The Indispensability of Conscious Access

A thesis submitted

To Kent State University in partial

Fulfillment of the requirements for the

Degree of Master of Arts

by

Steven Winterfeldt

August, 2016

© Copyright

All Rights Reserved

Except for previously published materials
Thesis written by

Steven Winterfeldt

B.S., Austin Peay State University, 2012

M.A., Kent State University, 2016

Approved by

Deborah Smith, Advisor

Deborah Barnbaum, Chair, Department of Philosophy

James L. Blank, Dean, College of Arts and Sciences
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................... v

CHAPTER 1: Knowledge as Justified True Belief ........................................... 1

   Section I: Knowledge Generally ....................................................... 1

   Section II: Causal Origin and Justification ......................................... 6

   Section III: The Gettier Problem ...................................................... 7

   Section IV: Overview ...................................................................... 13

CHAPTER 2: Externalist Theories of Epistemic Justification .......................... 14

   Section I: Alvin Goldman’s Causal Theory of Knowledge ..................... 14

   Section II: Armstrong’s Thermometer Model of Knowledge .................. 20

   Section III: Goldman’s Process Reliabilism ....................................... 26

   Section IV: Conclusion .................................................................... 30

CHAPTER 3: Internalist Theories of Epistemic Justification .......................... 31

   Section I: Accessibilism .................................................................. 31

   Section II: Mentalism ..................................................................... 35

   Section III: Deontologial Conceptions of Justification ....................... 40

   Section IV: Conclusion .................................................................... 44

CHAPTER 4: Arguments ......................................................................... 45

   Section I: The Unreflective Agent ..................................................... 46

   Section II: Objectively vs. Subjectively Good Reasons for a Belief ........ 48

   Section III: Clairvoyance ................................................................. 51

   Section IV: The Brain in a Vat ......................................................... 56

   Section V: Conclusion .................................................................... 59
ACKNOWLEDGEMENTS

I would first like to thank the members of my thesis committee: Dr. Deborah Smith, Dr. Michael Byron, and Dr. Polycarp Ikuenobe for their support and guidance in the process of writing this thesis. This thesis would have been impossible without the help of these individuals. I would especially like to thank my thesis adviser, Dr. Smith, for her insight, feedback, and patience throughout the writing and editing phases. I owe a debt of gratitude to many members of the philosophy department at Kent State University for their encouragement and support over the past two years. In particular, I would like to thank Dr. David Pereplyotchik for his advice and guidance. I would also like to thank my family and friends for encouraging me to pursue my academic goals even when it required that I move far away from home.
Knowledge as Justified True Belief

I. Knowledge Generally

Philosophers have long sought to account for what we are capable of knowing and how we might come to know it. When we say that Jones knows that the Earth revolves around the sun or knows that there is a book on the desk in front of him, we are attributing to him propositional knowledge. It is called propositional knowledge because the content of the belief can be expressed by a declarative sentence in some language (e.g. the English sentence ‘The book on the desk is red’) and these sentences express propositions. Propositional knowledge, or knowledge-that, can be distinguished from two other types of knowledge: knowledge by acquaintance, as when Smith knows Jones, and operational knowledge or knowledge-how, as when Smith knows how to ride a bike or drive a car. Although there are interesting questions concerning each kind of knowledge, for the purposes of this thesis, I am concerned exclusively with propositional knowledge.

One of the central projects in the practice of epistemology is to determine what conditions are independently necessary and jointly sufficient for knowledge. For the greater part of the past 2,400 years, it was generally agreed that knowledge involved three conditions: (i) belief, (ii) truth, and (iii) justification. From roughly the time of Plato until the 20th century, the so-called justified, true, belief theory of knowledge was the established model under which the epistemic debates throughout the history of philosophy raged, most notably in the modern era.
That the conditions of belief and truth are necessary conditions for knowledge is relatively unproblematic. In order for a proposition to qualify as knowledge, it must be believed. If Smith does not even believe that there is a red book on the desk, there is no sense in which he can know that there is a red book on the desk. So belief must be at least one of the necessary conditions for knowledge. But belief alone is not sufficient. Even if Smith believes that there is a red book on the desk, if there is not, in fact, a red book on the desk, his belief cannot count as knowledge. He cannot know because his belief is not true. To be sure, we might not always be aware of whether our beliefs are true. Whether a belief is true or not is a matter of objective fact, not subjective determination.¹ A belief may be true regardless of whether an agent is aware of the fact that it is true, and a belief (justified or otherwise) that is not true cannot count as knowledge, no matter how fervently an agent believes or how much justification he may have for that belief. While one may not be aware of whether one’s belief is true or not, one cannot know what is not true, and so truth appears to be one further necessary condition for knowledge.

It is often agreed that true belief is not sufficient for knowledge, however. Since it is possible to hold an accidentally true belief, something more than mere true belief is needed.² A belief must also be justified. Suppose that Smith accuses Jones of stealing a book from the bookstore. Smith has the belief that Jones stole the book. Suppose, further, that Jones did in fact

¹ While it is generally accepted that truth is one of the conditions for knowledge, there has been some debate over what makes a proposition true. Most often, it is argued that a proposition is true if it corresponds with the mind-independent world as it actually exists. Some philosophers have offered alternative views of truth. One need not take truth to be correspondence with a mind-independent reality in order to hold that truth is a matter of objective fact. Even on views such as Putnam’s idealized end of inquiry view (see Hilary Putnam, *Reason, Truth, and History* (Cambridge, MA: Cambridge University Press, 1981)), or Dummett’s verificationism (see Michael Dummett, “Truth”, *Proceedings of the Aristotelian Society*, 59 (1959), 141-162), or Wright’s superassertability view (see Crispin Wright, *Truth and Objectivity* (Cambridge, MA: Harvard University Press, 1992)), truth is still objective in that a person or even all people can be mistaken about what the truth is. For my purposes, I will simply assume that truth is correspondence with reality.

steal the book. Smith’s belief would also be true. But suppose that Smith assumes that Jones stole the book simply because Smith does not like Jones and wants to frame him for the theft. Although Smith’s belief is true, it falls short of knowledge because it is only accidentally true. Given Smith’s epistemically weak reasons for believing as he does, the fact that his belief is true seems to be merely a matter of luck. His belief therefore fails to satisfy the justification condition. Since it is unjustified, it cannot count as a bit of knowledge, despite the fact that Smith believes it and that the belief is in fact true.

We often use the word ‘justified’ (or speak of something’s being justified) in many different contexts. We might say, for instance, that someone is justified in keeping a promise. Such an action might be said to be ethically justified. Ethical justification aims at what is good. Alternatively, we might say that something is pragmatically justified. Pragmatic justification aims at self-interest or self-preservation. Someone who holds a belief in God on the basis of an argument such as Pascal’s Wager might be pragmatically justified in holding that belief, because holding such a belief arguably carries with it the greatest benefit for the least risk. Such a belief (on those grounds) would not be epistemically justified, however. A belief is epistemically justified, in the broadest sense, if there is a reason why the belief is highly likely to be true. Epistemic justification, in contrast to ethical and pragmatic justification, aims at what is true. Pascal’s Wager is a pragmatic argument, not an epistemic one: we should believe that God exists not because there is some reason why a belief that God exists is likely to be true, but because such a belief plausibly carries with it the most favorable cost-benefit ratio.

In some cases, these different types of justification might overlap. For example, to the extent that beliefs are action-guiding, one’s justification for acting a certain way might be based, at least in part, on one’s epistemically justified beliefs. One’s decision to extend the sphere of
moral consideration to animals might be based, in part, on one’s belief that they have central nervous systems and can therefore feel pain. In other cases, a belief can be both pragmatically and epistemically justified. The sciences seem to provide us with many examples of this sort of overlap. For example, scientists might posit the existence of some physical law of the universe, not only because such a belief might be epistemically justified by scientific observation, but also because positing the existence of that physical law results in a more explanatorily successful scientific theory. But such overlap will not always be the case. Our actions are not always motivated by epistemic reasons. There are instances in which we act out of passion, or anger, or faith without “thinking” about it. In many other instances, an epistemically justified belief might produce no action at all. My belief that a molecule of water is composed of two hydrogen atoms and one oxygen atom may produce no further actions on my part. There are also cases where a belief might be pragmatically justified but not epistemically justified (or vice versa), as in the example above of Pascal’s Wager. While there may often be overlap between the various kinds of justification, we can clearly distinguish between them. When epistemologists speak of a belief being justified, they are referring to epistemic justification, the sort of justification that aims at truth.

Epistemic justification is an important condition for knowledge because it is supposed to prevent those instances in which a belief is merely true by accident from counting as knowledge. A justified, true belief is thus something that is more valuable than mere true belief alone. Ernest Sosa offers us an analogy to the sport of archery that helps to illustrate the added value of justified, true belief.³ When engaged in the sport of archery, it is correct to say that the intrinsic goal of the sport is to hit the target as close to the bullseye as possible. It is not enough, however,

---
that the goal be achieved by mere chance or luck. Rather, the archer strives to hit the mark with a certain skill. The best archers are those that are able to adjust their aim to account for changes in range or weather or other conditions, thereby exhibiting a particular skillfulness in hitting the target. Similarly, the fact that a belief is justified reflects a certain epistemic skill.

If an agent (call him Smith) flips a coin and then comes to form the belief, perhaps on the basis of wishful thinking (or perhaps on no basis at all), that the coin will land on its edge and, if somehow the coin does land on its edge, we would not attribute to Smith knowledge that the coin would land on its edge because there was no reason why the belief was likely to be true. Similarly, Smith’s belief that Jones stole the book on the basis that he wants to frame Jones is not a belief that is highly likely to be true. In both situations, the agents’ belief is true merely by luck. This would be analogous to the archer who hits the bullseye merely by chance. However, if Smith were to, e.g. have visual evidence which suggests that Jones stole the book and then form the belief that Jones stole the book based on that evidence, such a belief would be highly likely to be true. Smith’s belief is epistemically appropriate given his visual evidence. This would be analogous to the archer who not only hits the bullseye but also does so with skill. Justification is thus a third condition for knowledge. In order for a true belief of Smith’s to count as knowledge, it must not only be believed and be true, but it must also be epistemically appropriate or justified.

