red. It seems that we are supposed to have an intuitive response to this case that favors this conclusion. And it does seem that it would be irrational for S to be in the cognitive state that Pollock describes, i.e., the state of both believing that he has ESP and on that basis can tell that the next marble will be black and believing that the next marble will be red. Following our standard procedure, we will read

“irrational” here as “unjustified.” So, we grant Pollock that S is unjustified in believing that he has ESP and on that basis can tell that the next marble will be black and believing that the next marble will be red. This is not enough to support justification-asymmetry since there are a number of interpretations of this intuition that are consistent with justification-symmetry.

First, the conclusion that an unjustified belief has prevented S from being justified in believing that the next marble will be red is ambiguous. An unjustified belief can be either completely without justification or merely without sufficient justification. Pollock writes that S believes without justification that he has ESP. This means either that S has no justification or that S has some justification but not enough to be justified *simpliciter*.

Say it's the latter. Perhaps S has played this game many times in the past. On some occasions, he has had a funny feeling followed by his drawing a black marble. Let's say that this is some reason for him to believe that he has ESP that allows him to know when the next marble will be black. However, S is ignoring the many times that he has had this feeling and the next marble was not black. Just the same, it is true that given only past frequencies S has a better-than-luck chance of being right when he believes that the next marble will be black after having the funny feeling. So, it is plausible to think that S has some justification for his belief that the next marble will be black but not much, certainly not enough for S to be sufficiently justified in holding that belief.

If something like this is what is going on in Pollock case, then even if S is unjustified in being in the cognitive state that Pollock describes, this does not favor justification-asymmetry. This is so because S has beliefs that jointly constitutes a defeater

with respect to the belief that the next marble will be red and that are efficacious reasons to believe. Since *S* has, by hypothesis, some justification for his belief that he has ESP and on that basis can tell that the next marble will be black, *S*'s reasons for believing this satisfy the efficacy conditions on reasons to believe. Furthermore, these reasons to believe also constitute a defeater for *S*'s belief that the next marble will be red. Thus, on this interpretation, this case could not possibly support justification-asymmetry. This is at least a plausible explanation for our intuition that *S* is unjustified in being in the cognitive state that Pollock describes and it must be ruled out by the defender of justification-asymmetry in order for our intuition about this case to favor their conclusion.

On the other horn of the dilemma, imagine that *S*'s belief that the next marble will be black has no justification whatsoever. We must still be cautious when interpreting our intuitions about this case. We must distinguish between the justificatory status of individual cognitive states and the justificatory status of sets of one's cognitive states. To move immediately from the claim that the conjunction of beliefs is unjustified to the claim that each belief is unjustified is to commit the fallacy of division. There is no easy route from the fact that *S* is unjustified in believing both that the next marble will be red and that [he has ESP and on that basis can tell that the next marble will be black] to the conclusion that if *S* unjustifiably believes that [he has ESP and on that basis can tell that the next marble will be black], then *S* is unjustified in believing that the next marble will be red. Presumably, if *S* is unjustified in believing that *r*, then *S* is unjustified in being in the following cognitive state: believing that *q* and that *r*. This is so even if *S* is justified in believing that *q*. So, our initial intuition might be due entirely to *S*'s unjustifiably



believing that he has ESP and on that basis can tell that the next marble will be black.<sup>42</sup>

The defender of justification-asymmetry can accommodate this by constructing a case that controls for this noise. Consider a case in which S unjustifiably believes that he had oatmeal for breakfast instead of believing that he has ESP and on that basis can tell that the next marble will be black, but which is otherwise identical to Pollock's case. It is intuitively plausible that S is less justified in being in the cognitive state of believing that [S has ESP and on that basis can tell that the next marble will be black] and that the next marble will be red than S is in being in the cognitive state of believing that S had oatmeal for breakfast and that the next marble will be red. The fact that S is unjustified in believing that he has ESP and on that basis can tell that the next marble will be black is not enough to explain this intuition since S's belief that he had oatmeal for breakfast is also unjustified. The relationship between S's belief that he has ESP and on that basis can tell that the next marble will be black and S's belief that the next marble will be red seems to be at least partially responsible for the epistemic mess in which S finds himself.

Still, even this is not enough to favor justification-asymmetry. Granting that our intuitions are specific enough to favor the claim that the unjustified status of S's cognitive state as described by Pollock is due to more than just the unjustified status of S's belief that the next marble will be black, it still doesn't follow that S's belief that the next marble will be red would be unjustified. It is plausible that our intuitions are partly explained by the conflict between the contents of S's beliefs itself. To see this, consider a case exactly like Pollock's above except that S's has drawn no marbles but believes the next marble

<sup>42</sup> The reader may balk at applying the terms "justified" and "unjustified" to anything other than one's holding of individual beliefs. This qualm can be accommodated. S's believing that *p* and that *r* has some epistemic status and, in Pollock's case, it's a bad one. The question under consideration, then is whether this badness due to S's belief that *r* rendering S's belief that *p* unjustified or to something else.

will be red on the basis of reading the tea leaves. Assume that S is as unjustified in believing on the basis of tea leaves as she is in believing on the basis of ESP. This is a case in which S's overall cognitive state is unjustified. Furthermore, this strikes me as a case in which the justificatory status of neither of S's beliefs fully accounts for the negative epistemic evaluation of S's overall cognitive state. At least some of S's epistemic problems are due to the fact that she believes conflicting things, i.e., that next marble will be black and that the next marble will be red.

If you're not convinced, contrast the previous case with the following one: instead of unjustifiably believing that the next marble will be red on the basis of the tea leaves, S unjustifiably believes that he will have oatmeal for breakfast the next day. Despite the fact that the beliefs in this case are as unjustified as the beliefs in the previous case, S's cognitive state in the previous case is epistemically worse than it is in this case. The difference in epistemic status must stem from the difference in contents: in the oatmeal case, they just don't conflict like they did in the tea leaf case. So, conflicting doxastic contents is at least sometimes sufficient to reduce the justification for one's overall cognitive state.

One might suspect at this point that, while this does reduce the justification for one's overall cognitive state, it does so by reducing the justification for each of the beliefs that make up that state. Where S's belief that  $p$  and S's belief that  $r$  conflict, you might think that S is less justified in believing that  $p$  and that  $r$  than S would be in believing that  $p$  or that  $r$  because said conflict renders both S's belief that  $p$  and S's belief that  $r$  less justified than it otherwise would be. This, the defender of justification-asymmetry might

maintain, would be enough to support Pollock's interpretation of his case.

This does not seem right. First of all, it is at least possible that *S* is less justified in believing that *p* and that *r* but not less justified in believing that *p* or that *r*. This possibility has not been ruled out and it is not at all clear how a case could be constructed to do so. But, more importantly, there is positive reason to believe that *S* is less justified in believing that *p* and that *r* but not less justified in believing that *p* or that *r* when *p* and *r* conflict. Imagine that *S* believes that *p* and that *r*. Let's stipulate that *S*'s beliefs that *p* is justified and *S*'s belief that *r* is unjustified. *S* comes to realize that *p* and *r* conflict. Intuitively, this realization<sup>43</sup> has a negative impact on *S*'s epistemic state. But this impact does not bear on the epistemic status of *S*'s belief that *p* or *S*'s belief that *r*. Instead it bears on the epistemic status of *S*'s believing that *p* and that *r*. Imagine that *S* ceases to believe that *r*. *S* would then be in a better epistemic state. But it is still true that *S* realizes that *p* and *r* conflict. So, it must be that this realization makes *S* less justified in believing that *p* and that *r* but not less justified in believing that *p* or that *r*. The realization of conflict between one's beliefs is a reason to not hold both of those beliefs but it is not necessarily a reason to not hold one or the other of those beliefs.<sup>44</sup>

So, in order for Pollock's case to support justification-asymmetry, it must be established that the unjustified status of *S*'s belief that the next marble will be black, conjoined with the conflict between the contents of *S*'s beliefs, is not enough to explain our intuitions concerning the epistemic status of *S*'s cognitive state. There are two routes they might take to do this, one general and one specific.

---

43 Or the fact of conflict; it doesn't matter which at this point. We address this in section 2.1 below.

44 This issue is further discussed in section 1.3 below.

On the general route, the defenders of justification-asymmetry would argue that even when taking into account the effect of S's unjustified belief that the next marble will be black and the conflicting contents of S's belief on the justificatory status of S's cognitive state, our intuitions concerning the justificatory status of S's cognitive state as described by Pollock are not wholly explained. On the specific route, the defenders of justification-asymmetry would argue that, instead of the general intuition that S's overall cognitive state is epistemically bad, we have the specific intuition that S's belief that the next marble will be red would be unjustified.

Neither of these routes are particularly attractive. Past experience indicates that those reacting to Pollock's case are quite willing to identify S's cognitive state as epistemically bad. They are also willing, though somewhat more hesitantly, to endorse the claim that this epistemic status is not explained wholly by the justificatory status of S's belief that he has ESP and on that basis can tell that the next marble will be black. But they are positively unwilling to endorse the claim that this is not explained wholly by the justificatory status of S's belief that he has ESP and on that basis can tell that the next marble will be black conjoined with the epistemic effect of the mere conflict of S's belief contents.

They are likewise unwilling to endorse the claim that if S believes that he has ESP and on that basis can tell that the next marble will be black, then S is unjustified in believing that the next marble will be red once they are made aware of the distinction between this claim and the claim that S is unjustified in believing that the next marble will be red and believing that he has ESP and on that basis can tell that the next marble

will be black.

Respondents might be willing to endorse a claim like, "If S believes that the next marble will be black, even unjustifiably, then S shouldn't believe that the next marble will be red." However, the scope of "shouldn't" in that sentence is unclear. It might be read as ranging over just S's believing that the next marble will be red, or it might be read as ranging over the whole conditional. Pollock has given us no reason to interpret this claim one way rather than the other. So, intuitions about this case are not specific enough to provide much support for Pollock's claim.

If this were all that could be said in support of justification-asymmetry, we could take the reasons-against-belief theory to be in no danger. Given the weakness of these intuitions and the success of the theory, we should be comfortable giving up these apparent cases of efficacious defeat as merely apparent. Still, it remains a possibility that there are reasons beyond just our intuitions to favor the possibility of unjustified beliefs playing a defeating role. We'll turn to Bergmann's attempts to provide those reasons now.

## 1.2 *The Second Argument*

Like Pollock, Bergmann's defense of justification-asymmetry begins with a thought-experiment:

Suppose that, while looking down at your hands, you form the perfectly sensible belief that you have hands. But then suppose you irrationally become convinced that you are a brain-in-a-vat under the control of some enemy of yours (a disgruntled former colleague) who is trying to mislead you into thinking you have hands when you don't. (2006a, 164)

We take it that this case provides no more intuitive support for justification-asymmetry than Pollock's case did and poses no problem for the arguments we offered in section 1.

However, Bergmann is not interested in resting his defense on intuitions alone. He writes:

[What you need to do to see] that your hand belief loses its justification is to consider whether your belief that you have a hand is an appropriate response to the rest of your evidence, which includes not only your perceptual experience but also your belief that that perceptual experience can't be relied on to indicate the truth about whether you have hands. And it seems intuitively that the belief that you have hands is not an appropriate response to that combined evidence. Given all that, it seems that your belief that you have hands is an inappropriate response to the evidence you have. It is, therefore, unjustified in the circumstances. (2006a, 165)<sup>45</sup>

David Alexander (unpublished) has argued persuasively that Bergmann's argument here either begs the question by assuming that unjustified beliefs are part of one's evidence or fails because the claim that one's beliefs must be appropriate responses to one's evidence is false. Consider what Alexander calls the epistemic conception of evidence:

**Epistemic Conception:** One's evidence is simply all the information that one is rationally required to be responsive to in forming one's beliefs. (unpublished, 4)

On this conception it is true that one's belief must be an appropriate response to one's

---

45 It's worth noting here that Pollock makes some similar remarks, indicating that he might not be putting as much weight on our intuitions as it might have seemed:

One must proceed on the basis of whatever epistemic connections one sees – it is irrational to do anything else. (1974, 40)

However, the context in which this quote appears renders it ambiguous; he may only mean to have asserted that one must recognize one's defeaters as defeaters for them to defeat, a point we will take up in section 2.2 below, not that they need not be justified. So, we will let Bergmann be the champion for the second argument.

evidence in order to be justified. But Bergmann cannot assume without argument that one's evidence, understood in this way, includes the unjustified belief that one's perceptual experience can't be relied on to indicate the truth about one's hands; to do so would be to just assume that defeaters need not be justified in order to be efficacious.