Historically, it was commonly thought that justification guaranteed truth, although most epistemologists now agree that it is possible to hold a justified but false belief, just as even a skilled archer might occasionally miss the target due to some unforeseen circumstance. Perhaps it was not Jones, but his identical twin, that Smith saw steal the book. Plausibly, his belief would still be justified, even if it is false. Most contemporary epistemologists now hold that justification is fallible—it makes a belief likely to be true, but it does not guarantee that the belief is true.
II. Causal Origin and Justification

Beliefs can be divided into two types based on their causal origin in the epistemic agent: non-inferential beliefs and inferential beliefs.\textsuperscript{4} Causally non-inferential beliefs typically include beliefs based on sense experience, introspection, intuition, and memory. Such beliefs are non-inferential in the sense that they are not the products of any explicit inferential process. Other beliefs are causally inferential in the sense that they \textit{are} products of some explicit inferential process. Such beliefs include beliefs based on the use of deduction or induction. Suppose that I have the belief that all humans are mortal and also the belief that Socrates is a human. I then conclude from these two beliefs that Socrates is mortal. The concluding belief is the product of a deductive inference from the premise beliefs and is thus a causally inferential belief.

To speak of a belief’s causal origin is to say nothing with regard to how or why that belief is justified. However, this distinction does often parallel a similar distinction in justification (though it need not always do so). It is often agreed that causally inferential beliefs are also justified inferentially (if they are justified at all).\textsuperscript{5} But there is considerable disagreement about how a causally non-inferential belief is justified. Since these beliefs are not the product of any inference, their justification might also be non-inferential in that it plausibly does not involve an inferences from any other beliefs. Foundationalists\textsuperscript{6} often take this to be the case with simple empirical beliefs. In proper conditions, if Smith observes a red book on the desk, then he seems prima facie justified, via sense experience, in believing that there is a red book on the desk. If he

\textsuperscript{4} Laurence BonJour, \textit{The Structure of Empirical Knowledge} (Cambridge, MA: Harvard University Press, 1985), 112-113. Here, I am borrowing a distinction made by Laurence BonJour between beliefs that are \textit{causally} (non-)inferential and beliefs that are \textit{justified} (non-)inferentially.

\textsuperscript{5} For a causally inferential belief to be justified inferentially, the inference used must be truth conducive and the beliefs that serve as the premises of the inference must themselves be justified.

\textsuperscript{6} Foundationalism is the view that causally non-inferential beliefs can be non-inferentially justified, i.e. these beliefs are self-justifying or otherwise possess some degree of justification that is immediate or intrinsic. See BonJour, \textit{The Structure of Empirical Knowledge}, 17.
is not observing the book but is instead remembering where he laid it, then assuming that his memory is accurate, he is at least prima facie justified, via introspection and memory, in holding his belief about the book. Highly complex and sophisticated beliefs, however, might be causally non-inferential but only justified inferentially. Smith’s belief that a certain work of art was made by M.C. Escher is likely justified only in virtue of its logical relationship to other beliefs that Smith holds, even if it was caused non-inferentially by viewing the work of art. The central question of the debate between foundationalists and coherentists⁷ is whether there are any causally non-inferential beliefs that are also justified non-inferentially. While that debate is indirectly related to the project that I am developing here, whether foundationalism or coherentism is ultimately correct has no direct bearing on my project. As such, I will leave it an open question as to whether causally non-inferential beliefs are justified inferentially or non-inferentially.

### III. The Gettier Problem

As was briefly mentioned earlier, the justified, true, belief account of knowledge has its roots in the Platonic dialogues. In the *Theaetetus*, after having determined that mere true belief is not sufficient for knowledge, the dialogue’s namesake suggests to Socrates that knowledge is “true belief with the addition of an account,” and that “[w]here no account could be given of a thing, it was not ‘knowable’ . . .”⁸ In the *Meno*, Socrates has this to say about true belief versus knowledge:

> True opinions are a fine thing and do all sorts of good so long as they stay in their place, but they will not stay long. They run away from a man’s mind; so they are not worth much until you tether them by working out the reason. . . . Once they are tied down, they become knowledge and are stable. That is why knowledge is

---

⁷ Coherentism is the view that a belief is justified if and only if it belongs to a coherent system of beliefs.

⁸ Plato, *Theaetetus*, 201d.
something more valuable than right opinion.\textsuperscript{9}

It is worth noting that Plato himself rejected the justified, true, belief conception of knowledge on the grounds that to have an “account” of something meant to have knowledge of the differences between and among things. The consequence of this, according to Plato, was that in order to have knowledge, we must already possess knowledge.\textsuperscript{10} If this were the case, then a vicious regress would ensue. On Plato’s view, knowledge is instead a matter of recollection. Nevertheless, the justified, true, belief conception of knowledge became the established model throughout most of the history of philosophy.

In 1963, Edmund Gettier published a paper that was as short as it was significant. Gettier challenged the established justified, true, belief account of knowledge by arguing that it was possible to have a justified, true belief that still did not amount to knowledge. Gettier’s counterexamples turn on two simple points: (i) it is possible to have a justified but false belief (as we have already seen), and (ii) if some agent \( S \) is justified in believing some proposition \( P \) and \( P \) entails another proposition \( Q \) and \( S \) infers \( Q \) from \( P \) and accepts \( Q \) on the basis of this inference, then \( S \) is justified in believing \( Q \).\textsuperscript{11}

Gettier presents two counterexamples in his paper, although I will only focus on one.\textsuperscript{12} He asks us to suppose that Smith has strong evidence for believing the following proposition about one of his colleagues, Jones:

\[(f) \text{ Jones owns a Ford.} \]

Perhaps Smith has seen Jones driving a Ford and Jones has even offered to give him a ride in the

\textsuperscript{9} Plato, \textit{Meno}, 98a.
\textsuperscript{10} Plato, \textit{Theaetetus}, 210a.
\textsuperscript{12} Ibid., 193.
car. In such a case Smith would appear to be justified in believing (f). From this, Smith constructs three disjunctive propositions:

(g) Either Jones owns a Ford or Brown is in Boston.

(h) Either Jones owns a Ford or Brown is in Barcelona.

(i) Either Jones owns a Ford or Brown is in Brest-Litovsk.

Each of these propositions follows deductively from (f). If Smith is justified in believing (f), and if he deduces any one of these propositions from (f) and accepts it on the basis of this inference, then he is justified in believing whichever proposition he deduces, (g), (h), or (i). Let us suppose that Smith comes to believe (h) on this basis even though he is wholly ignorant of Brown’s actual location at this time.

Suppose further that, entirely unbeknownst to Smith, Jones does not in fact own a Ford. He has merely been pretending to own a Ford. Brown, however, is in fact in Barcelona. Smith’s belief (h) is true and it is justified because it has been inferred from another justified (though false) belief (f). Smith thus has a justified, true belief, but not, it seems, knowledge. The evidence that justified (h) is in no way connected to the truth of the belief. His justification concerns Jones’ ownership of the car, while the truth of the proposition is guaranteed by Brown’s actual whereabouts. Smith’s justified belief just so happens to be true.

Gettier is often credited with being the first to refute the justified, true, belief account of knowledge, but other counterexamples to the established account of knowledge can be found in the literature prior to the publication of his influential paper. Perhaps the best example of this is Bertrand Russell’s stopped clock case.\(^{13}\) Suppose that you are walking by a clock tower, and you

decide to inquire as to the time of day. As you look at the clock tower, you see that it reads 5 o’clock. On this basis, you form the belief that it is 5 o’clock. The belief appears to be justified by virtue of your visual experience, and in fact it is 5 o’clock, so your belief is true. But unbeknownst to you, the clock stopped ticking exactly twelve hours ago, and it just so happens to read 5 o’clock. Russell’s claim was that, in such a situation, one does not have a justified belief that it is 5 o’clock. As I have described the case, it is not immediately clear that what goes wrong is the justification of the belief. And so it is easy to see how Russell’s stopped clock can be turned into a Gettier example. One can insist that such a belief is justified and true, but still falls short of knowledge for some other reason. It is worth nothing, however, that Russell’s example differs from Gettier’s in that it does not depend on an inference from a justified but false belief.

Gettier-type counterexamples effectively demonstrate that justification is not sufficient to rule out accidentally true belief as cases of knowledge. In other words, the established justified, true, belief account of knowledge might not be the whole story. Much of the work done by epistemologists in the decades after Gettier has been in response to the problem that he raised. Gettier effectively forced a complete re-examination of the established account of knowledge.

There are two broad approaches to addressing the Gettier problem. Some philosophers sought to supplement the established justified, true belief account with some fourth “anti-Gettierizing” condition. A variety of possible anti-Gettierizing conditions have been proposed. One such suggestion is that one’s justified, true belief cannot be based on a false premise. This sort of view has been advocated by Michael Clark and Gilbert Harman. On this view, $S$ knows that $P$ if and only if (i) $S$ believes that $P$, (ii) $P$ is true, (iii) $S$ is justified in believing that

---

$P$, and (iv) $P$ is not based on any falsehood. While such a condition would seem to address Gettier’s own examples, it is possible to construct a Gettier case that does not involve an inference from a false belief. Russell’s stopped clock case is one such example (though, again, this is not how it was originally conceived).¹⁶

Another possible anti-Gettierizing condition that has been proposed is the defeasibility condition. This sort of view has been articulated by Keith Lehrer and Thomas Paxson¹⁷ and by Peter Klein.¹⁸ On the defeasibility view, $S$ knows that $P$ if and only if (i) $S$ believes that $P$, (ii) $P$ is true, (iii) $S$ is justified in believing that $P$, and (iv) $P$ is undefeated. The justification for a belief is defeated if there is some true proposition $Q$ such that, if $S$ were to become aware of $Q$, $S$ would no longer be justified in believing that $P$. In Gettier’s case, the defeater is that Jones is merely pretending to own a Ford. In Russell’s case, the defeater is that the clock has stopped working.

It is worth noting that whether we adopt the no false premises condition or the defeasibility condition (or any of the other possible anti-Gettierizing conditions that are present in the literature) such a condition is an external condition on knowledge. A condition is external if an epistemic agent need not have conscious access to whatever satisfies the condition. A condition is internal if, when satisfied, the agent has conscious access to whatever satisfies the condition. Truth is an external condition on knowledge, because an agent can know a proposition without conscious access to the fact that it is true. Belief is generally taken to be an internal condition on knowledge, because an agent will presumably have conscious access to whether or

---

¹⁶ Harman is aware of Russell’s stopped clock case. His account of inference as abduction should guarantee that there is a false premise. The false premise in Russell’s case is that the visual evidence is best explained by the fact that it is 5 o’clock. See Gilbert Harmon, “Thought, Selections,” in Epistemology: An Anthology, 194-201.


not he believes that $P$. Both of the potential fourth conditions offered here are external conditions on knowledge. An agent generally will not have conscious access to the fact that none of his premises are false or to the fact that there is no defeater for his justification. Those who are inclined to respond to Gettier’s challenge by adopting a fourth, external condition on knowledge typically do so because they wish to preserve a more traditional interpretation of justification as an internal condition on knowledge.

Other philosophers have suggested that answering the Gettier problem requires that we rethink the justification condition for knowledge. Proponents of this approach typically advocate some theory of justification as an external condition on knowledge. On this view, an agent might be able to recognize that his belief satisfies the justification condition (however that is cashed out) and thus have conscious access to the factors that justify his belief. What the justificatory externalist rejects is that this will always be the case. As Alvin Goldman, a prominent externalist, writes:

> There are many facts about a cognizer to which he lacks “privileged access,” and I regard the justificational status of his beliefs as one of those things. This is not to say that a cognizer is necessarily ignorant, at any given moment, of the justificational status of his current beliefs. It is only to deny that he necessarily has, or can get, knowledge or true belief about this status. . . [H]e can have a justified belief without knowing that it is justified (or believing justifiably that it is justified).\(^{19}\)

The externalist thus eschews the requirement of conscious access as an important aspect of the concept of justification. Epistemic externalism was initially designed to counter the Gettier problem, although the theory can now be motivated independently of the problem and it has since taken on a life of its own.

Whatever the solution to the Gettier problem itself might be, the problem shows that

---

there is a difficulty with the justified, true, belief account of knowledge as it has traditionally been conceived. Those philosophers that adopt an external account of justification reject the traditional, internal account of justification. Those philosophers who wish to preserve the traditional interpretation of justification as an internal condition on knowledge may do so, but any internalist theory of justification will have to be supplemented by some fourth, external condition for knowledge in order to solve the Gettier problem.

IV. Overview

In this thesis, I will focus on the debate between internalists and externalists concerning epistemic justification. Epistemic externalism proposes that a belief may be justified by factors that are or may be external to the conscious awareness of the agent. In the next chapter, I will explore externalist theories of epistemic justification in more detail. Epistemic internalism, on the other hand, attempts to preserve the justification condition as it has traditionally been understood, as an internal condition on knowledge. In chapter three, I will explore internalist theories of justification. In the fourth chapter, I will examine two arguments in favor of externalism and two in favor of internalism. In the final chapter, I will defend the idea that conscious access to the factors that justify a belief is a crucial feature of epistemic justification. As my explicit concern is the debate between these two views concerning the nature of the justification condition for knowledge, from this point forward I will largely ignore discussion of knowledge more generally.
Externalist Theories of Epistemic Justification

As noted in the first chapter, the key feature of externalist theories of epistemic justification is that they do not require that an agent have conscious access to the factors that justify a belief in order for that belief to be justified. In this chapter, I will explore three different versions of externalism. In section I, I will discuss Goldman’s early causal theory of knowledge. D.M. Armstrong’s thermometer model of knowledge will be the subject of section II. Goldman’s later process reliabilism, the most popular version of externalism in the contemporary debate, will be discussed in section III. Each of these theories appeals to some external fact in determining whether or not a belief is justified. Causal theories appeal to causal connections. Armstrong’s theory appeals to law-like connections (which are broader than causal connections). Process reliabilism appeals to the reliability of the process or mechanism that produced the belief.

I. Alvin Goldman’s Causal Theory of Knowledge

Goldman’s causal theory of knowledge is an early version of externalism. As such, it was initially designed to counter the Gettier problem. His diagnosis of what goes wrong in Gettier-style cases is that a necessary causal connection is lacking between the fact that \( P \) and the agent’s belief that \( P \). Recall that Smith has strong reasons for believing the claim that Jones owns a Ford.

\[20\] Although (early) Goldman and Armstrong frame their theories as accounts of knowledge rather than justification simpliciter, in both cases their discussion is equally applicable to the justification debate. Nearly any instance of ‘knowledge’ can be replaced with ‘justified belief’ and their respective points remain the same.
From this, he deduces via a valid inference the further proposition that ‘either Jones owns a Ford or Brown is in Barcelona’ (call this disjunctive proposition $R$). Keep in mind that Smith is entirely ignorant of the actual whereabouts of Brown. Although $R$ is true, the reason that it is true is because Brown is in Barcelona, not because Jones owns a Ford. The truth of the proposition is in no way causally connected to Smith’s reason for believing that Jones owns a Ford. What makes the proposition true is the fact that Brown is in Barcelona. What leads Smith to believe the proposition is his evidence for the belief that Jones owns a Ford. There is no causal connection between Smith’s reason for his belief and the fact that makes his belief true.

On Goldman’s causal account of knowledge, “$S$ knows that $P$ if and only if the fact $P$ is causally connected in an ‘appropriate’ way with $S$’s believing $P$.“^21 To illustrate his point, let us consider a rather simple variant of Gettier’s case. Suppose that Jones does in fact own a Ford. In that case, Smith’s belief that $R$, which he has deduced from his belief that Jones owns a Ford, would then be causally connected in an appropriate way to the fact that makes it true. Smith would then have knowledge (justified, true belief) that $R$. There are a variety of ways that a belief can be causally connected in an appropriate way to the fact that makes it true. Appropriate knowledge producing causal processes include, but are not limited to, sense perception, memory, inference,^22 and testimony. Visual perception is the paradigm example of an appropriate knowledge producing causal process. In cases of visual perception, some kind of causal mechanism is clearly at work. While the exact specifics of the causal mechanism is a problem better left to perceptual scientists, the fact that some kind of causal process is at work can be demonstrated by the fact that, if such a causal process were absent, an agent would not form the

^22 Inferential relations are not commonly thought of as causal relations. I will address this point later in this section.
associated belief. An agent who was blind, for example, would not form the visually based belief that there is a book sitting on the table. He might form the belief that a book is sitting on the table via some other appropriate causal process, perhaps via another form of perception such as touch or via testimony from some third-party agent, but he would not form the belief via visual perception.

There would also appear to be a causal process at work when one remembers something. As Goldman says, “S remembers P at time t only if S’s believing P at an earlier time is a cause of his believing P at t.” Note that it is not sufficient that an agent believe P at some point in time and also believe P at a later time. An agent may have learned P at t₁, forgotten it, and then relearned it at t₂. In such cases, a causal connection is clearly lacking between the beliefs at t₁ and t₂, and so we would not be willing to attribute knowledge based on memory to the agent.

Only some of the things that we know are based on perception or memory, or some combination of the two. Much of our knowledge is based on inference, but it is not immediately clear how inference is itself a causal process. Although Goldman is inclined to say that it is, he is careful not to phrase his causal condition on knowledge in terms of direct causation. His explicit reasoning for this move is that, if we require that the fact that P must directly cause the corresponding belief in the agent, then the condition would be too strict and would wrongfully preclude instances of foreknowledge, unless we want to defend the plausibility of backward causation. Suppose that Smith tells Jones, on Friday, that he (Smith) will be driving to New York City the following Monday. Smith’s assertion leads Jones to believe that Smith will drive

23 Ibid., 359.
24 Ibid., 360.
25 Ibid.
26 It should be noted that, as Goldman uses the term ‘inference’, he is counting only truth-conducive (not fallacious) inferences. Additionally, to say that S knows that P by inference does not entail that S has gone through an explicit and conscious process of inferential reasoning.
27 Ibid., 362.
to New York City on Monday. And if Smith does drive to New York City on Monday, then Jones’ belief is true. But Smith’s driving to New York City on Monday cannot directly cause Jones’ belief to be true on Friday (again, unless we want to defend the plausibility of backward causation). Since there is no direct causal connection between Jones’ belief and the fact that makes it true, then conceiving of a causal connection in this way would seem to wrongfully preclude Jones’ belief from counting as knowledge.

On Goldman’s view, it is sufficient that both the fact and the belief have a common cause. Suppose that $S$ comes to believe that $Q$ on the basis of an inference from $P$, in combination with other background beliefs. If a continuous causal chain exists between the fact $Q$ and $S$’s belief that $Q$ and, if the inference is an objectively truth-conducive one, then $S$ knows that $Q$. To borrow Goldman’s example, suppose that a volcano erupts, coating the surrounding countryside in lava. At some later time, $S$ perceives that the lava is now solidified ($P$) and concludes that the volcano erupted ($Q$). The sort of continuous causal chain involved here includes the eruption of the volcano that causes the lava flow that solidified and that is presently the cause of $S$’s perception. There is a causal connection, via visual perception, between the fact that $P$ and $S$’s belief that $P$. There is a further causal connection between the presence of the solidified lava and the past fact $Q$. The presence of the solidified lava and $S$’s belief that $P$ have a common cause: the eruption of the volcano. Since the appropriate sort of causal chain does exist, and since the inference from $P$ to $Q$ is an objectively truth-conducive one, $S$ knows that $Q$, namely that the volcano erupted, on the basis of an inference from $P$. If this causal chain is not present, then $S$ does not know that $Q$. And so Goldman can address the objection that inference is not itself a causal process. He need not claim as much. His claim is only that, if a chain of

---

28 For a diagrammed example of this sort, see ibid., 364-365.
29 Ibid., 361.
objectively good inference is added to a causal chain, the whole chain then becomes causal.