Next, consider what an interpretation of Bergmann's claim that involves what Alexander calls the non-epistemic conception of evidence:

**Non-Epistemic Conception:** One's evidence is not necessarily the kind of thing that one's justified beliefs must be responsive to. (unpublished, 4)

On this conception, Bergmann's assumption that one's evidence includes the unjustified belief that one's perceptual experience can't be relied on to indicate the truth about one's hands does not beg the question. But now Bergmann cannot assume without argument that this belief has any effect on the justification one has for believing that you have hands; to do so would be, again, to just assume that which was to be shown.

### 1.3 *Defeat and Proper Function*

Alexander's objection is a conclusive refutation of the most obvious interpretations of Bergmann's argument. However, it may be that there is a less obvious interpretation that doesn't commit Bergmann to question-begging. Bergmann's account of justification relies heavily on the notion of a properly functioning cognitive system, such that were a belief not the result of a properly functioning cognitive system, it would not be justified. Perhaps the idea is that were one to believe that one's perceptual experience can't be relied on to indicate the truth about whether one has hands, and were one's cognitive system to be functioning properly, one wouldn't believe that one has hands. On

this interpretation, Bergmann is assuming the non-epistemic conception of evidence and taking his proper function account to provide support for the claim that your belief that you have hands must be an appropriate response to your belief that your perceptual experience can't be relied on to indicate the truth about your hands belief. In this way, he avoids begging the question.

But there is no good reason to think that a properly functioning cognitive system would cease believing that you have hands when you happen to believe that you're a brain in a vat. Given that your belief that you're a brain in a vat is unjustified, this belief must not be a response produced by your properly functioning cognitive system. So, your believing that you are a brain in a vat and not believing that you have hands would not be a response produced by a properly functioning cognitive system, either. The only response that would be produced by a properly functioning cognitive system is your believing that you have hands and ceasing to believe that you're a brain in a vat. If this is an option, then believing that you have hands is a response that would be produced by your cognitive system functioning properly and, so, is justified after all.

If this isn't an option, then we are left with three bad responses: (i) believing that you're a brain in a vat and believing that you have hands, (ii) believing that you're a brain in a vat and ceasing to believe that you have hands, and (iii) ceasing to believe that you're a brain in a vat and ceasing to believe that you have hands. Since we're concerned with the justificatory status of your belief that you have hands given that you believe that you're a brain in a vat, let's focus on the response in which you continue to believe that you have hands, (i).



While it is true that this response is not the result of a properly functioning system, it does not follow from this that continuing to believe that you have hands is not the result of a properly functioning system *given* that you believe that you are a brain in a vat. It is true that a cognitive system that produces the belief that you have hands and the belief that you are a brain in a vat is functioning worse than a cognitive system that produces just one or the other of those beliefs. But, as we saw in section 1 above, this sort of thing can be explained by reference to the conflict between the beliefs without assuming that the belief that you are a brain in a vat is a defeater with respect to your belief that you have hands.

If you're saddled with an unjustified belief that you cannot get rid of, the best thing to do might be to not allow that belief to affect your other beliefs. This is at least plausible, and so, while it allows Bergmann to avoid begging the question, so long as no additional defense is provided for the claim that the proper function account favors justification-asymmetry, Bergmann is not helped by adverting to the responses of properly functioning cognitive systems.<sup>46</sup>

#### 1.4 *In Support of Justification-Symmetry*

Neither of the arguments in favor of justification-asymmetry are compelling. This does not show that justification-asymmetry is false, however. In this section, we will first consider the so-called “arbitrariness objection,” which asserts that reasons must be given for deviating from justification-symmetry and that no such reasons have been given, so we should assume that justification-symmetry is true. We will then develop further a

---

<sup>46</sup> Thanks to Aaron Bronfman for this suggestion.

further argument in favor of justification-symmetry relying on the fact that the efficacy of a defeater co-varies with the degree of justification one has for that defeater.

According to Bergmann, the arbitrariness objection is as follows:

Clearly, unjustified beliefs can't confer any justification on other beliefs inferred from them, even if those inferred beliefs are properly inferred from those unjustified beliefs. Unjustified reasons, therefore, can't confer justification. But then something similar must be true of defeaters: unjustified defeaters can't remove justification. There is no good reason to treat reasons differently than defeaters in this regard. To say otherwise is arbitrarily to subject reasons to higher standards than defeaters. (2006a, 166)

Bergmann rejects this objection on the grounds that reasons to believe function by transmitting something to the beliefs they support, i.e., justification, while defeaters do not function by transmitting something to the beliefs with which they conflict. Thus, while it makes sense to think that a belief must possess something in order to transmit it, and so efficacious reasons to believe must be justified, efficacious defeaters do not transmit anything and so the considerations that support the justification-condition on efficacious reasons to believe do not support a justification-condition on efficacious defeaters.

In response to Bergmann, Alexander (unpublished, 6-7) argues that while reasons to believe and defeaters differ in the respect that Bergmann points out, they are similar in another important respect: they are both the sorts of things that should be taken into account when forming beliefs. Since not everything is the sort of thing that should be taken into account when forming beliefs, there must be epistemic credentials that

efficacious reasons to believe and defeaters possess that other things do not. Insofar as a reason to believe must be justified in order to be the sort of thing that should be taken into account when forming beliefs, it is reasonable to assume that, in the absence of arguments to the contrary, a defeater must also be justified in order to be the sort of thing that should be taken into account when forming beliefs.

While it may be true that there are epistemic credentials that reasons to believe and defeaters share, there are also epistemic credentials that reasons to believe must have that defeaters need not. For example, something is not a reason to believe that  $q$  unless it has content that somehow augers in favor of  $q$ . However, something can be a defeater for  $q$  without having content that augers in favor of  $q$ ; in fact, in standard cases of defeat, defeaters with respect to the belief that  $q$  have content that conflicts with  $q$ . Since reasons to believe and defeaters do not share all epistemic credentials, Alexander cannot simply assume that being justified is one credential shared between the two, so if there is a presumption in favor of justification-symmetry, it is not established by Alexander's argument.

Still, Bergmann's rejection of the arbitrariness objection is flawed. He asserts that a reason to believe justifies by transmitting justification, but this claim is problematic. In standard cases in which A transmits  $x$  to B are cases in which A possesses  $x$  and then imparts some of A's  $x$  on B, thus leaving A with less of  $x$ . But a justified belief that is a reason to believe that  $q$  loses none of its justification in justifying one's belief that  $q$ . Furthermore, as Bergmann has already asserted, a visual experience possesses no justification but can provide justification. So, it simply cannot be that anything that

provides justification must possess justification.

Perhaps the model of disease transmission would work well for Bergmann since A must usually have the disease in order to transmit it to B, A doesn't become less diseased by so transmitting, and in special cases one needn't be diseased in order to transmit the disease, e.g., Typhoid Mary.<sup>47</sup> But the point of all this is not that there is no model of transmission that can make Bergmann's claim that reasons to believe transmit justification plausible. The point is that just because we accept the transmission model of justification, we needn't accept the claim that this explains why a reason to believe must be justified and denying that defeaters work on a transmission model doesn't do anything to overturn the arbitrariness objection.

There are other reasons to reject justification-asymmetry, too. Consider the following cases:

**Case 1:** On the basis of the truth-table she just completed, a student, S, is justified to a certain degree in believing that a given proposition,  $q$ , is contingent. Soon, a fellow student tells S that there is a mistake in her truth-table. On the basis of this, S comes to believe that her truth-table contains a mistake.

**Case 2:** On the basis of the truth-table she just completed, a student, S, is justified to a certain degree in believing that a given proposition,  $q$ , is contingent. Soon, her professor tells S that there is a mistake in her truth-table. On the basis of this, S comes to believe that her truth-table contains a mistake.

It to be obvious that S's belief that her truth-table contains a mistake is better justified in

---

<sup>47</sup> It is actually implausible to claim that a carrier without symptoms is therefore not diseased, but it doesn't really matter here.

case 2 than it is in case 1, that S is less well justified in believing that  $q$  is contingent in case 2 than in case 1, and that S's belief that her truth-table contains a mistake is a defeater for her belief that the  $p$  is contingent.

These cases illustrate a covariance between how well justified a defeater is and the strength of that defeater. These cases are not anomalous; similar cases can be developed involving increased justification for defeaters over time, involving perceptual rather than testimonial evidence, and involving the defeat of *a posteriori* rather than *a priori* justified beliefs.

On the face of it, justification-symmetry provides an explanation for this phenomenon. The very thing that makes S's belief that her truth-table contains a mistake a stronger defeater in case 2 than in case 1 also makes that belief a stronger justifier, i.e., the degree of justification one has for a belief impacts efficacy reasons to believe and defeaters. As we argued in the previous chapter, the reasons-against-belief theory is particularly well-situated to explain this similarity since according to that theory, reasons to believe that  $q$  and defeaters with respect to the belief that  $q$  are mirror images of one another, the latter being reasons to not believe that  $q$  on some basis. Something is a reason to not believe that  $q$  on some basis because it is an epistemic reason to believe something, e.g., that not- $q$  or that one's basis for believing that  $q$  is unreliable. It is consonant with this view that as the justificatory efficacy of those reasons to believe increase so, too, would the efficacy of the reasons to not believe that they constitute.

One might object that it isn't obvious that the difference in how well justified S is in believing that  $q$  is contingent between case 1 and case 2 is due not to a difference in the

degree of justification S has for her belief that her truth-table contains a mistake, but rather to a difference between the two cases in the content of S's experiences. Perhaps S is less well justified in believing that  $q$  is contingent in case 2 than she is in case 1 because the experience of being told by her professor that her truth-table contains a mistake is a stronger defeater for her belief that  $q$  is contingent than is the experience of being told this by her fellow student. One might conclude that it is the content of a defeater, and not its justificatory efficacy, that is responsible for its efficacy as a defeater. So, one might conclude, we need not appeal to justification-symmetry to explain this covariance.

This much is true: were the cases above changed so that S never forms the belief that her truth-table contains a mistake, it would still be the case that she is better justified in believing that her truth-table contains a mistake in case 2 than she is in case 1 and that S is less well justified in believing that  $q$  is contingent in case 2 than in case 1. So-far, so-good for the objection. Notice, however, that this introduces a new covariance: the efficacy of S's experiences *qua* justifiers and the efficacy of S's experiences *qua* defeaters. The better S's experiences are at justifying beliefs that conflict with the belief that  $q$  is contingent, the better they are at being defeaters with respect to the belief that  $q$  is contingent. Justification-symmetry provides an explanation for this phenomenon, too. The very thing that makes S's experience in case 2 a stronger justifier than her experience in case 1, i.e., its content, also makes it a better defeater.

### 1.5 Coda

We've put justification-symmetry between reasons to believe and defeaters on

sound footing with respect to degree of efficacy. The very same thing that makes something a better justifier also makes it a better defeater. However, having established this much, we have not yet established justification-symmetry between justifiers and defeaters with respect to efficacy *simpliciter*. For all we've said so far, it might be that, although reasons to believe and defeaters increase and decrease in efficacy for the same reasons, a reason to believe cannot be efficacious at all without being justified, but a defeater can be.

However, given the co-variance in between how well justified a defeater is and how efficacious it is, the natural position to take would be that defeaters, like reasons to believe, must be justified to some degree in order to be efficacious to any degree. The plausibility of this might seem to depend on whether or not the principles of justification are weighted in favor of avoiding the false over believing the true. One might think that whenever one has a conflicting belief, no matter what its justificatory efficacy, there is some reason to doubt the proposition that belief conflicts with, even if the efficacy of a defeater is otherwise a factor of justificatory efficacy. That would seem to be the safer route if we're more concerned with avoiding the false than believing the true.

Despite appearances, a system of principles of belief formation and revision that puts more weight on avoiding false beliefs than it does on forming true beliefs would not necessarily endorse justification-asymmetry. This is so because defeaters don't just result in the withholding of judgment; they also remove roadblocks to beliefs one would otherwise hold. To see how, consider the following case:

A scientist, S, has significant evidence in favor of a proposition, *q*. However, S

has more evidence in favor of not- $q$ . Being a responsible scientist, S does not believe that  $q$ , despite her desire to do so. S then gets herself to believe, with no justification whatsoever, that all the evidence that has been provided for not- $q$  is misleading.

Note that the reason S is not justified in believing that  $q$  is that she has a defeater with respect to  $q$  in the form of reasons to believe that not- $q$ . If all defeaters are efficacious to some extent, regardless of whether or not they are justified at all, then S has managed to acquire an efficacious defeater with respect to her this defeater, i.e., S has acquired an efficacious defeater-defeater. Her defeater with respect to her belief that  $q$  defeated, S becomes justified in believing that  $q$ . So, if all defeaters are efficacious to some extent, defeaters could promote the formation of false beliefs because misleading defeaters can remove roadblocks to false beliefs. So, regardless of whether or not one thinks there is a bias in favor of avoiding the false over believing the true, one should find it plausible that an unjustified defeater is an inefficacious defeater.

## 2. *Other Asymmetries*

The foregoing addresses the most obvious potential asymmetry between reasons to believe and defeaters. However, having ruled that asymmetry out, we have not shown that there are no asymmetries between the two. There is a family of efficacy conditions that might be thought to apply differently between reasons to believe and defeaters. These conditions concern one's awareness of one's defeaters or reasons to believe. Here are three further possible asymmetries:<sup>48</sup>

---

<sup>48</sup> We will continue to limit the discussion to beliefs unless explicitly stated. We assume that these asymmetries could be re-written to include perceptual experiences and, although it would complicate



**Awareness-Asymmetry:** While (1a) S must believe/realize that  $p$  in order for the fact that  $p$  to be efficacious as a reason for S to believe that  $q$ , it is not the case that (1b) S must believe/realize that  $r$  in order for the fact that  $r$  to be efficacious as a defeater with respect to the belief that  $q$ .

**Meta-Awareness-Asymmetry:** While (2a) S need not believe/realize that  $p$  is a reason for S to believe that  $q$  in order for the belief that  $p$  to be efficacious as a reason for S to believe that  $q$ , it is not the case that (2b) S need not believe/realize that S's belief that  $r$  is a defeater with respect to S's belief that  $q$  in order for S's belief that  $r$  to be efficacious as a defeater with respect to S's belief that  $q$ .

**Basing-Asymmetry:** While (3a) S must base S's belief that  $q$  on S's belief that  $p$  in order for S's belief that  $p$  to be efficacious as a reason for S to believe that  $q$ , it is not the case that (3b) S must base S's failure to believe/ceasing to believe that  $q$  on S's belief that  $r$  in order for S's belief that  $r$  to be efficacious as a defeater with respect to the belief that  $q$ .

Awareness-asymmetry concerns whether or not one must be aware of a reason to believe or defeater *simpliciter* in order for that reason to believe or defeater to be efficacious. If only mental states can play the role of an efficacious reason to believe but some non-mental states can be efficacious defeaters, then there is awareness-asymmetry between reasons to believe and defeaters.<sup>49</sup>

Meta-awareness-asymmetry and basing-asymmetry, on the other hand, both concern the relationship between reasons to believe/defeaters and the beliefs with respect

---

the discussion, it would not raise any insurmountable challenges.

49 To be clear, we are not suggesting that one must be aware that one believes that  $p$  in order to believe that  $p$ . Rather, we are suggesting that one must be aware of something, perhaps  $p$ , in order to believe that  $p$ . In other words, the awareness is not of the belief; it is the belief.

to which they are reasons to believe/defeaters. Meta-awareness-asymmetry concerns whether or not being aware that one's reasons to believe/defeaters are reasons to believe/defeaters is a necessary condition on those reasons to believe/defeaters being efficacious. Basing-asymmetry concerns whether or not the relationship that efficacious reasons to believe must bear to the beliefs for which they are reasons to believe differs from the relationship that efficacious defeaters must bear to the beliefs with respect to which they are defeaters.

### 2.1 *Awareness-Asymmetry*

There is a distinction drawn in the defeat literature between psychological, or doxastic, defeaters and normative defeaters. Jennifer Lackey puts it this way:

A psychological defeater is a doubt or a belief that is had by S and that indicates that S's belief that *p* is either false or unreliably formed or sustained. Defeaters in this sense function by virtue of being had by S, regardless of their truth-value or epistemic status...A normative defeater is a doubt or a belief that S ought to have and that indicates that S's belief that *p* is either false or unreliably formed or sustained. Defeaters in this sense function in virtue of being doubts or beliefs that S *should have* (whether or not S does have them) given the presence of certain available evidence. (2008, 44-45)

A psychological defeater is a belief that need not be justified in order to be efficacious *qua* defeater. If we continue to assume that one's beliefs must be justified in order to be efficacious as reasons to believe, then the argument from section 1 rules out the possibility of such defeaters. Still, if the fact that S should believe something can be

an efficacious defeater with respect to some belief but such facts can never be reasons for S to believe something, then there is a significant asymmetry, an awareness-asymmetry, between reasons to believe and defeaters. So, what can be said in favor of the existence of normative defeaters?<sup>50</sup>

John Gibbons, in defense of what he calls access externalism, gives us a case that might be taken to make the existence of normative defeaters plausible:

The other morning, I went downstairs to make a mushroom, jalapeño, and cream cheese omelette. I had checked the night before to make sure we had all of the ingredients. Since Sunny rarely eats breakfast, it was reasonable for me to believe that the ingredients were still there. I went to the refrigerator and pulled out the eggs and mushrooms. While chopping, I firmly believed that I would soon have a mushroom, jalapeño, and cream cheese omelette. Unfortunately, in plain sight on the door of the refrigerator, there was a note. 'We're out of cream cheese.' I didn't notice the note, but I should have. After all, this is where we leave notes of this sort in our house. I thought I was having cream cheese for breakfast, but I should have known better. If I should have known that not-*p*, then I'm not justified in believing that *p*. (2006, 457)

We might be dubious of even the intuition that there is something epistemically wrong with John believing that he will eat cream cheese for breakfast. Even granting that if S knows that not-*p*, then S ought not believe that *p*, i.e., it ought to be the case that if S knows that not-*p* then S does not believe that *p*, and granting that S ought to know that

<sup>50</sup> Gilbert Harman's (1973, 143-144) assassination example is often cited as supporting the existence of normative defeaters. But Harman is discussing defeaters for knowledge, not justification, in that case. In fact, since it is supposed to be a Gettier-style case, Harman explicitly says that one does retain justification.

not- $p$ , it still doesn't follow that  $S$  ought not believe that  $p$ . This inference confuses that scope of the ought in the conditional between the wide-reading, i.e., [it ought to be the case that (if  $S$  knows that not- $p$ , then  $S$  does not believe that  $p$ )], and the narrow-reading, i.e., [if  $S$  ought to know that not- $p$ , then  $S$  ought not believe that  $p$ ]. Consider the following illustration:

Al wants to eat a delicious pasta dinner at 7:30 pm. In order to do this, he needs to start cooking the sauce at 6:00 pm and to start cooking the pasta at 7:15 pm.

However, Al gets distracted reading an engrossing and exciting dissertation and doesn't put the sauce on until 6:30 pm.

In this case, Al (practically) ought to start cooking the sauce at 6:00 pm and, if he does start cooking the sauce at 6:00 pm, he ought to start cooking the pasta at 7:15 pm. But that doesn't mean that he ought to start cooking the pasta at 7:15 pm, *simpliciter*. Given the way things actually go for Al, he ought not to start cooking the pasta at 7:15 pm because it will be cold and less than delicious by the time the sauce is ready. So, the wide-scope reading of the conditional is true, i.e., it is true that [it ought to be the case that (if Al starts the sauce at 6:00 pm, then Al starts the pasta at 7:15 pm)], but the narrow-scope reading of the conditional is false, i.e., it is false that [if Al ought to start the sauce at 6:00 pm, then Al ought to start the pasta at 7:15 pm]. There is no obvious reason to think that the move from the wide-scope reading to the narrow-scope reading in Gibbons' case is any more warranted than it is in the pasta case.

But even granting Gibbons that transition, his case still doesn't support awareness-asymmetry.<sup>51</sup> Gibbons' focuses on what he should have known, but we can safely assume

---

51 Note that Gibbons never says he is trying to support anything like awareness-asymmetry. For all he

that what is doing the work here, if any work is being done, is that he should have believed that he would not be eating cream cheese for breakfast. That John should have believed that he would not be eating cream cheese for breakfast either entails that John is justified in believing that he will not be eating cream cheese for breakfast or it doesn't.

Assume, first, that the entailment holds. On this horn of the dilemma, although Gibbons' case might support the existence of normative defeaters and the denial of (1b), it does not support awareness-asymmetry. Along with the denial of (1b), it relies on the denial of (1a), i.e., even though John does not believe that there is a note on the refrigerator indicating that there is no cream cheese, that fact is still efficacious as a reason for John to believe that he will not be eating cream cheese for breakfast. So, on this reading of Gibbons' case, there is no awareness-asymmetry and thus no challenge to the reasons-against-belief theory.

Assume, next, that the entailment doesn't hold. In this case, although John should have believed that he will not be eating cream cheese for breakfast, it is not the case that John is justified in believing that he will not be eating cream cheese for breakfast. It is important to clarify here just what Gibbons thinks is true. He asserts that:

**Justify:** If S should have believed that not- $p$ , then S is unjustified in believing that  $p$ .

This is distinct from the claim that:

**Should:** If S should have believed that not- $p$ , then S should not believe that  $p$ .

We contend that both *Should* and *Justify* explain the intuition that there is something

---

says in Gibbons, 2006, he may deny (1a) along with (1b). My concern here is the support he is trying to provide for the denial of (1b).

epistemically bad about John believing that he will eat cream cheese for breakfast, that *Should* doesn't entail *Justify*, that *Should* is more plausible than *Justify*, and that only *Justify* supports the existence of normative defeaters.

On the current interpretation, there's a difference between *Should* and *Justify* because there are two epistemic oughts at play. First, there's the epistemic ought at play where John should have believed that not-*p*; call this ought<sub>s</sub>. Second, there's the epistemic ought at play where John is unjustified in believing that *p*; call this ought<sub>j</sub>. So, here are *Should* and *Justify* rewritten:

**Justify\***: If S ought<sub>s</sub> to have believed that not-*p*, then S ought<sub>j</sub> not to have believed *p*.

**Should\***: If S ought<sub>s</sub> to have believed that not-*p*, then S ought<sub>s</sub> not to have believed *p*.

It is plausible that:

**Equal\***: If S ought to [not do something], then S ought not to [do that thing].

To illustrate this, consider:

**Equal\***: If I morally-ought to [not take the \$20 on your desk], then I morally-ought not to [take the \$20 on your desk].

But, while *Equal\** is plausible, perhaps even tautological, the following is not:

**Equal\*\***: If I morally-ought to [not take the \$20 on your desk], then I practically-ought not to [take the \$20 on your desk].

*Equal\*\** is far from obvious. If it is true, it's true because there is some theoretical bridge that takes us from the moral-ought to the practical-ought, a controversial claim. Similarly,

*Should\** is plausibly, perhaps even tautological, since it follows from *Equal*. Furthermore, assuming that  $\text{ought}_s$  is a genuine epistemic ought, then the truth of *Should\** would explain why we think there is something bad about John's epistemic position when he believes that he will eat cream cheese for breakfast; he  $\text{ought}_s$  not to believe this.

*Justify\**, on the other hand, if it were true, would only be true if there were some theoretical bridge that takes us from  $\text{ought}_s$  to  $\text{ought}_j$ . One candidate for that bridging claim would be something like:

**Bridge:**  $S \text{ ought}_s \text{ to believe that } p \text{ if and only if } S \text{ ought}_j \text{ to believe that } p$

The suggestion here is that there is some strong connection between  $\text{ought}_s$  and  $\text{ought}_j$ , perhaps identity. However, the defender of *Justify\** cannot avail themselves of *Bridge* on this interpretation since, by hypothesis, that John should have believed that not- $p$  does not entail that John is justified in believing that not- $p$ . Thus, on this interpretation, *Bridge* fails when read left to right. Instead, the defender of *Justify\** will have to make the following plausible:

**Bridge\*:**  $\text{If } S \text{ ought}_j \text{ to believe that } p, \text{ then } S \text{ ought}_s \text{ to believe that } p$

But *Bridge\** looks an awful lot like *Justify\**, so much so that it is not clear how one could support *Bridge\** without begging the question with respect to *Justify\**. Gibbons' case cannot be employed since, as demonstrated above, *Should\** provides a perfectly good explanation for our intuitions about that case. We suggest that unless one is willing to endorse *Bridge*, which would vitiate the support that Gibbons' case provides for awareness-asymmetry, there is no easy way to support *Justify\**.

So, it looks like *Should\** is a better candidate for an explanation of our intuitions

about Gibbons' case. But *Should\** doesn't support the existence of normative defeaters with respect to justification; instead, *Should\** supports the existence of normative defeaters with respect to epistemic-oughts, which falls outside the scope of this project. Even granting that Gibbons' case works, it remains true that on any plausible interpretation of Gibbons' case it either doesn't provide support for normative defeaters or else it supports the existence of normative defeaters but does not support asymmetry.