Knowledge via testimony can also be characterized in causal terms in much the same way that inference was. Suppose that the fact that \( P \) causes (perhaps through the causal mechanism of perception) some agent \( T \) to believe that \( P \). Suppose that \( T \)’s belief that \( P \) then causes \( T \) to assert that \( P \) and that \( T \)’s assertion causes another agent \( S \) to come to believe that \( T \) has asserted that \( P \). From this, \( S \) infers (in combination with other background beliefs, such as beliefs about \( T \)’s sincerity) that \( T \) believes that \( P \). \( S \) then infers from this his own belief that \( P \). \( S \) thus has knowledge that \( P \) based on testimony. Smith’s intention to drive to New York on Monday causes him to tell Jones that he will do so and causes him to actually drive to New York on Monday. Jones’ testimonial knowledge is thus related to the fact that makes it true by a common cause: Smith’s intention. As before, if any part of the causal chain is not present, then \( S \) does not know.\(^{30}\)

Knowledge can be obtained by perception, memory, inference, testimony, or any combination thereof, but all can be characterized in causal terms. An important feature of Goldman’s causal account of knowledge and the feature that makes his theory characteristically externalist is that it is not necessary that an agent have conscious access to the fact that a causal connection exists between his belief and the corresponding fact.\(^{31}\) If this causal connection exists, a belief will count as knowledge despite the fact that an agent may be incapable of reflecting on or articulating the fact that it exists.

Perhaps the most famous counterexample to Goldman’s causal theory is the case of Henry and the barn façades.\(^{32}\) Suppose that Henry is driving through the countryside with his

\(^{30}\) Ibid., 362-363.
\(^{31}\) Ibid., 370.
son. It is a bright and clear day. As they are driving, Henry identifies the various structures that come into view. Henry has excellent vision and an unobstructed view of each of the structures and he is driving slowly enough that he has ample time to examine them reasonably carefully. As they pass a barn, Henry asserts ‘That’s a barn’. Moreover, the structure that Henry identifies is in fact a barn. Under these circumstances, we would readily attribute knowledge about the barn to Henry.

But let us modify the case such that, unbeknownst to Henry, the county that they are driving through is riddled with structures that are not in fact barns, but only barn façades cleverly disguised to look like actual barns from the road. Given this additional information, we would not so readily attribute knowledge about the barn to Henry, even if all of the other facts of the case remain the same. Henry seems to have simply gotten lucky in successfully identifying the barn. His belief would have been false if he had made the same assertion about any of the other structures that they encountered. But this additional information does nothing to affect the causal ancestry of the belief and so the causal theorist is still committed to saying that Henry knows that the structure before him is a barn. So it does not appear that an appropriate causal chain between a belief and the fact that makes it true is, after all, sufficient for knowledge.33

This counterexample led Goldman to reject the causal account of knowledge that he had helped to develop, though his later theories of knowledge and justification are in the same vein. Goldman uses the barn façade example as a motivation for introducing his discriminatory theory of knowledge.34 Goldman, however, has since rejected his discriminatory theory in favor of

33 One can preserve the externalist claim that Henry’s belief is justified by interpreting this case as a Gettier-type case. On Goldman’s later process reliabilist theory of justification, Henry’s belief will turn out to be a justified belief that still falls short of knowledge.

34 On Goldman’s discriminatory account, S knows that P only if he can distinguish or discriminate between the actual state of affairs in which P is true and a relevant state of affairs in which P is false. See “Discrimination and Perceptual Knowledge,” 774.
process reliabilism, which will be explored in section III. Goldman’s process reliabilism is not the only species of reliabilism, however. One of the earliest reliabilist accounts of knowledge was D.M. Armstrong’s thermometer model of knowledge.

II. Armstrong’s Thermometer Model of Knowledge

D.M. Armstrong presents his thermometer model of knowledge in his book *Belief, Truth, and Knowledge*. 35 Like Goldman’s causal theory, Armstrong’s theory is an account of knowledge rather than of justification simpliciter. Unlike Goldman, however, he does not frame his theory as a response to the Gettier problem. Instead, he frames his account as a foundationalist response to the infinite regress problem. The infinite regress problem is a skeptical argument regarding epistemic justification. The threat of an infinite regress seems to arise when we consider what it means to have evidence or good reasons for a belief. When one claims to have a justified belief that \( P \), the first question usually asked is what justifies the belief? In response to this question, one might cite as justification some other belief (or some set of beliefs) \( Q \) such that \( Q \) implies \( P \). If the inferential relationship, whether it is inductive or deductive, between the belief that \( P \) and the (set of) belief(s) that \( Q \) is an objectively strong one and not an instance of fallacious reasoning and if \( Q \) is itself justified, then most people would be willing to grant that the agent’s belief that \( P \) is justified. So one’s justified belief that \( P \) is often based on or inferentially supported by some other belief that \( Q \). But, as mentioned, in order for a belief that \( Q \) to provide sufficient support for a belief that \( P \), \( Q \) must itself be justified. Usually, this is on the basis of some other belief that \( R \). But again, before a belief that \( R \) can provide sufficient support for a belief that \( Q \), it must itself be justified, usually on the basis of some other

belief that $S$. And so on ad infinitum. We seem to have an infinite regress of beliefs which require justification. The skeptic concludes that, since we can never reach the end of the regress, no beliefs can ever be justified. Thus, if knowledge requires justification, we cannot have knowledge. The foundationalist responds to this problem by maintaining that there are some beliefs that are non-inferentially justified.\textsuperscript{36}

Armstrong then divides foundationalist responses to the problem into two camps: “initial credibility” theories and “externalist” theories.\textsuperscript{37} He then divides externalist theories into “causal” theories (a la Goldman) and “reliability” theories.\textsuperscript{38} Armstrong rejects the causal version of externalism because, as we saw with Henry and the barn façades, there would seem to be cases in which a belief might be caused by the state of affairs that makes it true, but the belief would not count as knowledge. Armstrong’s account was introduced as a reliabilist alternative to Goldman’s causal theory of knowledge.

On Armstrong’s view, a belief that is produced non-inferentially is a case of knowledge (is justified) if and only if there is “a law-like connection between the state of affairs $Bap$ [$a$ believes that $p$] and the state of affairs that makes ‘$p$’ true such that, given $Bap$, it must be the case that $p$. “\textsuperscript{39} What kind of connections, then, are these law-like connections? Firstly, they are nomological laws, i.e. laws of nature that can, in principle, be investigated by the natural sciences.\textsuperscript{40} Given that they are the sort of connections that can be investigated by the sciences,

\textsuperscript{36} While the internalism/externalism debate often parallels the debate between foundationalists and coherentists, internalists and externalists can be either foundationalists or coherentists. It is for that reason that the solution to the foundationalism/coherentism debate is of no direct consequence to my current project.
\textsuperscript{37} Ibid., 157. Initial credibility theories hold that certain sorts of non-inferential beliefs have an intrinsic claim to credibility. This sort of view is a version of internalism only if the initial credibility of these claims is something that an agent has conscious access to.
\textsuperscript{38} Ibid., 158-159.
\textsuperscript{39} Ibid., 166.
\textsuperscript{40} Even though these laws may be metaphysically contingent, metaphysicists can still speak of nomological necessity. Nomological necessity is a restricted sense of necessity according to which something is necessary if it holds in all possible worlds that satisfy the actual physical laws.
there must be a probability or at least possibility that the conditions can be recreated and the situation repeated. 41 Secondly, they are the sorts of connections that yield counterfactual conditionals: If \( P \) were not the case, then an agent would not believe that \( P \). These sorts of connections are thus reliable indicators of truth. Lastly, the connection is objective in that it holds independently of anyone who may or may not be aware of its existence. 42 An agent can know some proposition whether or not he has conscious access to the fact that a law-like connection exists between his belief and the state of affairs that makes it true. This appeal to reliable law-like connections is what characterizes Armstrong’s theory as an externalist theory of justification.

On Armstrong’s view, non-inferential knowledge (justified belief) is a lot like the correct readings that are produced by a reliable thermometer. When a thermometer is reliable, there is a law-like connection between its reading ‘\( T^\circ \)’ and the temperature of the environment being \( T^\circ \). In this case, the fact that it is \( T^\circ \) in the environment causes the thermometer to read ‘\( T^\circ \)’. Furthermore, if it were not the case that the temperature of the environment were \( T^\circ \), then the reliable thermometer would not read ‘\( T^\circ \)’. Similarly, a non-inferential belief that counts as knowledge exhibits a law-like connection between the belief and the state of affairs that makes it true. If it hadn’t been the case that \( P \), the agent would not have believed \( P \). Unlike the thermometer, however, it is not always the case that the state of affairs causes the belief. It is not generally the case that, if a certain state of affairs obtains, then an agent will believe, as a matter of nomic necessity, that the state of affairs obtains. The agent might, after all, form no belief at all. Rather, Armstrong’s claim is that, if an agent believes that \( P \), then it is nomically necessary that \( P \) obtains.

41 Ibid., 173.
42 Ibid., 168-169.
A thermometer that rarely, if ever, instantiates this law-like connection between the reading and the actual state of affairs might produce the occasional correct reading, but the thermometer would be unreliable and we would not be able to trust the readings that it produced. Similarly, in a case where there is no nomologically necessary connection between a belief and the fact that makes the belief true, these beliefs will be unreliable. Occasionally, one of these beliefs might be true, but they will fall short of knowledge precisely because the belief is an unreliable belief.