There's a more general reason to deny awareness-asymmetry, too. Imagine that (1a) is true and (1b) is false. This means that one can get into a case where one is neither justified in believing that *q* nor justified in not believing that *q*. Consider Gibbons' case again. Since we're assuming that (1b) is false, John would have an efficacious defeater with respect to his belief that he will eat cream cheese for breakfast, so he would not be justified in believing that he will eat cream cheese for breakfast. But since we're assuming that (1a) is true, John would not be justified in believing that he will not eat cream cheese for breakfast. So, John also wouldn't be justified in not believing that he will eat cream cheese for breakfast. The possibility of such an epistemic tragedy is controversial<sup>52</sup> but if one believes that epistemic tragedies are impossible then one should also believe that awareness-asymmetry is false.

One way to avoid epistemic tragedy but endorse awareness-asymmetry would be to allow that something can be an efficacious reason to not believe something without meeting the conditions necessary to be efficacious reasons to believe something. This way out either is theoretically unmotivated and thus unacceptably *ad hoc* or else is theoretically motivated and thus renders awareness-asymmetry unproblematic for the

---

52 See Jim Pryor (2015).



reasons-against-belief theory since it would simultaneously discharge whatever explanatory burden that asymmetry had placed on the reason-against-belief theory. If it's the former, then the epistemic tragedy re-emerges. If it's the latter, then we have no problem with it nor with awareness-asymmetry.

So, either there are no normative defeaters, their existence does not support awareness-asymmetry, or awareness-asymmetry is explained in way that renders it unproblematic for the reasons-against-belief theory. In any of these cases, awareness-asymmetry poses no threat to the reasons-against-belief theory of defeat.

## 2.2 *Meta-Awareness-Asymmetry*

Assume that one must be aware of one's reasons to believe/defeaters in order for them to be efficacious. Awareness of one's reasons to believe/defeaters *simpliciter*, however, is not the same as awareness of them *qua* reasons to believe/defeaters. It might be that, while one need not recognize one's reasons to believe *qua* reasons to believe in order for them to be efficacious, one must recognize one's defeaters *qua* defeaters in order for them to be efficacious. Let us call this a meta-awareness efficacy condition. The closest thing in the extant literature to support for this meta-awareness-asymmetry appears in remarks made by Pollock when discussing whether or not having reasons to believe a defeater is enough to reduce one's justification:

It might also seem that, if S has a good reason for believing some defeater but does not do so, this is sufficient to prevent the prima facie reason from justifying the belief that Q. But upon reflection, we think that this is wrong. Suppose R is a

defeater for P as a reason for believing that Q, and suppose S has a good reason for believing R but does not believe R. This can only be because he does not realize that he has a good reason for believing R. But then surely it would be irrational to withhold belief in Q on this basis [...] One must proceed on the basis of whatever epistemic connections one sees – it is irrational to do anything else. It only becomes unjustified once he realizes that he has good reason for believing a defeater, and that proposition (that he has good reason for believing a defeater) is itself a defeater. (1974, 44)

Pollock's claim that S's belief that Q only becomes unjustified once S realizes that he has good reason for believing a defeater is ambiguous. On one reading, the *de dicto* reading, it supports basing-asymmetry, but on the other, the *de re* reading, it does not.

Read *de dicto*, Pollock is asserting that, in order for the defeater R to be efficacious, one must believe that one has good reason to believe a defeater, which happens to be R. Read *de re*, Pollock is merely asserting that, in order for the defeater R to be efficacious, one must believe that one has good reason to believe R, which happens to be a defeater. In other words, read *de dicto*, Pollock is asserting that a defeater is not efficacious unless it is recognized as a defeater while, read *de re*, Pollock is asserting only that a defeater is not efficacious unless it is believed *simpliciter*.

Read *de re*, Pollock's claim does not support the denial of (2b). In fact, if anything, it supports the truth of (1b) since it is asserting that only reasons one is aware of can be efficacious in removing justification. So, if there is support for the denial of (2b) here, Pollock's assertion must be read *de dicto*: one must recognize one's reasons as

reasons to believe a defeater *qua* defeater, not just as reasons to believe something, which happens to be a defeater.

In support of the *de dicto* interpretation, note that Pollock asserts that, “One must proceed on the basis of whatever epistemic connections one sees – it is irrational to do anything else.” It is clear from the context that, although Pollock explicitly refers to irrationality, what Pollock means is that one is justified only in proceeding on the basis of whatever epistemic connections one sees. “Proceeding” here just means believing or ceasing to believe. So, to put it into the terminology we've been using, Pollock is asserting that S's belief that *r* is an efficacious reason for S to cease to believe that *q* only if S believes that S's belief that *r* is a defeater with respect to S's belief that *q*. But then, by the arbitrariness objection considered in section 1.4 above, it would seem that S's belief that *p* would justify S in believing that *q* only if S believes that S's belief that *p* is a reason for S to believe that *q*. Pollock writes:

I may claim that my dog is ill. After staring at my dog in puzzlement for a while, I may suddenly notice that he has a glazed look in his eye and it will come to me that that is what led me to think he was ill. But on a conscious level I was previously unaware of the glazed look. In other words, my reason for thinking he is ill is that he has a glazed look in his eye, but as I did not consciously note the latter, I did not have the belief that he has a glazed look in his eye. (1974, 60)<sup>53</sup>

Pollock says that in this case he is justified in believing that his dog is sick. But, although he is aware of the glazed look in his dog's eye, he does not realize that he is aware of this, much less realize that this is a reason for him to think that his dog is sick. Still, Pollock

---

53 Pollock credits this case to Anthony Quinton (1955).

maintains, his awareness of the glazed look in his dog's eye justifies him in believing that his dog is sick.

In this case, however, he would be proceeding on the basis of epistemic connections he doesn't see. If the claim that one is justified in proceeding on the basis of only those epistemic connections that one sees is the principle that is supposed to support the denial of (2b), then, absent an argument to the contrary, it would also support the denial of (2a) and, thus, the denial of Pollock's conclusion in the case above. The reasons-against-belief theory is not committed to the denial of both (2a) and (2b) nor to the affirmation of both (2a) and (2b), but it is committed to the disjunction [(2a) & (2b)] or [not-(2a) & not-(2b)]. Nothing Pollock says poses a significant challenge to this disjunction.

### 2.3 *Basing-Asymmetry*

It is generally maintained that S's belief that *p* is not an efficacious reason to believe that *q* unless S believes that *q* on the basis of S's belief that *p*. It might be thought that the same cannot hold for defeaters since they don't produce belief. After all, there is no relevant belief to be based on anything. But it is not obvious that there isn't a relation analogous to the basing relation at play in cases of defeat. Consider the following pair of cases:

**Paulo:** Paulo believes that his most recent philosophy paper will pass, and on the basis of good evidence, too. However, after turning the paper in he hears on good authority that his philosophy instructor dislikes him and will fail his paper regardless. Because of this information, Paulo stops believing that his philosophy

paper will pass.

**Felix:** Felix believes that his most recent philosophy paper will pass, and on the basis of good evidence, too. However, after turning the paper in he hears on good authority that his philosophy instructor dislikes him and will fail his paper regardless. Like Paulo, Felix stops believing that his philosophy paper will pass but, unlike Paulo, Felix does not do so because of the information he received; instead, he does so because he suffers from imposter syndrome and the information that his instructor doesn't like him sends him into a spiral of self-abuse, ending in the unjustified belief that he has no philosophical talent. On the basis of that unjustified belief, Felix ceases to believe that his paper will pass.

It seems that Paulo is in a better epistemic position than Felix with respect to ceasing to believe that the paper will pass. Here's one way of explaining this: having a defeater with respect to  $S$ 's belief that  $q$  promotes the propositional justification one has to cease to believe that  $q$ , but if one ceases to believe that  $q$  for some reason other than the defeater, then the defeater is unable to promote the doxastic justification for that belief. So, both Paulo and Felix are propositionally justified in ceasing to believe that their papers will pass but only Paulo is doxastically justified in ceasing to believe that his paper will pass. If this were right, then there is a basing-relation active in cases of defeat, too.

The idea here is that the basing relation holds between one's reasons to believe and one's forming or maintaining a belief, not between one's reasons to believe and the belief state itself. Similarly, the analogous relation for defeaters holds between one's defeater and one's ceasing to believe or continuing to not believe, not between one's

defeaters and the belief state itself.

This interpretation of the cases is, obviously, convivial to the reasons-against-belief theory of defeat and so to simply assert that it is the right way to think about these cases would be question-begging. But this interpretation of the cases does raise the challenge for the defender of basing-asymmetry to explain the difference in epistemic status between Paulo and Felix. Of course, the intuitions about Felix's epistemic situation are open to many of the same criticisms that were raised above in section 1.1 with respect to Pollock's marble case. That should not trouble us because the interpretation we've given of these cases still demonstrates that there is a relation analogous to the basing relation that can plausibly be posited to hold for defeaters.

However, the issue of basing-asymmetry still raises concerns about the reasons-against-belief theory of defeat. Recall from chapter 3 that in order to establish the attractively simple picture of defeat given by the reasons-against-belief theory, it was necessary to posit that reasons to believe are self-promoting, i.e., S's belief that  $p$  is never a reason for S to believe that  $q$  *simpliciter* but rather a reason for S to believe that  $q$  on the basis of S's belief that  $p$ . This was necessary because the reasons-against-belief theory asserts that S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if S's belief that  $r$  is a reason for S to not believe that  $q$  on some basis. Defeat, then, could be characterized as a conflict between reasons for S to believe that  $q$  on the basis of S's belief that  $p$  and reasons for S to not believe that  $q$  on the basis of S's belief that  $p$ .

If basing-asymmetry is false because (3b) is true, i.e., because defeaters can stand in a basing relation to the cessation of belief in the same way that reasons to believe can

stand in a basing relation to the formation of belief, then the picture above is muddled. To maintain symmetry, it cannot be that S's belief that  $r$  is a defeater with respect to S's belief that  $q$  if and only if it is a reason for S to not believe that [( $q$  on some basis) *simpliciter*]. Instead, it must be that S's belief that  $r$  is a reason for S to [(not believe that  $q$  on some basis) on the basis of S's belief that  $r$ ].

Recall the worry from chapter 3 that led to the defense of the claim that reasons to believe are self-promoting:

There seems to be a mismatch between S's reasons for and against believing that  $q$ . S's belief that  $p$  seems to be a reason for S to believe that  $q$  *simpliciter*, but S's belief that  $r$  might well be a reason for S to not believe that  $q$ , not *simpliciter*, but rather only on the basis of S's belief that  $p$ . So, this does not seem to be a simple case of reasons for S to do something, i.e., believe that  $q$ , versus reasons for S to not do it. Obviously, there is no straight-forward function from S's reasons to take some action,  $\phi$ , and S's reasons to not take some other action,  $\psi$ , to whether or not S should  $\phi$ .

As it turns out, if defeaters are self-promoting, this does not spoil the simple competition-between-reasons-for-and-against-belief picture of defeat. The worry from the previous chapter was that believing that  $q$  and believing that  $q$  on the basis of S's belief that  $p$  are not obviously the same thing. As a result, having reason to not do one of them doesn't immediately tell us anything about our reasons for or against doing the other.

If defeaters are self-promoting, no similar problem arises. This is because of the scope of the negation involved in reasons against belief. It is not that S's belief that  $r$  is a

reason for S to not [(believe that  $q$  on the basis of S's belief that  $p$ ) on the basis of S's belief that  $r$ ]. That would indeed be problematic since not doing that thing is consistent with believing that  $q$  on the basis of S's belief that  $p$ . If S believes that  $q$  on the basis of S's belief that  $p$  and ignores S's belief that  $r$ , then S has not [(believed that  $q$  on the basis of S's belief that  $p$ ) on the basis of S's belief that  $r$ ]; rather, S has simply (believed that  $q$  on the basis of S's belief that  $p$ ).

Instead, S's belief that  $r$  is a reason for S to [(not believe that  $q$  on the basis of S's belief that  $p$ ) on the basis of S's belief that  $r$ ], and doing that is not consistent with S's believing that  $q$  on the basis of S's belief that  $p$ . So, the reason-against-belief theory of defeat remains consistent with the simple conflict-of-reasons-for-and-against-belief picture of defeat.

### 3. *Conclusion*

Principled arguments in favor of asymmetry are scarce. Without a principled argument, defenders of asymmetry are left with only intuitions about cases. But we've argued that for several kinds of asymmetry, the intuitions are simply not strong enough to support asymmetry. Furthermore, the principles that are cited in support of asymmetry actually only support a claim about the conditions on the efficacy for defeaters. Without an argument to the effect that these principles don't support the same claim about the conditions on the efficacy for reasons to believe, they do not actually support asymmetry.