Armstrong thinks that his account of non-inferential knowledge can be expanded to capture cases of inferential knowledge as well.\(^\text{43}\) Consider the inference from ‘a is F’ to ‘a is G’. Suppose that an agent knows that a is F. Since the agent knows that a is F, there will be a law-like connection between the belief that a is F and the fact that a is F. In order for the inference from ‘a is F’ to ‘a is G’ to count as an instance of inferential knowledge on Armstrong’s view, (i) the agent must believe that a is G, (ii) as a matter of objective fact, ‘a is F’ must function in the agent’s mind as a conclusive reason for the belief that a is G (though he need not be aware of this), and (iii) the general inferential principle used by the agent must be an objectively truth-conducive one. In this way, the belief that ‘a is F’ causally brings about or sustains the belief that ‘a is G’. And so there is a further law-like connection (the inference) between his belief that ‘a is F’ and his belief that ‘a is G’.

The general inferential principle applied in this example is (x)(if x is F, then x is G), where if… then expresses a nomological necessity. Since the general inferential principle is true, there will be a law-like connection in nature between the fact that something is F and the fact that it is G. Since by hypothesis the agent knows that a is F, it is a fact that a is F. Thus, there

\(^{43}\) Ibid., 206-208.
will therefore be a law-like connection in nature between the fact that \( a \) is \( F \) and the fact that \( a \) is \( G \). So there is a law-like connection at the level of facts and another law-like connection at the level of beliefs. But then the agent’s knowing that \( a \) is \( G \) lies in a further law-like connection that exists between his belief that \( a \) is \( G \) and the fact that \( a \) is \( G \). The inference serves to transfer the reliability of the belief that \( a \) is \( F \) to the belief that \( a \) is \( G \). The law-like connection between the agent’s belief that \( a \) is \( F \) and the fact that \( a \) is \( F \) ensures that a certain law-like connection obtains between the agent’s belief that \( a \) is \( G \) and the fact that \( a \) is \( G \).

Armstrong suggests a modification to his thermometer analogy to accommodate inferential knowledge.\(^\text{44}\) Suppose that you have a thermometer with an attached pressure gauge and the conditions in the environment are such that the temperature is directly correlated with the air pressure. If the temperature is \( T^\circ \), then air pressure is \( P \). Furthermore, the thermometer registering \( T^\circ \) brings it about that the pressure gauge reads \( P \). It might be the case that the thermometer is reading unreliably, and in such a case the pressure gauge would consequently also read unreliably. But if the temperature gauge is reading reliably, then the pressure gauge is reading reliably. The reliability of the gauges nomically ensures the existence of the correlated state of affairs.

Armstrong argues that these law-like connections are both necessary and sufficient for knowledge. In arguing that a law-like connection is necessary for knowledge, he asks us to imagine a case in which some agent \( S \) comes to believe that \( P \) non-inferentially, perhaps through visual perception or memory. Suppose further that the belief is true. But suppose that no law-like connection exists between the belief that \( P \) and the state of affairs \( P \). Surely in such a case we would hesitate to attribute knowledge to the agent. We would hesitate because there is no reason

\(^{44}\) Ibid., 207.
why the belief is likely to be true. The belief just happens to be true. On Armstrong’s view, in order for a belief to count as knowledge, the belief must be a perfectly reliable indicator of truth. It is a matter of nomological necessity that, given \( Bap \), it must be the case that \( P \). If such a law-like connection does obtain, then the existence of the belief guarantees that the state of affairs obtains. It is impossible for the agent to be mistaken.

In arguing that these law-like connections are sufficient for knowledge, Armstrong asks us whether there might be anything else that needs to be added to his analysis to yield knowledge. One suggestion is that the belief needs to be of a certain sort; it needs to be produced in a standard way, e.g. by the standard perceptual mechanisms. Armstrong’s response is that we presumably rely on our ordinary sense perceptions because of their superior truth-gaining capacities. But suppose that our ordinary sense perceptions are a lot less reliable than they are in fact taken to be and suppose that there is some other mechanism for arriving at non-inferential beliefs that was more reliable than our standard perceptual mechanisms. Beliefs produced by such a mechanism would seem to be better candidates for knowledge than beliefs produced by ordinary perception. These beliefs seem to be better candidates precisely because of the stronger and more reliable law-like connection that exists between the belief and the corresponding state of affairs.\(^{45}\)

Armstrong only considers this one suggestion, but there appear to be independent reasons for rejecting Armstrong’s account. On Armstrong’s view, knowledge demands certainty. He holds that there is a conceptual link between knowledge and certainty because, from his point of view, it seems just as paradoxical to say “I know that \( P \) is true, but I am not certain that \( P \) is true,” as it is to say “I am certain that \( P \) is true, but I do not know that \( P \) is true.”\(^{46}\) In order for a

\(^{45}\) Ibid., 190.
\(^{46}\) Ibid., 216.
belief to count as knowledge on his view, the belief must be a perfectly reliable indicator of truth; It must be a matter of nomological necessity that, given \( Bap \), it is the case that \( P \). If such a law-like connection does obtain, then the existence of the belief guarantees that the state of affairs obtains and it is impossible for the agent to be mistaken. This feature of Armstrong’s theory is what allows him to address the Gettier problem.

Truth is of course a necessary condition on knowledge, but most contemporary epistemologists now hold that it is possible to hold a justified but false belief, that justification can be fallible. It is often thought that justification demands only a high degree of probability. As long as it makes sense to interpret Armstrong’s account as regarding the debate about epistemic justification (as opposed to simply denying that knowledge even requires justification), then his view seems to conflict with this fallibilistic tradition. 47 Later versions of reliabilism have relaxed this demand on certainty and allow that an agent may have a justified belief that still falls short of knowledge, although it is worth nothing that this move does open up these later versions of reliabilism to the Gettier problem. With that in mind, I turn now to Goldman’s fallibilistic process reliabilism.

III. Goldman’s Process Reliabilism

Whereas Goldman’s early causal theory was a theory of knowledge rather than justification simpliciter, Goldman explicitly formulates his version of reliabilism as a theory of justification. In developing his process reliabilist theory, Goldman is concerned with the various cognitive processes or mechanisms that produce beliefs. His idea is that whether or not a belief is justified is a function of the reliability of the belief-forming mechanism(s) that produced the

47 Armstrong does argue against relaxing the requirement of certainty, although an examination of this aspect of his view is outside the scope of this thesis.
belief. Unlike Armstrong, however, Goldman does not require perfect reliability. Rather, a reliable belief-forming mechanism is simply one that produces a majority of true beliefs as opposed to erroneous ones in appropriate circumstances. Unreliable belief-forming mechanisms are those mechanisms that are likely to produce erroneous beliefs as or more often than true beliefs. On this view, a belief is justified if it is produced by a reliable belief-forming mechanism, such as vision or deductive inference. A belief is not justified if it is produced by an unreliable belief-forming mechanism, such as wishful thinking or mere guesswork.

A belief-forming mechanism is something that takes a certain input and yields beliefs as output. Belief-forming mechanisms can be either belief-independent or belief-dependent mechanisms. What differentiates the mechanisms is what they take as input. Some types of mechanisms are belief-independent in that they take as their input something other than beliefs. The processes involved in beliefs based on sense perception and introspection would generally be considered to be belief-independent mechanisms in that neither takes other beliefs as input. The standard example of a putatively reliable, belief-independent mechanism is sense perception, and in particular visual perception. If my visual mechanisms produce in me the belief that there is a book on the desk and, if a variety of enabling and defeating conditions have been met (there is adequate lighting, there is an appropriate distance between the object and the observer, the object is an appropriate size, etc.) and, if that mechanism in those circumstances has a tendency to produce true beliefs, then any belief produced by that mechanism in those circumstances is justified.

Other types of mechanisms are belief-dependent in that they can take as input some other

---

49 Ibid., 340-341.
beliefs (though they need not always do so). Memory and logical reasoning would both be examples of belief-dependent mechanisms. The mechanism of memory sometimes takes other beliefs as input (as when one recalls an earlier belief) and sometimes takes sensory or introspective experience as input (as when one recalls what something looked or felt like). Logical reasoning, however, will always take other beliefs as input. A belief-dependent mechanism is said to be conditionally reliable if and only if it tends to produce true beliefs given that the beliefs it takes as input are true. A belief produced by a conditionally reliable belief-dependent mechanism is justified if and only if the input beliefs are justified (independently of whether they are true).\textsuperscript{50} Valid deductive or strong inductive reasoning would be an example of a belief-dependent mechanism that is plausibly taken to be conditionally reliable.\textsuperscript{51}

There is a further question as to the extent of a belief-forming mechanism. Consider the belief-forming mechanism of visual perception. There are multiple ways that we might describe the process that is involved. We could say that it begins with the light rays that bounce off of an object in the environment. The light rays then enter the lens and form an image on the retina. The optic nerve then carries an electrical impulse to the brain, where a conscious experience is produced. Some elements of this process are external to the organism that is the epistemic agent, while other elements are internal to the organism (though not necessarily conscious). Are the events that are external to the organism to be counted among the inputs of the belief-forming mechanism? Goldman’s response to this question is to restrict the extent of the belief-forming process to things that are internal to the organism. He writes: “A justified belief is, roughly speaking, one that results from cognitive operations that are, generally speaking, good or

\textsuperscript{50} Ibid., 341.

\textsuperscript{51} Goldman’s distinction between belief-independent and belief-dependent mechanism in some ways parallels the distinction made in Chapter One between causally inferential and causally non-inferential beliefs.
successful. But ‘cognitive’ operations are most plausibly construed as operations of the cognitive faculties, i.e. ‘information-processing’ equipment internal to the organism.” So the input of the visual mechanism, in this case, includes the light rays that enter through the lens of the eye and form an image on the retina. Although the extent of the mechanism involved is restricted to things that are internal to the organism, the fact that a mechanism is (or is not) reliable is a fact that is (or may be) external to the conscious awareness of the agent. Since the fact of whether or not a belief is produced by a reliable mechanism is external to the agent’s conscious awareness, a natural consequence of process reliabilism is that it is possible for an agent to have a justified belief without thereby having any conscious access to the fact that his belief was reliably produced. This is the rejection of the internalist requirement of conscious access that is characteristic of externalism.