So, there is no extant reason to accept asymmetry and some reason to deny it, even independent of the success of the reasons-against-belief theory. However, the



success of that theory does provide significant reason to accept symmetry. Given the robust success of the reasons-against-belief theory at handling various puzzles concerning defeaters and explaining otherwise unexplained phenomena, there is fairly strong reason to reject any hypotheses that do not fit well with that theory. Such hypotheses would have to be strongly supported and, in the absence of further arguments, this is simply not the case for the asymmetries considered above.

Up to this point the case for an exploration of the fundamental nature of defeat has been made (chapter 1) and existing theories of defeat have been rejected (chapter 2). A new theory of defeat, the reason-against-belief theory, was then developed (chapter 3) and, now, defended against significant objections. We take the reason-against-belief theory to be plausible and so, in the remaining chapter, we will explore the various applications to which the reason-against-belief can be put.

# CHAPTER 5

## APPLICATIONS & IMPLICATIONS

In the preceding chapters we've seen the development and defense of the reasons-against-belief theory of defeat, according to which:

**Reasons-Against-Belief Theory of Defeat:** *S*'s belief that *r* would be a defeater with respect to *S*'s belief that *q* if and only if *S*'s belief that *r* would be a reason for *S* to not believe that *q* on some basis, *E*.

It was argued in chapter 1 that this project is important because the concept of defeat is employed in a number of debates in epistemology but that without a clear and carefully developed theory of defeat the applications are bound to be fruitless at best and misleading at worst. In particular, it was argued that although it is widely maintained that externalism about justification faces a difficulty in accommodating defeaters, that conclusion is unwarranted prior to a theory of defeat. It was argued further that as the concept of defeat is applied in the literature on higher-order evidence, it is misapplied there, too.

In this closing chapter we will see the application to these controversies of the lessons learned in developing the reasons-against-belief theory of defeat. The implications of the reasons-against-belief theory of defeat turn out to be fairly modest. This is a virtue of the theory; it would be surprising, and not a little suspicious, if a theory of defeat settled by itself long standing epistemological controversies.

## 1. *Externalism & Defeat*

We concluded in chapter 1 that externalist theories of justification do not obviously face a problem accommodating defeat merely because they are externalist. The apparent difficulty that externalism faces in accommodating defeat is due not to the fact that the theory is externalist, *per se*, but rather to the bifurcation problem:

**Bifurcation Problem:** If a theory of justification (i) motivates different conditions on justification in different ways and (ii) provides no explanation for this difference, then (iii) this theory of justification is explanatorily inadequate.

The threat of the bifurcation problem arises for externalism because, as we argued in chapter 1, externalist theories must place both a positive condition on justification and a negative condition on justification in order to accommodate defeat. Recall, for example, the simple externalist theory:

**Appended Simple Externalism:** S is justified in believing that *q* if and only if S's belief that *q* is the result of a reliable process and S has no defeaters with respect to the belief that *q*.

This theory places two conditions on justification. First, in order to be justified in believing that *q*, S's belief that *q* must result from a reliable process; call this the positive condition on justification. Second, in order to be justified in believing that *q*, S must have no defeaters with respect to the belief that *q*; call this the negative condition on justification. Without the negative condition, the theory would be extensionally inadequate; it would erroneously classify as justified S's believing that *q* on the basis of a reliable process even when S possesses a defeater with respect to that belief.

We should not limit the current discussion to our simple externalist theory from above. Instead, let us refer to whatever conditions on justification any externalist theory puts in place to accommodate defeat as negative conditions and whatever other conditions on justification an externalist theory puts in place as positive conditions.<sup>54</sup> Externalism writ large, which we will henceforth simply refer to as externalism, incurs an explanatory burden via the bifurcation problem only if the underlying motivation for the positive conditions necessarily differs from the underlying motivation for the negative condition; only then will externalism satisfy condition (i) of the bifurcation problem.<sup>55</sup>

Whether or not externalist theories of justification that accommodate defeat necessarily satisfy condition (i) of the bifurcation problem depends on both the underlying motivation for the positive condition and the underlying motivation for the negative condition. We've seen that, in the absence of a thoroughgoing theory of defeat, it is impossible to determine the underlying motivation for the negative condition and, thus, impossible to evaluate whether or not externalism incurs an explanatory burden.

In the intervening chapters, we've seen the development of just such a theory, the reasons-against-belief theory of defeat, and it might be thought that under this theory defeaters have an internalist character. The reasons-against-belief theory of defeat proceeds in terms of epistemic reasons, both for and against belief, and epistemic reasons are thought by some to be fundamentally internalist in nature. For example, Laurence Bonjour writes of an externalist theory:

---

54 This might not be the best way to make this division since a theory certainly could have negative conditions other than no-defeaters conditions. We will ignore that possibility in the current work.

55 Of course, specific externalist theories might run afoul of the bifurcation problem even if externalism writ large need not. We are here concerned with externalism writ large.

What makes such a view externalist is the absence of any requirement that the person for whom the belief is justified have any sort of cognitive access to the relation of reliability in question. Lacking such access, such a person will in general have no reason for thinking that the belief is true or likely to be true, but will, on such an account, none the less be epistemically justified in accepting it. Thus such a view arguably marks a major break from the modern epistemological tradition, stemming from Descartes, which identifies epistemic justification with having a reason, perhaps even a conclusive reason, for thinking that the belief is true. (1992, 133)

The suggestion here is that externalists are committed to the claim that one can satisfy the positive conditions on justification with respect to believing that  $q$  without having any reason to believe that  $q$ . If that is correct, and if defeaters just are reason to not believe, then there is a disconnect between the positive and negative conditions that externalism places on justification. Externalists would be committed to the puzzling claim that, although you need not have any reasons for belief in order to be justified in believing, you must lack reasons against belief. Absent an explanation for why the negative conditions proceed in terms of reasons but the positive conditions do not, externalism is rendered explanatorily inadequate.

The internal character of epistemic reasons, the passage above suggests, is supposed to be that one has an epistemic reason to believe that  $q$  only if one has cognitive access to the epistemic relation between one's basis for believing that  $q$  and one's belief that  $q$ . We can call this a meta-awareness condition on having reasons. The characteristic

claim of externalism is that there is no meta-awareness condition on justification. Thus, BonJour concludes, externalists are committed to the claim that one can be justified in believing without having epistemic reason to believe.

Either there is a meta-awareness condition on having epistemic reasons or there is not. If there is not, then externalists are not committed to the claim that one can be justified in believing without having epistemic reason to believe. It is open to an externalist to maintain that satisfying the positive conditions on justification gives one epistemic reason to believe. For example, it might be maintained that if one has a reliable belief forming process that indicates that  $q$ , then one has reason to believe that  $q$ . Thus, if there is no meta-awareness condition on having epistemic reasons, there is no immediate threat of externalism running afoul of the bifurcation problem.

If, on the other hand, there is a meta-awareness condition on having epistemic reasons, then externalists are committed to the claim that one can be justified in believing without having epistemic reason to believe. However, this would only be problematic if the reasons-against-belief theory of defeat is necessarily committed to defeat proceeding specifically in terms of epistemic reasons, but it is not. What is essential to the theory is that defeaters are entities that function by supporting one's not believing. The reasons-against-belief theory could just as easily have been called the factors-against-belief theory. Nothing hangs on the use of the term "epistemic reason," so the nature of epistemic reasons is actually a red herring in this debate. If BonJour is right that there is a meta-awareness condition on having epistemic reasons, this does not result in externalists incurring an explanatory burden. The externalist can do without epistemic reasons.

What would need to be established to show that externalism incurs an explanatory burden via the bifurcation problem is that there is a meta-awareness condition on having a defeater. Since this is where the real action is, and not with the nature of epistemic reasons, there is no harm in shifting back to our familiar terminology of epistemic reasons. On the face of it, the cases that impel externalists to include a negative condition on justification do suggest that there is a meta-awareness condition on having a defeater. Consider a classic case developed by Bonjour:

Maud believes herself to have the power of clairvoyance, though she has no reasons for this belief. She maintains her belief despite being inundated by her embarrassed friends and relatives with massive quantities of apparently cogent scientific evidence that no such power is possible. One day Maud comes to believe, for no apparent reason, that the President is in New York City, and she maintains this belief despite the lack of any independent evidence, appealing to her alleged clairvoyant power. Moreover, her belief about the President did result from the operation of that power. (1985, 40)

In this case, the evidence that Maud has that clairvoyance is impossible is supposed to be a defeater with respect to Maud's belief that the President is in New York. Under the reasons-against-belief theory of defeat, this would be because that evidence is reason for Maud to not believe on the basis of her clairvoyance that the President is in New York. A contrast is drawn between Maud's position with respect to the epistemic relation between her clairvoyance and her believing that the President is in New York and Maud's position with respect to the epistemic relation between the evidence provided by her

acquaintances and her believing that the President is in New York. In the latter case, Maud has cognitive access to the epistemic relation between her evidence and her believing, i.e., she has cognitive access to the fact that her evidence gives her reason to not believe. In the former case, Maud does not have cognitive access to the epistemic relation between her clairvoyance and the formation of belief, i.e., she does not have cognitive access to the fact that her clairvoyance gives her reason to believe.

One might be led by this case into thinking that there is a meta-awareness condition on having a defeater, one that cannot be accommodated by the considerations that motivate the externalist's positive conditions on justification. However, one would be mistaken in so thinking. The case above only shows that one can have a defeater such that one has cognitive access to the epistemic relation between that defeater and one's believing; it has not shown that one must have such access. Consider this modification of Bonjour's case:

Maud believes herself to have the power of clairvoyance, though she has no reasons for this belief. Maud also believes that the patterns in her tea leaves are evidence for distant events, though she has no reason for this belief. One day Maud comes to believe, for no apparent reason, that the President is in New York City, and she maintains this belief despite the lack of any independent evidence, appealing to her alleged clairvoyant power. Moreover, her belief about the President did result from the operation of that power. However, Maud also sees a pattern in her tea leaves that she takes to indicate that the President is not in New York. Moreover, reading the tea leaves really is a reliable way to form beliefs and



the pattern in Maud's tea leaves really does support not believing that the President is in New York.

In this case, Maud has no more cognitive access to the epistemic relation between the configuration of her tea leaves and her believing that the President is in New York than she does to the epistemic relation between her clairvoyance and her believing that the President is in New York. If Maud's perception of the pattern in her tea leaves<sup>56</sup> is a defeater with respect to Maud's belief that the President is in New York, then there is no meta-awareness condition on having a defeater and externalism does not incur an explanatory burden via the bifurcation problem.

We can certainly imagine philosophers insisting at this point that Maud's perception of the pattern in Maud's tea leaves really does not constitute a defeater. But it's one thing to insist on this and another thing entirely to support it. Note that we are taking no position one way or the other about the possibility of providing such support. Rather, the thought is that the prospects are dim for providing such support that is independent of an argument for the inclusion of a meta-awareness condition on satisfying the positive conditions on justification. We saw in chapter 4 that there are no compelling reasons to think that there is an asymmetry in meta-awareness conditions between justifiers and defeaters. Thus, if it is granted to the externalist that there is no problem with their positive conditions on justification independent of the bifurcation problem, then bifurcation poses no problem, too.

To put the point another way, imagine Bonjour's objection proceeding in this way:

---

<sup>56</sup> Or her belief about the pattern in her tea leaves or, if you don't have a problem with epistemic reasons being things other than mental states, the pattern of the tea leaves itself.

(P1) If there is a meta-awareness condition on having a reason to not believe, then there is a meta-awareness condition on having a reason to believe.

(P2) There is a meta-awareness condition on having a reason to not believe.

(C1) So, there is a meta-awareness condition on having a reason to believe.

Since no support is provided for (P2), it is open to the externalist to accept the first premise, but to then turn Bonjour's *modus ponens* into a *modus tollens*:

(P1) If there is a meta-awareness condition on having a reason to not believe, then there is a meta-awareness condition on having a reason to believe.

(P2\*) There is no meta-awareness condition on having a reason to believe.

(C1\*) So, there is no meta-awareness condition on having a reason to not believe.

My suggestion is that if (P2) is true, then the problem facing externalism is not bifurcation of motivations. The externalist can maintain a single motivation for both the positive and negative conditions. The problem facing externalists would be that they would be wrong about whether or not there is a meta-awareness condition on reasons, for or against, belief. Accommodating defeat does not put any explanatory burden on externalism over and above the explanatory burden it already has in virtue of its positive conditions.

It is now clear that accommodating defeat under the reason-against-belief theory of defeat does not pose any special problem for externalism. Any problems externalism faces do not come from that quarter. In fact, the reason-against-belief theory provides externalism with a tool to tackle another of Bonjour's problem cases. Consider the case of

Norman:

Norman, under certain conditions which usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power or for or against the thesis that he possesses it. One day Norman comes to believe that the President is in New York City, though he has no evidence either for or against this belief. In fact the belief is true and results from his clairvoyant power under circumstances in which it is completely reliable.