There are a number of ways in which Goldman’s process reliabilism differs from his earlier causal view and from Armstrong’s own version of reliabilism. As noted earlier, both Goldman’s early causal view and Armstrong’s version of reliabilism were both cast as theories of knowledge generally. Goldman’s process reliabilism, in contrast, is a theory of justification simpliciter. The fact that a belief is reliably produced will be sufficient to satisfy the justification condition, but it will not be sufficient for knowledge. This is because the reliability of a belief producing mechanism will not guarantee that any belief produced by that mechanism is true. The fact that a belief is produced by a reliable cognitive mechanism ensures only that the belief is highly likely to be true. This is the key contrast with Armstrong, who required the existence of a reliable law-like connection between a belief and a state of affairs that guarantees that the belief is true.

---

Additionally, Goldman’s early causal view and Armstrong’s own version of reliabilism address (or were originally designed to address) the Gettier problem. It turns out, however, that Goldman’s early causal view falls victim to Gettier cases such as the barn façade example. Recall that the barn façade example is a case where a belief is caused by the state of affairs that makes it true, but intuitively the belief still falls short of knowledge. In that case, Goldman’s causal condition on knowledge is satisfied, but it still seems to be merely a matter of luck that the belief is true. Because process reliabilism allows for fallibilistic justification, Goldman’s version of reliabilism can explain this problem away as a Gettier case. The belief is justified, since it was produced by a reliable belief-forming mechanism (visual perception) in appropriate circumstances (adequate lighting, distance, size of the object, etc.), but the belief still falls short of knowledge because it fails to satisfy some further necessary condition on knowledge (perhaps it is defeated by the fact that the county that they are driving through is riddled with fake barn façade structures).

IV. Conclusion

I take Goldman’s process reliabilism to be the strongest version of externalism in the contemporary debate between internalists and externalists. Hereafter, any mention of the term ‘externalism’ is meant to refer specifically to Goldman’s process reliabilism. In the next chapter, I will explore internalist theories of justification. Chapters four and five will be devoted to analyzing the debate between internalists and externalists.
Internalist Theories of Epistemic Justification

Epistemic internalists, in contrast to externalists, wish to preserve the traditional conception of justification as an internal condition on knowledge, although in doing so they must opt for some fourth condition on knowledge in response to the Gettier problem. The key feature of internalist theories of justification is that justification is a matter of a cognitive agent’s internal mental states. In this chapter, I intend to explore internalist theories of justification in detail. In section I, I will focus on accessibilist accounts of internalism. Section II will be devoted to mentalist versions of internalism. Deontological accounts of internalism will be the subject of the third section.

I. Accessibilism

Perhaps the strongest proponent of an accessibilist version of internalism is Laurence BonJour. On a BonJourian account of justification, an agent’s belief is justified by virtue of his having conscious access to evidence or good reasons that support the belief in question. Where an agent is lacking conscious access to satisfactory evidence in support of a belief, or in the face of strong counter-evidence that is consciously available to the agent, the demands of epistemic rationality and responsibility give him a strong reason for doubting the belief. Epistemic

---

rationality would thus seem to require that an agent form and maintain beliefs in accordance with the evidence to which he has conscious access. The accessibilist holds that there is an intimate link between epistemic justification and epistemic rationality and responsibility. One is justified in believing only what is rational and epistemically responsible for one to believe. In order for a belief to be both rational/epistemically responsible and justified for a particular agent, the accessibilist requires that any factors that justify the belief must be somehow consciously accessible to the agent. A belief is justified for a particular agent if and only if it accords with the evidence that is consciously available to him.

The accessibilist conception of internalism has considerable intuitive force behind it, because it mirrors what happens when questions of justification arise in ordinary circumstances. When someone claims to know something, the question naturally arises: “On what basis can you make that claim?” Of an agent with the capability to question the justification for his own beliefs or the beliefs of others, the accessibilist will insist that it makes no sense for him to hold a belief without some understanding of why (in the evidentiary sense) he holds that belief. Perceptual experience, introspection, memory, logical and probabilistic inferences, and coherence with other background beliefs are all things that an agent might cite as evidence for a belief.

Having conscious access to evidence in favor of a belief is valuable because having such access makes it likely that our beliefs are true. Take, for example, beliefs based on sense experience. Suppose that some agent S comes to believe P that there is a book on the table. The evidence to which S has conscious access is the content of the immediate sense experience.

---

55 These sorts of beliefs are sometimes (though not always) taken to be non-inferentially justified in that they are justified not by means of some inference, but by means of some direct experience. Some internalists hold that the only beliefs that are justified non-inferentially are beliefs about the content of one’s sensory experience.
namely the apparent color, size, and shape of the book. Suppose further that, over time, all of S’s other conscious sense experiences putatively about the book tend to cohere together in an intelligible manner and change in a predictable way. The possession of such evidence (i.e. the conscious experience itself) makes S’s belief that there is a book on the table much more likely to be true than if S did not have conscious access to such evidence. S would not be justified in forming such a perceptual belief if he were blind, for example (though there might be other ways that S could justifiably come to hold a belief that P). The conscious mental states that make up the perceptual experience provide a good reason, and one that the agent obviously has conscious access to, for thinking that the correlated belief is true.

In the case of beliefs explicitly justified via some inference (either deductive or inductive), the premises of the inference function as part of the evidence for the conclusion. Where the premise beliefs are themselves justified in an internalist way and the inference is an objectively truth-conducive one, the premises in effect transfer their justification, via the inference, to the concluding belief. In the case of an explicit inference, the agent has conscious access to the evidence that functions as the premises as well as to the fact that he has drawn his conclusion from those premises.\textsuperscript{56} A belief that is inferred in this way via an objectively truth-conducive inference from other beliefs that are themselves justified in an accessibilist way will be much more likely to be true than a belief that is inferred from an unjustified belief or via an inference that is not truth-conducive. It is clear that one cannot be said to have properly made an explicit inference if one does not have conscious access to the premises used in that inference or the inferential relationship between the premises and the conclusion.

The justifying factors that are consciously available to the agent may take the form either

\textsuperscript{56} Additionally, the agent must have conscious access to the features of the inference that make it an objectively truth-conducive one.
of some direct experience or of some inferential relation between the belief and other beliefs, depending on whether the belief is justified non-inferentially or inferentially. Whether the belief is non-inferentially or inferentially justified, however, the important point is that, on an internalist account of justification, conscious access to the factors that justify the belief is essential to that belief’s being justified.

Just as there are multiple versions of internalism, there are also multiple versions of accessibilism. Some species of accessibilism, such as that endorsed by Roderick Chisholm, hold that, in order for an agent to have a justified belief at $t$, the agent must have conscious access to the factors that justify his belief at that time.\(^57\) If a believer cannot consciously access or put forth his justification for a belief at $t$, then he does not have a justified belief at $t$.

But this requirement seems to be too strict. At any given time, there are a plethora of justified beliefs that we may hold, consciously or otherwise. Given the limits of the human mind, we could not possibly reflect on the justification for every belief that we hold (consciously or otherwise) at the same time, nor even could we actively reflect on the justification of any given belief $P$ at every moment that the belief is held.\(^58\) Moreover, in many cases, we may not even actively reflect on the justification for every belief that we consciously entertain. Most people have a justified belief that the Earth orbits the sun, even though they are not consciously accessing that belief on a daily basis. Even where they are consciously accessing the belief, they may not always be actively reflecting on their justification for that belief. In these cases, a Chisholmian internalist is committed to saying that such beliefs are not justified. But this seems


\(^{58}\) Here it is useful to distinguish between holding a belief and consciously entertaining a belief. A belief that is consciously entertained is sometimes called an occurrent belief. I am consciously entertaining the belief that I am writing my thesis. A belief that is not consciously entertained is sometimes called a dispositional belief. I have a dispositional belief that the Earth orbits the sun even when I am asleep or otherwise not bringing the belief to conscious awareness.
counter-intuitive. Rather, it seems sufficient that one’s justification is consciously accessible, even if one is not actively reflecting on it at $t$. It is plausible that one is justified in believing some proposition at $t$ if one’s evidence for the belief is consciously accessible within some suitable period of time, whether this reflection is short and immediate or prolonged and contemplative, depending on the belief in question.

My preferred account of accessibilist internalism, which is based on BonJour’s, shares the Chisholmian emphasis on conscious access. What differentiates BonJour’s account from the Chisholmian account is that this awareness of or access to the factors that justify a belief need not be achieved at every given moment during which the justified belief is held. What is important is that a believer be able to recall or reflect on the justifying features for his belief. This requires the potential for having conscious access to those features, even if it would take prolonged reflection to bring them to light.

II. Mentalism

In their paper “Internalism Defended,” Richard Feldman and Earl Conee present a version of internalism that they call ‘mentalism.’ They contrast their version of internalism with accessibilist accounts of internalism, according to which the justification of an agent’s belief is determined by things to which the agent has conscious access. On a mentalist account of justification, the justification for an agent’s belief is determined by things that are internal to the agent’s mind. The key difference between accessibilism and mentalism is that not all things that are internal to the agent’s mind will be things to which the agent has conscious access.

Feldman and Conee’s central claim is that “every variety of change that brings about or enhances

justification either internalizes an external fact or makes a purely internal difference. . . . [T]here is no need to appeal to anything extramental to explain any justificatory difference.60 Mentalism is thus committed to two theses. First, the justification of an agent’s belief strongly supervenes on the agent’s occurrent and dispositional mental states, events, and conditions. Secondly, if two agents are exactly alike mentally, then each of their beliefs will be justified to the same extent.61 Feldman and Conee present several examples which are intended to motivate their brand of internalism. I will borrow what I take to be the three most illustrative examples.