(1985, 41)

Intuitions are split about this case, but Bonjour thinks it is clear that Norman is not justified in believing that the President is in New York. However, the negative conditions that externalism places on justification cannot explain this, it is supposed, because Norman has no defeater with respect to the belief that the President is in New York. So, it is concluded, externalism is extensionally inadequate.

For externalists troubled by this case the reasons-against-belief theory of defeat provides a possible explanation for Norman's lack of justification that is consistent with externalism. Recall from chapter 3 the discussion of being justified in not believing in the apparent absence of reasons for or against belief. It was suggested there that an absence of reason to believe that  $q$  is itself a reason to not believe that  $q$ . By extension, a reason to believe that there is an absence of reason to believe that  $q$  would also be a reason to not believe that  $q$ . Although Norman does have reason to believe that the President is in New York, Norman might plausibly be seen to have reason to believe that he has no reason to

believe that the President is in New York and, thus, might plausibly be seen to have reason to not believe that the President is in New York. This suggestion is tentative, but then so too are the intuitions that the case elicits.

The result of applying the reasons-against-belief theory of defeat to the internalism/externalism debate is only negative. The nature of defeat does not settle that debate. It does place constraints on the construction of theories of justification, e.g., Pollock's theory of justification is problematic because it treats reasons for belief and reasons against belief asymmetrically, but it does not raise any special problems for externalist or internalist theories of justification merely in virtue of their being externalist or internalist.

## 2. *Higher-Order Evidence*

In chapter 1 we saw that the relationship between higher-order evidence and defeaters cannot be determined in the absence of a thorough-going theory of defeat. Now that we have such a theory before us, we can see whether and how defeaters and higher-order evidence might differ from one another.

Before proceeding, we should note an important variation in types of higher-order evidence. Some higher-order evidence concerns one's ability to properly form beliefs on the basis of one's evidence but does not challenge the value of the evidence itself.

Consider, for example:

**Questionable Judgment:** You are a prosecutor assigned to a local robbery case.

The detectives have been investigating four main suspects: Smith, Jones, Parker,

and Brown. You have now set out to see whether there is a plausible case to be brought against any of them and you've obtained a copy of a file containing the evidence that the detectives have gathered about each of the suspects as it relates to the crime. You open it, and after reading for several minutes, come to see that there is an obvious and extremely compelling case to be made that Jones is guilty. There is nothing tricky about the case and you can tell that any sincere, rational adult should see that Jones is very likely the perpetrator of the crime. But just then you get a telephone call from someone at the police department. There is a problem. A crooked detective is looking to frame someone that he knows is innocent. He somehow procured a quantity of the illegal reasoning-distorting drug and, through some of his own experimentation, figured out that it will cause those who take it to judge one of the innocent suspects in the robbery case as guilty. He tried to drug you by mixing the substance into the office coffee that you drank this morning. But he didn't realize that there were two pots of coffee and he only put the drug into one of them. Although you did not actually drink the laced coffee, no one, yourself included, yet knows this.<sup>57</sup>

In this case, the information regarding the drugged pot of coffee is evidence that you are incompetent with respect to forming beliefs on the basis of criminal evidence. Notice that it does not call into question your evidence; it only calls into question your ability to process your evidence. This higher-order evidence is self-oriented.

Contrast this with higher-order evidence that does not challenge one's ability to properly form beliefs on the basis of one's evidence but does challenge the value of the

---

<sup>57</sup> This and the case that follows are due to Andrew Rotondo (2013, 565-566).

evidence itself. Consider, for example:

**Questionable Evidence:** You are a prosecutor assigned to a local robbery case. The detectives have been investigating four main suspects: Smith, Jones, Parker, and Brown. You have now set out to see whether there is a plausible case to be brought against any of them and you've obtained a copy of a file containing the evidence that the detectives have gathered about each of the suspects as it relates to the crime. You open it, and after reading for several minutes, come to see that there is an obvious and extremely compelling case to be made that Jones is guilty. There is nothing tricky about the case and you can tell that any sincere, rational adult should see that Jones is very likely the perpetrator of the crime. But just then you get a telephone call from someone at the police department. There is a problem with the investigation into the robbery. Two files about the robbery have been discovered. One contains the actual evidence that was collected. The other was fabricated by a crooked detective looking to frame someone that he knew to be innocent. This detective unsuccessfully tried to swap the fake file for the actual one. The result is that there are now two files on the crime - the real one and the fake one – and, although you actually have the real file, no one, yourself included, knows this.

Notice that the evidence you receive in the telephone call does not call into question your ability to process your evidence; it only calls into question the quality of your evidence. Although this is a kind of higher-order evidence in that it does not bear directly on your belief about who committed the crime, it is not self-oriented but rather evidence-oriented.

Furthermore, this evidence-oriented higher-order evidence is a standard case of source-specific defeat. Since we are exploring the similarities and differences between standard cases of defeat and higher-order evidence, it will be helpful to reserve the term “higher-order evidence” for instances of self-oriented higher-order evidence.

Although *Questionable Judgment* and *Questionable Evidence* differ in interesting respects, the reasons-against-belief theory of defeat suggests an underlying commonality. The telephone call in both cases gives you reason to not believe that Jones is guilty on the basis of the contents of the file and, by so doing, reduces your justification for believing that Jones is guilty. In order for higher-order evidence to differ fundamentally from standard cases of defeat, higher-order evidence must not function in virtue of being a reason to not believe on the basis of your evidence.

We shall proceed by laying out the various differences that have been suggested to hold between standard defeaters and higher-order evidence. We will then sort the merely apparent differences from the actual differences. Finally, we will evaluate whether or not the actual differences support the claim that standard defeaters and higher-order evidence are fundamentally different. We will see that they do not.

### 2.1 *Supposed Differences*

Higher-order evidence is taken to differ from standard defeaters in four primary ways: its justification-stripping force is agent-relative, its justification-stripping force is temporally anomalous, arguments dismissing higher-order evidence are often circular in a peculiar way, and higher-order evidence does not nullify the evidential power of one's evidence.

First, higher-order evidence is supposed to be agent-relative in a way that standard defeaters are not. In *Questionable Evidence*, were another investigator to be looking at the same evidence as you, the telephone call you receive would challenge her justification for believing that Jones is guilty just the same as it does for you. However, in *Questionable Judgment*, the analogous claim would not be true. The other investigator, upon being told that you might be affected by a drug, would not lose justification for believing that Jones is guilty.<sup>58</sup> David Christensen (2010) takes these to be general features of higher-order evidence and source-specific defeaters.<sup>59</sup>

Second, higher-order evidence is supposed to be sensitive to time in a way that differs from standard cases of defeat. Consider the following cases:

**Standard Science:** S is a scientist investigating some phenomenon experimentally. Were S to get evidence, E, from the experiments, S would be justified in believing that *q*. What's more, S is justified in believing that were S to get E, S would be justified in believing that *q*. S is going into the lab tomorrow and will find out whether or not E obtains. So, tonight, S is justified in believing that if S sees that E tomorrow, it is likely that *q*, and S is also justified in believing that if S sees that E tomorrow, then S will be justified in believing that *q*.

However, later at night, S receives a call from a reliable source informing S that here will be a power-outage tomorrow morning that will reset the experiment in such a way that E will no longer reliably indicate that *q*.

The intuitive effects of this phone call are as follows: (i) S is now unjustified in believing

---

58 Unless, of course, she drank the coffee, too.

59 Though it should be pointed out that he says that higher-order evidence often has this property, not that it always does



that [if S sees E tomorrow, then it is likely that  $q$ ], (ii) S is now unjustified in believing that [if S sees E tomorrow, then S will be justified in believing that  $q$ ], and (iii) S will tomorrow be unjustified in believing that  $q$  on the basis of E. Contrast this case of standard defeat with a case of higher-order evidence:

**Higher-Order Science:** S is a scientist investigating some phenomenon experimentally. Were S to get evidence, E, from the experiments, S would be justified in believing that  $q$ . What's more, S is justified in believing that were S to get E, S would be justified in believing that  $q$ . S is going into the lab tomorrow and will find out whether or not E obtains. So, tonight, S is justified in believing that if S sees that E tomorrow, it is likely that  $q$ , and S is also justified in believing that if S sees that E tomorrow, then S will be justified in believing that  $q$ .

However, later at night, S receives a call from a reliable source informing S that she will be slipped a drug in her morning coffee that will render her incompetent with respect to forming beliefs on the basis of evidence like E.

The intuitive effects of this call are the same as the intuitive effects of the call in *Standard Science* with respect to (ii) and (iii), but different with respect to (i). In other words, just like in the case of *Standard Science*, the call in *Higher-Order Science* makes it the case that (ii) S is now unjustified in believing that [if S sees E tomorrow, then S will be justified in believing that  $q$ ], and (iii) S will tomorrow be unjustified in believing that  $q$  on the basis of E. However, unlike the case of *Standard Science*, the call in *Higher-Order Science* does not make it the case that (i) S is now unjustified in believing that [if S sees E tomorrow, then it is likely that  $q$ ]; in *Higher-Order Science* S remains justified

now, prior to the time of the supposed drugging, in believing that S's seeing E tomorrow would make  $q$  likely.

The temporal aspect of higher-order evidence is not just forward-looking; a similar effect can be achieved with backward-looking cases. Consider:

**Standard Backward Science:** S is a scientist investigating some phenomenon experimentally. Were S to get evidence, E, from the experiments, S would be justified in believing that  $q$ . What's more, S is justified in believing that were S to get E, S would be justified in believing that  $q$ . S went into the lab yesterday and found out that E obtains. However, just then, S receives a call from a reliable source informing S that there was a power-outage yesterday morning that reset the experiment in such a way that E no longer reliably indicate that  $q$ . As a matter of fact, there was no power outage.

In this case it is intuitive to believe that (i\*) S was justified in believing that  $q$  prior to the phone call, (ii\*) S ceased to be justified in believing that  $q$  after the phone call, and (iii\*) S is justified after the phone call in believing that [S was justified in believing that  $q$  prior to the phone call]. Contrast this with a case of higher-order evidence:

**Higher-Order Backward Science:** S is a scientist investigating some phenomenon experimentally. Were S to get evidence, E, from the experiments, S would be justified in believing that  $q$ . What's more, S is justified in believing that were S to get E, S would be justified in believing that  $q$ . S went into the lab yesterday and found out that E obtains. However, just then, S receives a call from a reliable source informing S that she was slipped a drug in her coffee yesterday

morning that rendered her incompetent with respect to forming beliefs on the basis of evidence like E. As a matter of fact, S was slipped no such drug.

In this case it is intuitive to believe, just like in *Standard Backward Science*, that (i\*) S was justified in believing that *q* prior to the phone call and (ii\*) S ceased to be justified in believing that *q* after the phone call. Unlike *Standard Backward Science*, however, it is not intuitive to believe that (iii\*) S is justified after the phone call in believing that [S was justified in believing that *q* prior to the phone call]. Rather, it seems like the higher-order evidence calls into question not just the current epistemic status of S's belief that *q* but also the epistemic status of S's belief that *q* yesterday. The standard defeater from *Standard Backward Science* does not seem to do this. As Maria Lasonen-Aarnio puts it:

Assume that I am told that what I believe based on perception to be a red object is in fact being illuminated by red trick lighting. Such evidence doesn't seem to cast any doubt on the epistemic rationality of my originally believing the object to be red based on my perceptual experience as of a red object. By contrast, defeat by higher-order evidence has a retrospective aspect, providing a subject with evidence that her belief was never rational, reasonable, or justified to start out with. (2014, 317)

In *Questionable Evidence*, the information about the file-swap causes a loss of your justification, but it does not call into question your justification prior to the phone call. You are justified in believing that [you were justified in believing that (Jones is guilty)]. However, in *Questionable Judgment*, the information you receive concerning the drugged coffee causes both a loss in your justification for believing that Jones is guilty and calls

into question your justification prior to the phone call. You cease to be justified in believing that [you were justified in believing that (Jones is guilty)].