In the first case, suppose that two individuals, Bob and Ray, are both sitting in an air-conditioned hotel lobby reading the daily newspaper. As they are reading the weather report for that day, they both read that it will be very warm, and on this basis they both form the belief that it is very warm on that day. Bob then decides to step outside, while Ray continues to read the paper. As Bob leaves the hotel lobby and goes outside, he immediately feels the warmth of the day on his skin. Bob’s belief that it is very warm on that day is better justified than Ray’s belief, because Bob’s conscious experience has internalized an external fact. Ray, by contrast, has only the newspaper report to rely on.62

In the next case, suppose that a logic TA and a student are looking over a logic problem. The TA can easily “see” that some proposition is a logical entailment of some given premise. She is familiar enough with the rules of logic to be able to see what follows from what. She is thus justified in believing that the entailed proposition is true. The student, on the other hand, is completely oblivious and as such would not be justified in believing that the entailed proposition is true. He could, however, come to be justified in holding that belief if he were to come to share

60 Ibid., 410.
61 Ibid, 408.
62 Ibid., 409.
the TA’s mental states, including her web of various background beliefs concerning logical entailment. The difference between the TA and the student is a purely mental one.  

The last case that I will explore is a variation on Gettier’s “Jones owns a Ford” example that was discussed in chapter one. Suppose that Smith has ample evidence to believe that one of his colleagues, Jones, owns a Ford. Jones claims to own a Ford, and Smith has seen Jones driving around in one. From this, Smith deduces that someone in his office owns a Ford. Unbeknownst to Smith, however, Jones is merely pretending to own the car. At the same time, someone else in the office, of whom Smith does not know, does in fact own a Ford. Smith’s belief that someone in the office owns a Ford would be true, but his belief that Jones owns a Ford would be false. This is sufficient, as Gettier demonstrated, to prevent Smith from knowing that someone in the office owns a Ford, although his belief is true and would still appear to be justified. Suppose that at a later time, Smith amasses evidence that Jones is merely pretending to own the car. At this point, Smith would no longer be justified in believing that someone in the office owns a Ford, although in this case the belief would still be true. The epistemic change, Feldman and Conee argue, occurs when a suitable external fact, the fact that Jones is merely pretending to own the Ford, is brought to Smith’s attention. The relevant change is one that is wholly internal to Smith.  

Given these and the other examples presented in Feldman and Conee’s paper, it might be difficult to see any relevant difference between mentalist versions of internalism and accessibilist versions of internalism. In each of the three cases presented above, the accessibilist can plausibly account for the relevant epistemic difference in terms of having conscious access to evidence and reasons for a belief. This should not be surprising, since both theories are versions of internalism...
and much of Feldman and Conee’s paper is dedicated to defending internalism, whether of the mentalist or accessibilist variety, from the various criticisms that have been levied against it. What, then, is the relevant difference between the two theories, and what reason might we have for preferring one account over the other?

Based on the descriptions of the two accounts above, the relevant difference lies in the fact that mentalism is the broader theory. Anything that is consciously accessible to an agent will be something that is internal to the agent’s mind, but not all things that are internal to an agent’s mind are things to which he will have conscious access. Initially, this would seem to allow the theory to judge as justified certain beliefs that an accessibilist account of internalism might have difficulty with. For example, one may plausibly be said to hold any number of beliefs while sleeping and these beliefs are plausibly still justified (if they ever were) even though they are not consciously entertained. On a BonJourian account of accessibilism, these beliefs would not be justified because one would be unable to actively reflect on the justification for such a belief, even when given a reasonable period of time for reflection.

Mentalist accounts, however, can account for the justification of non-occurrence beliefs. The justification for non-occurrence beliefs, such as the beliefs that we hold while sleeping, would be other non-occurrence beliefs and memories. The other beliefs and memories would be mental states, but, while the agent is sleeping, they would not be consciously accessible mental states. I suggest, however that this makes mentalism too broad; so broad, in fact, that it risks collapsing into externalism. If a mental state is capable of conferring justification on an agent’s belief at time $t$ regardless of whether or not the mental state is consciously accessible at $t$, then mentalism becomes characteristically externalist in that it rejects the requirement of conscious accessibility that is so fundamental to accessibilist accounts of internalism.
Moreover, it is not immediately clear how Feldman and Conee’s notion of mental states differs relevantly from Goldman’s notion of cognitive processes. Recall from chapter two that Goldman restricts the extent of a belief-forming cognitive process to things that are internal to the organism itself. Take, for example, an agent’s visual belief that there is a red book on the desk. On a reliabilist account of justification, the belief will be justified if the process that produced it is reliable, even though no part of the process except the resulting belief may be consciously accessible to the agent. Part of the cognitive process involved in a case of visual perception includes the transferring of an electrical impulse, via the optic nerve, from the retina to the brain. These impulses, at least once they reach and are processed by the brain, can plausibly be construed as an unconscious mental state. So there are at least some aspects of the cognitive process involved in a Goldmanian case of, e.g. visual perception, that can plausibly be construed as being internal to the agent’s mind (even if they are not conscious).

In order for the agent’s belief to be justified on an accessibilist view, a conscious experience would need to be produced by this process. But let us suppose that the agent is sleepwalking and that no conscious experience is produced. The accessibilist is committed to saying that such beliefs are not justified. The reliabilist can still say that the belief is justified because it is produced by a reliable cognitive mechanism. What is the mentalist to say about this case? If the mentalist requires that a conscious experience is necessary to justify the belief, then his view seems to collapse into accessibilism. If he insists that the belief is justified despite the lack of a conscious experience, then his view seems to collapse into externalism.

Given this, combined with the fact that mentalism rejects the requirement of conscious access on the part of the epistemic agent, it is difficult to see how mentalism and externalism differ in any relevant way. If Feldman and Conee’s mental states can be understood in externalist
terms, then a question can be raised about why a belief derives its justification from the fact that a mental state is internal to the agent rather than from the fact that it was produced by a reliable cognitive mechanism, if it is. The accessibilist seems to have an answer to this question: the agent has conscious access to the factors that justify a belief. The mentalist, on the other hand, does not.

III. Deontological Conceptions of Internalism

As its name would imply, the deontological conception of justification strongly ties epistemic justification to one’s wider obligations and duties. This would initially seem to be a plausible view to take, as talk of whether or not an agent’s belief is justified often invokes other normative notions: right and wrong, obligations and duties, and the attribution of praise or blame. On a deontological conception of justification, the notion of the justification of a belief is inseparably tied up with these other normative notions. A belief will be justified inasmuch as an agent has fulfilled his epistemic duties.

It is not difficult to find cases in which the justification of a belief seems to be bound up with performing one’s intellectual duties. Suppose that an agent has set up an experiment in an attempt to confirm some hypothesis that he has formed. As the experiment progresses, the agent will amass evidence either for or against the hypothesis that is being tested. It is plausible to suppose that the agent has an intellectual duty to form his beliefs in accordance with the evidence. If the evidence supports the hypothesis, he ought to affirm it. If the evidence does not support the hypothesis, then he ought to reject it. The agent’s belief would seem to be justified inasmuch as he has fulfilled his epistemic duty to form beliefs that accord with the evidence that he has amassed from the experiment. If the agent has fulfilled his epistemic duty, then there would seem to be something praiseworthy about his belief. If, on the other hand, he insists on
believing that his hypothesis is true despite having evidence to the contrary, then there would seem to be something blameworthy about his belief. An agent who insists on holding a belief in the face of cogent counterevidence has violated his epistemic duties, and it is for that reason that his belief is not justified according to the deontologist.

The notion of performing one’s epistemic duties makes no essential reference to one’s internal mental states or what things one has conscious access to. For that reason, one might initially wonder why the deontological conception of justification is often grouped with other internalist theories of justification. The notion of fulfilling one’s epistemic duties can be characterized as an internalist condition on justification by bringing in the Kantian distinction between acting (in this case, forming beliefs) from a sense of duty and acting (forming beliefs) in accordance with duty. On Kant’s view, an agent fulfills his duty only if he acts from a sense of that duty. On a deontological account of justification, an agent fulfills his epistemic duties, and thus has a justified belief, only if he forms his beliefs in accordance with the evidence because he recognizes that it is his duty to do so. If an agent lacks conscious access to the relevant evidence, then he may get lucky and his beliefs may in fact accord with the evidence. However, he will be unable to form any beliefs from a sense of his epistemic duty simply because he will be unable to evaluate the evidence that is before him. If the agent is unable to fulfill his epistemic duties, then he will not have a justified belief on this version of internalism.

The deontological conception of justification is highly problematic, however, and many epistemologists, whether of the internalist or externalist stripe, have had reason to reject it. The strongest criticisms of the deontological conception generally fall into one of two types: it is argued either (a) that doxastic voluntarism is false, or (b) that deontological facts have no bearing.

---

Arguments of the first sort typically begin by noting that a crucial factor in holding an agent responsible for his beliefs is the idea that the agent could have done otherwise. This is the notion of doxastic voluntarism. It is difficult to see how one might be held accountable for forming a belief if one could not have done otherwise. We assign blame to agents that could have done otherwise and withhold it from agents who could not have done otherwise. The deontological conception of justification thus implies that agents have some significant degree of control over their belief formation. It is not clear, however, that we have any significant degree of control over our beliefs. And yet we are still able to subject the agent’s belief to epistemic evaluation.\(^\text{67}\) Take, for example, a normally functioning individual who is standing in a room and comes to believe that the lights in the room are on. Plausibly, the agent does not voluntarily choose to believe that the lights are on, nor for that matter could he voluntarily choose to not believe that the lights are on. In such cases, it seems inappropriate to attribute praise to the agent’s belief, because he has no control over the belief’s formation. The belief simply strikes him, spontaneously and involuntarily. Nevertheless, his belief is capable of being epistemically evaluated and may well be justified. Contrast this with a paranoid man who comes to believe that he is being stalked due to his uncontrollable schizophrenia. In such a case, we might not attribute blame to him for holding that belief. Nevertheless, we would insist that he is not justified in believing that he is being stalked. These two cases together show that beliefs can be subject to epistemic evaluation independently of whether the agent could have done otherwise.