Third, there is a peculiar circularity that can arise when dismissing higher-order evidence that does not arise when dismissing standard defeaters. In *Questionable Evidence*, imagine reasoning in the following way:

- (P1) The evidence in the file is such-and-such
- (C1) So, Jones is guilty
- (P2) I judged that Jones is guilty based on the evidence in the file I have
- (C2) So, I formed an accurate judgment about who is guilty of the crime
- (P3) I wouldn't have formed an accurate judgment about who is guilty of the crime if I had the fabricated the evidence
- (C3) So, I don't have the fabricated evidence

This line of reasoning is clearly problematic. (P1) is evidence for (C1) only if (C3) is true. So, S is licensed to infer (C1) from (P1) only if (C3) is true. So, this reasoning is rule-circular; it uses an inference rule that is only reliable in this case if the conclusion, (C3), is true. Now imagine employing parallel reasoning in *Questionable Judgment*:

- (P1) The evidence in the file is such-and-such
- (C1) So, Jones is guilty
- (P2) I judged that Jones is guilty on the basis of the evidence in the file I have
- (C2) So, I formed an accurate judgment about who is guilty of the crime
- (P3\*) I wouldn't have formed an accurate judgment about who is guilty of the crime if I had been drugged

(C3\*) So, I'm not drugged

Christensen and Andrew Rotondo (2013) argue that the reasoning here is not so obviously problematic. The inference rule, if (P1) then (C1), remains reliable even if (C3\*) is false. If this line of reasoning is circular, it is circular in a peculiar way. While Christensen and Rotondo disagree about the significance of these results, they agree that they point to an important difference between standard defeaters and higher-order evidence.

Lastly, and related to the previous point, Christensen argues that the admission of a standard defeater to one's evidence changes what that evidence supports in a way that admission of higher-order evidence to one's evidence does not. He considers a case similar to *Questionable Judgment* above:

**Deductive Drug:** I'm asked to be a subject in an experiment. Subjects are given a drug, and then asked to draw conclusions about simple logical puzzles. The drug has been shown to degrade people's performance in just this type of task quite sharply. In fact, the 80% of people who are susceptible to the drug can understand the parameters of the puzzles clearly, but their logic-puzzle reasoning is so impaired that they almost invariably come up with the wrong answers.

Interestingly, the drug leaves people feeling quite normal, and they don't notice any impairment. In fact, I'm shown videos of subjects expressing extreme confidence in the patently absurd claims they're making about puzzle questions. This sounds like fun, so I accept the offer, and, after sipping a coffee while reading the consent form, I tell them I'm ready to begin. Before giving me any

pills, they give me a practice question:

Suppose that all bulls are fierce and Ferdinand is not a fierce bull.

Which of the following must be true? (a) Ferdinand is fierce; (b) Ferdinand is not fierce; (c) Ferdinand is a bull; (d) Ferdinand is not a bull.

I become extremely confident that the answer is that only (d) must be true. But then I'm told that the coffee they gave me actually was laced with the drug. My confidence that the answer is "only (d)" drops dramatically. (2010, 187)

Of this case, Christensen writes:

[...] in the case where I'm immune, it is not obvious why my total evidence, after I learn about the drug, does not support my original conclusion just as strongly as it did beforehand. After all, the parameters of the puzzle are not rendered doubtful by my new information. The undermining is directed only at the simple deductive reasoning connecting these parameters to my answer. So there is a clear sense in which the facts which are not in doubt - the parameters of the puzzle - leave no room for anything other than my original answer. Or, to put it another way, the undoubted facts support my answer in the strongest possible way - they entail my answer - and this kind of connection cannot be affected by adding more evidence. Moreover, I even correctly see the entailment, and initially believe my answer in virtue of seeing the entailment. (2010, 187)

Contrast this with what he writes about a case involving a standard source-specific defeater:

[...] it seems to me that [the case involving standard defeaters] is very different

from the one in which I learn I've been drugged. In the [case involving standard defeaters], my reason for giving up my former belief does not flow from any evidence that my former belief was rationally defective. And insofar as I lack reason to worry about my epistemic malfunction, I may still use [...] the form of inference behind my original belief, whole-heartedly. It's just that with my present, enlarged pool of evidence, [that form of inference] no longer supports the [the conclusion I drew]. So the undercutting evidence does not prevent me from giving all of my evidence its due. (2010, 198)

The idea is that, where S believes that  $q$  on the basis of E, (E & higher-order evidence) supports S's believing that  $q$  as much as E alone does, but that (E & standard source-specific defeater) does not support S's believing that  $q$  as much as E alone does. It is this feature that Christensen takes to be fundamental in explaining the differences between higher-order evidence and standard defeaters.

## 2.2 *Dismissing Differences*

The existence of differences between higher-order evidence and standard defeaters is not, in and of itself, evidence that instances of higher-order evidence are not defeaters. What would count as evidence for this would be differences that indicate that higher-order evidence does not give one reason to not believe on some basis. There can, of course, exist broad differences between one kind of reason to not believe on some basis and another kind of reason to not believe on some basis. It is possible that some of these differences capture the distinction that is sometimes drawn between higher-order evidence and defeaters.

In the case of the first three supposed differences, this is particularly clear since the apparent differences are explicable under the reasons-against-belief theory and are not without exception. First, consider the supposed agent-relativity of higher-order evidence. It is suggested that higher-order evidence affects the justification of the subject it is about, but not the justification of other subjects who are aware of the higher-order evidence, whereas standard defeaters affect the justification of everyone aware of them. This suggestion is false. There are standard defeaters that are agent-relative in this way.

Consider two subjects, Sara and Tom. They are both working on a logic problem for class. They are instructed to find real-life premises using logical operators and see what they can prove. Sara and Tom together find that (1) if their friend Donny goes to the party, then Tasha won't go. What's more, they know that (2) Donny is going to the party. Sara hasn't yet learned *modus ponens*, so Sara proves that (3) Tasha won't go to the party by application of material implication, double negation, and disjunctive syllogism. Tom, on the other hand, has learned *modus ponens* as well as material implication, double negation, and disjunctive syllogism. Tom uses *modus ponens* to prove that (3) Tasha won't go to the party. Both Sara and Tom are justified in believing that (3) on the basis of (1) and (2). So, Sara and Tom have the same evidence. Later, Sara and Tom see in their text book that replacement rules like material implication are not valid in some contexts. Sara and Tom are both poorly informed enough about logic to be unable to rule out that this is case involves such a context. So, since the only inference rule Sara knew to prove that (3) was material implication, which Sara knows to be a replacement rule, Sara ceases to be justified in believing that (3). Tom, on the other hand, remains justified in believing



that (3).

The information from the text book is a standard source-specific defeater with respect to S's belief that (3), but this defeater is agent-relative since it has no effect on Tom. According to the reason-against-belief theory of defeat, what's going on here is that some reasons are agent-relative and some are not. Any source-specific defeater will be an agent-relative reason of this sort since different agents can have different sources for the same belief. In fact, it is not implausible to think that this is exactly what accounts for the agent-relativity of higher-order evidence. Since the application of my faculty for drawing conclusions from my evidence is necessarily a part of the source of my belief, any evidence that this faculty is flawed will be a source-specific defeater with respect to beliefs formed using that faculty. Necessarily, no one else can form their beliefs on the basis of exactly the same source as me since they must utilize their own faculty for drawing conclusions on the basis of their evidence. However, when higher-order evidence calls into question both my faculty and your faculty for drawing conclusions, e.g., you drank the potentially laced coffee, too, then it will affect both of us, just as a standard source-specific defeater can affect both of us if we share, even in part, a source for belief. While agent-relativity might be a feature of all instances of higher-order evidence, it cannot distinguish higher-order evidence from standard defeaters since some standard defeaters are also agent-relative.

Second, consider the supposed time-insensitivity of higher-order evidence. It has been suggested that higher-order evidence and standard defeaters differ in how they affect one's justification across time. But this suggestion, too, is not without exceptions.

Standard defeaters are sometimes time-sensitive. Recall the standard defeater from *Standard Science*:

**Standard Science:** S is a scientist investigating some phenomenon experimentally. Were S to get evidence, E, from the experiments, S would be justified in believing that  $q$ . What's more, S is justified in believing that were S to get E, S would be justified in believing that  $q$ . S is going into the lab tomorrow and will find out whether or not E obtains. So, tonight, S is justified in believing that if S sees that E tomorrow, it is likely that  $q$ , and S is also justified in believing that if S sees that E tomorrow, then S will be justified in believing that  $q$ . However, later at night, S receives a call from a reliable source informing S that there will be a power-outage tomorrow morning that will reset the experiment in such a way that E will no longer reliably indicate that  $q$ .

Imagine that, instead of being told that there will be a power-outage tomorrow morning that will reset the experiment in such a way that E will no longer reliably indicate that  $q$ , S is told that S will receive very convincing but misleading evidence tomorrow that there has been such a power-outage. Assume, also, that S is very forgetful and knows that this phone call will be forgotten tomorrow. In this case, S ceases to be justified in believing that if S sees that E tomorrow, S will be justified in believing that  $q$ , but S remains justified in believing that if S sees that E tomorrow, then it is likely that  $q$ . This is the same result delivered in *higher-order science*, but here the information S receives on the phone is not higher-order evidence in the sense that it is not evidence about S's competence with respect to forming beliefs on the basis of things like E. This is also, no

doubt, a strange case, but it is a strange case of standard defeat. What S receives is a reason to not believe that if S sees E tomorrow, then S will be justified in believing that  $q$ . However, this is not a reason for S to not believe now that if S sees that E tomorrow, then it is likely that  $q$ . This is precisely how the higher-order evidence in *Higher-Order Science* would be characterized by the reasons-against-belief theory, too. We will look at the retroactive aspect of higher-order evidence suggested by Lasonen-Aarnio in the next sub-section.

Third, consider the issue of circularity. It was suggested that, in the case of dismissing a standard defeater using circular reasoning, the kind of circularity employed was familiar, i.e., it involves rule-circularity. In the case of dismissing higher-order evidence, on the other hand, the circularity was supposed to be unfamiliar. However, this is not generally the case. Consider, first, a case of dismissing higher-order evidence:

(P1) It seems to me that  $p$  entails  $q$

(P2) It is the case that  $p$

(C1) So,  $q$

(P3) I judged that  $q$  on the basis of  $p$

(C2) So, I formed an accurate judgment about whether or not  $q$

(P4) If I had been incompetent with respect to forming beliefs on the basis of  $p$ ,  
then I wouldn't have formed an accurate judgment about whether or not  $q$

(C3) So, I am competent with respect to forming beliefs on the basis of  $p$

This argument involves circular reasoning in the following form: the inference from (P1) and (P2) to (C1) is only licensed if (C3) is true. Consider, next, an analogous case

involving standard defeaters:

(P1) It seems to me that  $p$  entails  $q$

(P2) It is the case that  $p$

(C1) So,  $q$

(P3) I judged that  $q$  on the basis of  $p$

(C2) So, I formed an accurate judgment about whether or not  $q$

(P4\*) If  $p$  did not entail  $q$ , then I wouldn't have formed an accurate judgment  
about whether or not  $q$

(C3\*) So,  $p$  entails  $q$

In this case, the reasoning appears circular, but it isn't clear how. In fact, it is unclear in the same way that cases of dismissing higher-order evidence were supposed to be. The inference from (P1) and (P2) to (C1) remains reliable even if (C3\*) is false. So, the way in which dismissing higher-order evidence and standard defeaters is circular depends not on whether it is higher-order evidence or standard defeaters that are being dismissed, but rather on just what premises and inferences one is relying on in dismissing them.

### 2.3 *Bracketing & Defeat*

The differences suggested by Lasonen-Aarnio and Christensen, however, seem to pose more of a challenge to the picture of higher-order evidence as being a kind of defeater. This is so because their suggestions point to a different relationship between higher-order evidence and first-order evidence than the reasons-against-belief theory posits between defeaters and first-order evidence for belief. According to the reasons-against-belief theory, defeaters are reasons to not believe. These reasons compete with

one's reasons to believe and the outcome of this competition explains the various phenomena of defeat. According to this theory, a standard defeater with respect to S's belief that  $q$  does not give S reason to believe that S's basis for believing that  $q$  does not, in fact, support believing that  $q$ . Rather, it merely gives S reason to not believe on that basis. For example, if you discover that the table before you that appears red is actually being illuminated by a red light, you cease to be justified in believing that the table is red on the basis of the table's appearance. Still, that a table appears red is still reason for you to believe that the table is red. It's just that you have reason to not believe that the table is red on that basis in this cases.

Contrast that with the picture of higher-order evidence suggested by Lasonen-Aarnio's comment that higher-order evidence functions by "providing a subject with evidence that her belief was never rational, reasonable, or justified to start out with" (2014, 317). The most natural way to explain this supposed contrast would be by positing that higher-order evidence with respect to S's belief that  $q$  gives S reason to believe that S's basis for believing that  $q$  does not actually support believing that  $q$ . This is why, in *Backward Higher-Order Science*, S not only ceases to be justified in believing that  $q$  but also ceases to be justified in believing that [S was justified in believing that  $q$ ].

We should be careful to spell out just what is and is not going on in *Backward Higher-Order Science*. Getting evidence that S was drugged yesterday does not change the actual epistemic relation between E and S's belief that  $q$  any more that getting evidence that there was a power-outage does. Rather, getting evidence that S was drugged yesterday gives S reason to believe that E does not support S's believing that  $q$ . And,

quite plausible, reason to believe that E does not support S's believing that  $q$  is reason for S to not believe that  $q$  on the basis of E, even if E really does support S's belief that  $q$ . If S's higher-order evidence is a reason for S to believe that E is not an indicator of  $q$  at all, then surely that's a reason for S to not believe that  $q$  on the basis of E. Since E really is a reason for S to believe that  $q$ , the higher-order evidence that S was drugged can function through the competition of reasons for and against S's believing that  $q$ .

This explains the retrospective aspect of higher-order evidence, too; since higher-order evidence is a reason for S to not believe that  $q$  on the basis of E in virtue of being a reason for S to believe that E does not support S's believing that  $q$ , the higher-order evidence is also a reason for S to believe that [S was not justified in believing that  $q$  on the basis of E]. But a reason for S to believe that [S was not justified in believing that  $q$  on the basis of E] is a standard defeater with respect to S's belief that [S was justified in believing that  $q$  on the basis of E], i.e., it is also a reason for S to not believe that [S was justified in believing that  $q$  on the basis of E].

Contrast this with the standard defeater that S has in *Standard Backward Science*. In that case, S acquired evidence that the power had gone out, thus making unreliable an otherwise reliable indicator of  $q$ , E. This evidence does not give S reason to believe that E does not support S's belief that  $q$ ; it just gives S reason to not believe that  $q$  on the basis of E in this case. Because of this, it does not give S a reason to believe that [S was not justified in believing that  $q$  on the basis of E], nor reason to not believe that [S was justified in believing that  $q$  on the basis of E]. There is no mystery about the retrospective aspect of higher-order evidence and it does not tell us anything deep about the structure

of justification. It obtains merely because higher-order evidence is a defeater both with respect to first-order beliefs and, unsurprisingly, also with respect to higher-order beliefs.

Christensen's suggestion that higher-order evidence requires one to bracket one's first-order evidence seems to introduce a new epistemic mechanism, one that is at odds with the competition-of-reasons-for-and-against-belief mechanism given by the reasons-against-belief theory of defeat. To bracket one's evidence is to rule it out-of-bounds, to not rely on it. But if higher-order evidence requires one to ignore one's first-order evidence, then the loss of justification due to higher-order evidence is not the result of a competition of reasons for and against belief but rather the result of losing the use of one's reasons for belief.

This mechanism is appealing to Christensen because he endorses something like the reason-neutralizing theory of defeat with respect to standard defeaters. He maintains that in a case of standard defeat, such as *Questionable Evidence*, one is free to utilize all of one's first-order evidence, but that when one's total evidence includes a standard defeater, one's total evidence does not support believing (2010, 198). Since it seems to him that one's total evidence does not cease to support believing when one's higher-order evidence is added, but one ceases to be justified in believing nonetheless, it must be that the higher-order evidence somehow prevents one from utilizing one's total evidence, i.e., it forces one to bracket one's first-order evidence.

However, as we saw in chapter 2, the reasons-neutralizing theory of defeat is flawed. Standard defeat can occur even when the addition of the standard defeater to one's total evidence does not result in one's total evidence no longer supporting belief.

Defeat cannot be explained at the level of total evidence; it requires finer distinctions than can there be drawn. When we gave defeat a closer look, we saw that defeat is best explained in terms of competition between one's reasons for believing on a given basis, which are not affected in any way by the acquisition of a defeater, and one's reasons for not believing on that given basis.

Given all of this, there is little pressure to adopt Christensen's bracketing mechanism. Some proper subset of one's evidence can support believing and one can give one's evidence its due while being unjustified in believing. Justification is not a function just of the support one has for believing; it is also a function of the support one has for not believing. Even where  $p$  entails  $q$ , one can still acquire reasons to not believe that  $q$  on the basis of one's belief that  $p$ .

Contrary to appearances, the reasons-against-belief theory of defeat faces no serious problems in explaining standard defeaters and higher-order evidence with a single underlying mechanism. This is not to say that there are no important differences between higher-order evidence and standard defeaters. Rather, it is to say that any differences that do exist between the two do not stem from some difference in the underlying mechanism responsible for their ability to remove one's justification.

### 3. *Conclusion*

The implications of the reason-against-belief theory that we've seen above are negative: externalism does not face any special problem in accommodating the phenomenon of defeat and there is no problem explaining higher-order evidence with the



same underlying mechanism that explains standard defeaters. This is not to say that externalism faces no problems at all or that the sorts of things that are identified as higher-order evidence can actually cause a loss of justification. Rather, it is just to say that these substantive epistemological issues do not seem to be settled by appeal to the phenomenon of defeat.

The reason-against-belief theory of defeat does, however, suggest directions for further research that might help to settle such substantive issues. For example, it has been asserted by some<sup>60</sup> that *a priori* justification is not susceptible to defeat or are not susceptible to defeat from *a posteriori* sources. According to the reasons-against-belief theory of defeat, this could only be because one cannot have reason, or perhaps just a *posteriori* reason, to not believe on the basis of a source of *a priori* justification. To vindicate or refute these claims, then, we must explore just what it takes to have an epistemic reason, for or against belief, and whether or not there is anything that could play the role of a reason against believing on the basis of an *a priori* source.

We should take note, however, that the reasons-against-belief theory of defeat does not offer an easy way to substantive epistemological conclusions. To the contrary, it suggests that the hard, fundamental work of understanding the nature and structure of epistemic reasons cannot be avoided; it suggests that there is no royal road to understanding.

---

60 See Philip Kitcher (1983, 1989), Hilary Putnam (1983), and Hartry Field (2000, 2005).

# WORKS CITED

- Alexander, David. 2015. "Can an Unjustified Belief Be a Defeater?" Unpublished. ([http://www.public.iastate.edu/~alexand/Research\\_files/Can%20an%20Unjustified%20Belief%20be%20a%20Defeater.pdf](http://www.public.iastate.edu/~alexand/Research_files/Can%20an%20Unjustified%20Belief%20be%20a%20Defeater.pdf))
- Alston, William. 1989. *Epistemic Justification: Essays in the Theory of Knowledge*. Ithaca: Cornell.
- Bergmann, Michael. 1997. "Internalism, Externalism, and the No-Defeater Condition." *Synthese* 110: 399-417.
- \_\_\_\_\_. 2005. "Defeaters and Higher-Level Requirements." *The Philosophical Quarterly* 55: 419-36.
- \_\_\_\_\_. 2006a. *Justification Without Awareness*. New York: Oxford.
- \_\_\_\_\_. 2006b. "BonJour's Dilemma." *Philosophical Studies* 131: 679-93.
- BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard.
- \_\_\_\_\_. 1992. "Externalism/Internalism." In *A Companion to Epistemology* Jonathan Dancy and Ernest Sosa eds. Malden, MA: Blackwell: 132-136.
- \_\_\_\_\_. 2006. "Replies." *Philosophical Studies* 131: 743-59.
- Bonjour, Laurence and Ernest Sosa. 2003. *Epistemic Justification*. Malden, MA: Blackwell.
- Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago: Chicago University Press.
- \_\_\_\_\_. 1963. "Replies and systematic expositions." In *The Philosophy of Rudolf Carnap* P. A. Schilpp ed. La Salle, IL: Open Court. 859-1013.
- Casullo, Albert. 2003. *A Priori Justification*. New York: Oxford.
- \_\_\_\_\_. 2005. "Epistemic Overdetermination and A Priori Justification." *Philosophical Perspectives* 19: 41-58.

- Chandler, Jake. 2013. "Defeat Reconsidered." *Analysis* 73: 49-51.
- Chisholm, Roderick. 1977. *Theory of Knowledge: 2<sup>nd</sup> Edition*. Englewood Cliffs: Prentice Hall.
- Christensen, David. 2010. "Higher-Order Evidence." *Philosophy and Phenomenological Research* 81: 185-215.
- Cohen, Stewart. 1983. "Justification and Truth." *Philosophical Studies* 46: 279-95.
- Dancy, Jonathan. 1993. *Moral Reasons*. Cambridge, MA: Blackwell.
- Feldman, Richard. 2005. "Respecting the Evidence." *Philosophical Perspectives* 19: 95-119.
- Field, Hartry. 2000. "A Priority as an Evaluative Notion." In *New Essays on the A Priori* Paul Boghossian and Christopher Peacocke eds. New York: Oxford University Press. 117-149.
- \_\_\_\_\_. 2005. "Recent Debates about the A Priori." In *Oxford Studies in Epistemology Volume 1* Tamar Gendler and Jonathan Hawthorne eds. New York: Oxford University Press. 69-88.
- Foley, Richard. 1987. *The Theory of Epistemic Rationality*. Cambridge: Harvard.
- Gettier, Edmund. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23: 121-3.
- Gibbons, John. 2006. "Access Externalism." *Mind* 115: 19-39.
- Goldman, Alvin. 1979. "What is Justified Belief?" In *Justification and Knowledge* George Pappas ed. Dordrecht: D. Reidel. 1-24.
- \_\_\_\_\_. 2009. "Replies to Discussants." *Grazer Philosophische Studien* 79: 245-88.
- Grundmann, Thomas. 2009. "Reliablism and the Problem of Defeaters." *Grazer Philosophische Studien* 79: 65-76.
- \_\_\_\_\_. 2011. "Defeasibility Theory." In *The Routledge Companion to Epistemology* Sven Bernecker and Duncan Pritchard eds. New York: Routledge. 156-66.
- Harman, Gilbert. 1973. *Thought*. Princeton: Princeton University Press.

- Kelly, Thomas. 2005. "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology Volume 1* Tamar Gendler and Jonathan Hawthorne eds. New York: Oxford University Press. 167-196.
- Kitcher, Philip. 1983. *The Nature of Mathematical Knowledge*. New York: Oxford University Press.
- \_\_\_\_\_. 1989. "Explanatory Unification and the Causal Structure of the World." in *Scientific Explanation*, P. Kitcher and W. Salmon eds. Minneapolis: University of Minnesota Press. 410–505.
- Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kvanvig, Jonathan. 2007. "Two Approaches to Epistemic Defeat." In *Alvin Plantinga Deane-Peter Baker* ed. New York: Cambridge. 107-124.
- Lackey, Jennifer. 2008. *Learning from Words: Testimony as a Source of Knowledge*. New York: Oxford University Press.
- Lasonen-Aarnio, Maria. 2014. "Higher-Order Evidence and the Limits of Defeat." *Philosophy and Phenomenological Research* 88: 314-45.
- Lycan, William. 2006. "On the Gettier Problem Problem." In *Epistemology Futures* Stephen Cade Hetherington ed. New York: Oxford.
- Moser, Paul. 1985. *Empirical Justification*. Dordrecht: D. Reidel.
- Olsson, E. J. 2012. "Reliabilism as Explicating Knowledge: A Sketch of an Account." In *Epistemology: Contexts, Values, Disagreement*. C. Jäger and W. Löffler eds. Germany: Ontos Verlag. 189-202.
- \_\_\_\_\_. 2014. "Gettier and the Method of Explication: a 60 Year Old Solution to a 50 Year Old Problem." *Philosophical Studies* Open Access.
- Pollock, John. 1974. *Knowledge & Justification*. Princeton: Princeton.
- \_\_\_\_\_. 1986. *Contemporary Theories of Knowledge*. Savage: Rowman & Littlefield.
- Pollock, John and Joseph Cruz. 1999. *Contemporary Theories of Knowledge: 2<sup>nd</sup> edition*. Lanham: Rowman & Littlefield.
- Pryor, Jim. 2015. "The Merits of Incoherence." Unpublished. (<http://www.jimpryor.net/research/papers/Antilogic5.pdf>)

- Putnam, Hilary. 1983. "Two Dogmas' Revisited." Repr. in his *Realism and Reason: Philosophical Papers, Vol. 3*: 87-97. Cambridge: Cambridge University Press.
- Quinton, Anthony. 1955. "The Problem of Perception." *Mind* 64: 28-51.
- Rotondo, Andrew. 2013. "Undermining, Circularity, and Disagreement." *Synthese* 190: 563-584.
- Steup, Matthias. 1998. *An Introduction to Contemporary Epistemology*. Upper Saddle River: Prentice Hall.
- Wheller, Gregory. 2014. "Defeat Reconsidered and Repaired." *The Reasoner* 8: 15.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. New York: Oxford University Press.
- van Inwagen, Peter. 1996. "It is Wrong Always, Everywhere, and for Anyone, to Believe on the Basis of Insufficient Evidence." In *Faith, Freedom and Rationality* J. and D. Howard-Synder eds. Maryland: Rowman and Littlefield. 137-154.