Arguments of the second sort aim to show either that the fulfillment of one’s epistemic


\(^{67}\) The examples that follow are borrowed from Feldman and Conee, “Evidentialism,” 311.
obligations is not sufficient to confer justification on a belief, or that it is not a necessary factor in determining whether or not a belief is justified. Either way, the conclusion of these arguments is that the deontological conception of justification fails to capture some important aspect of our notion of epistemic justification. Cases in which the fulfillment of one’s epistemic duty is not sufficient to confer justification can be illustrated by appeal to what BonJour calls epistemic poverty. These are cases in which the kind of evidence that is available or the method of inquiry is so limited as to make it difficult, if not impossible, to gather together enough evidence or strong enough reasons in favor of a belief. Otherwise educated individuals in the middle ages might be forgiven for believing that the sun orbited the Earth. After all, the Bible affirms it and it is the sun that appears to move while the Earth appears to remain stationary. If we would hesitate from attributing praise to their beliefs, at the very least we would refrain from faulting them for believing as such. Given the evidence that was available to them, they appear to have been fulfilling their epistemic duty to form beliefs in accordance with the evidence. On a deontological view, their beliefs would be justified. Yet it is plausible to hold that the kind of evidence that they had amassed would not be sufficient to make the belief highly likely to be true. Their beliefs do not appear to be justified even though they had fulfilled their epistemic obligations and duties.

Furthermore, epistemologists on both sides of the aisle seem to agree that the fulfillment of one’s epistemic obligations is not even a necessary factor in determining whether or not a belief is justified. Ernest Sosa briefly mentions the example of an agent who forms his beliefs through “uncaring negligence.” Suppose that an agent makes an effort to form his beliefs as

68 BonJour, “Reply to Sosa,” in Epistemic Justification, 176.
69 Sosa, “Skepticism and the Internal/External Divide,” in Epistemic Justification, 148. Laurence BonJour agrees that this is a plausible example. See BonJour, “Reply to Sosa,” 177.
chaotically and randomly as possible. It is clear that the agent has failed to fulfill his epistemic duty to form his beliefs based on the relevant evidence. It is still possible that some small set of the agent’s beliefs might be justified—those beliefs that he forms based on the relevant evidence—however small the set and however disconnected its members might be. This shows that the agent might still hold some justified beliefs despite his rampant failure to fulfill his epistemic duty.

IV. Conclusion

Given the considerations offered above, the most plausible conception of internalism seems to me to be accessibilism. If mentalism is indeed distinct from accessibilism, then the former view seems to risk collapsing into a form of externalism. The deontological view of justification appears to be even less promising, as epistemologists on both sides of the aisle have rejected the view on the basis that it presupposes that agents have some significant degree of control over their beliefs and, moreover, because the fulfillment of one’s epistemic duties appears to be neither necessary nor sufficient for epistemic justification. Ultimately, I intend to adopt and defend an accessibilist account of internalism. Hereafter, any use of the terms ‘internalism’ or ‘internalist’ is meant to refer to accessibilist accounts of internalism.
Arguments

In the previous two chapters, I have articulated the commitments of what I consider to be the strongest versions of both internalism and externalism. I take the strongest version of internalism to be accessibilism. I take the strongest version of externalism to be process reliabilism. If the externalist wants to argue that the justification condition is not an internal condition on knowledge, as it has traditionally been conceived, then a proponent of that view would need to demonstrate that his criterion for justification, reliably produced belief, is both necessary and sufficient to confer epistemic justification on a belief. The internalist can push back at two points. He can deny either (i) that reliably produced belief is sufficient for epistemic justification, or (ii) that reliably produced belief is necessary for epistemic justification. In this chapter, I will present four arguments, each of which engages this question in one form or another. In section I, I will present an externalist argument that affirms (i). In the second section, I will discuss an externalist argument that affirms (ii). In the third section, I will discuss an internalist argument that denies (i). Finally, in section IV, I will present an internalist argument that denies (ii). These four arguments, taken together, seem to yield some fairly powerful yet seemingly conflicting intuitions on both sides of the question. I do not intend to critically evaluate the arguments in this chapter. That will be the subject of the next chapter.
I. The Unreflective Agent

One strength of externalism is that it can easily attribute justified beliefs to certain unreflective agents—namely young children, cognitively disabled adults, or higher-order non-human animals. To the extent that it is plausible to attribute beliefs to these agents, we seem to want to say that certain of their beliefs (e.g. perceptual beliefs) have an important epistemic status. Such a status would seem to be derived from the reliability of the cognitive faculties that they use. These agents are able to reliably discriminate in their responses to objects in the world, and such behavior is plausibly indicative of the agent’s having justified beliefs about such objects. For example, we seem to want to attribute justified beliefs about red objects to the young child who can sort a collection of objects into piles based on whether the object is red or not, in spite of the fact that the child would be incapable of making the kinds of inferences that are required by the internalist. On a purely externalist account of justification, the fact that a belief is produced by a reliable cognitive mechanism is sufficient to confer epistemic justification on that belief. Externalism thus faces no difficulty in attributing justified beliefs to these unreflective agents.

On an internalist view, justification would seem to be a more complicated affair. According to the internalist, in order to have a justified belief, one must be able to reflect on and, in many cases, reason from what one takes to be evidence in favor of a belief. But this is a sophisticated and complex exercise, even for relatively simple perceptual beliefs. On an internalist view, what forms the justification for some perceptual belief is, in part, the conscious experience that one is appeared to in a certain way. Although the conscious experience may

---

directly justify the belief that one is being appeared to in this way, it would not seem to directly justify the belief that there is an object in the external world that is the cause of the experience. In order to justify that belief on an internalist view, one must be justified (in an internalist way) in holding a wide variety of other beliefs, including the belief that one’s conscious experience is caused in the right sort of way by the presence of an object in the external world. If one is justified in holding these other beliefs, then one can then use them as premises in an inference from the experience of being appeared to in a certain way to the existence in the external world of an object that causes that experience. But this requires a sophisticated level of reflection on one’s beliefs and a complicated inferential argument from the existence of the conscious experience to the existence of the external world. Setting aside the question of how one could be internally justified in accepting any of the beliefs that would serve as premises in such an argument, it is highly questionable whether unreflective agents are capable of making, yet alone articulating, such complicated inferences. And so, on an internalist account of justification, such unreflective thinkers will not have justified perceptual beliefs and thus cannot be said to know anything about the external world.

The criticism here is that internalism sets the bar too high for epistemic justification and consequently it runs the risk of entailing a fairly radical kind of skepticism. This worry can be made even more salient if we carry the objection to its logical end. It seems as though even sophisticated thinkers rarely, if ever, make these sorts of inferences when forming and maintaining most ordinary beliefs. Most ordinary people, or for that matter even many philosophers, rarely (if ever) explicitly form the sorts of beliefs or explicitly make the sorts of

71 Laurence BonJour attempts to provide an argument of this sort in *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985), 111-138; see also “The Conceptualization of Sensory Experience and the Problem of the External World,” in *Epistemic Justification*, 77-96.
inferences that are required by the internalist. In most ordinary situations, we make no explicit reference to our reliability as perceivers in forming our perceptual beliefs, nor for that matter do we explicitly appeal to the existence of sense data. It is clear, the externalist argues, that we simply do not routinely (if ever) engage in the sort of reasoning that the internalist requires. Internalism thus leads us into a troubling sort of skepticism whereby few, if any, agents actually have any justified beliefs. BonJour concedes that this objection is intuitively forceful, although he is ready to bite the bullet and claim that epistemic justification is an ideal that is almost never achieved. Since internalism would seem to set the bar too high, particularly in the case of unreflective agents, the powerful intuition at play here is that reliably produced belief is sufficient for epistemic justification.

II. Objectively vs. Subjectively Good Reasons for a Belief

Epistemic justification seems to require that an agent possess objectively good reasons for a belief. There are instances in which an agent might take herself to have very strong reasons for thinking that her belief is true and no reasons for thinking that it is false. But those reasons in favor of a belief might not actually be truth-conducive. In such cases, we might ask whether or not an agent’s belief is really justified. In cases where an agent takes herself to have objectively good reasons for her belief, but those reasons are not in fact objectively good, her belief is not really justified. This seems to suggest that there may be factors external to an agent’s conscious awareness that are necessary for determining whether or not a belief is justified.

Ernest Sosa has put forth an argument of this kind. Sosa considers the case of two

---

72 No one has yet been able to explicitly formulate a justificatory argument for perceptual beliefs on the basis of sense data. Most, if not all, such attempts have been found to be lacking in one way or another.
73 See BonJour, “Externalist Accounts of Justification,” in *Epistemic Justification*, 34-35.
74 Sosa, “Skepticism and the Internal/External Divide,” in *Epistemic Justification*, 151.
logicians, Mary and Jane, who have independently arrived at the same conclusion by way of what they each take to be a valid proof. Both of them have conscious access to evidence, e.g. their respective proofs, that suggests that their conclusion is true. Moreover, neither of them has conscious access to any reason for thinking that she has made an error in her proof. But suppose that Mary’s proof is in fact valid and deduced from premises that she is justified in accepting. Her proof constitutes an objectively good reason for thinking that her belief is true. By contrast, although the axioms that Jane starts with are justified, at least one of the inferences that she uses to draw her conclusion from those premises is, unbeknownst to her, invalid. Jane’s “proof” does not constitute an objectively good reason for thinking that her belief is true.

In this case, there seems to be an important epistemic difference between Mary and Jane. Mary’s proof is flawless, whereas Jane’s is fallacious. Mary’s belief is justified by the fact that her proof is in fact valid. Jane, on the other hand, merely believes (falsely) that her proof is valid. Even though she takes herself to have made good inferences, at least one of her inferences is objectively bad. Consequently, only Mary is really justified. And Jane is not really justified, despite her apparent reason for thinking otherwise. Even though Jane takes her evidence to be truth-conducive, it is not in fact truth-conducive. Externalists seem to be able to account for this epistemic difference. Mary’s belief was produced by a conditionally reliable, belief-dependent mechanism, in this case valid deductive inference. Jane’s belief was produced by an unreliable belief-dependent mechanism, in this case faulty inference. Internalists, however, do not seem to be able to account for this difference. There is nothing that is consciously accessible to either agent that would distinguish the epistemic status of their respective beliefs. Both agents have conscious access to what they take to be evidence in favor of their conclusion and no reason for thinking that their reasoning is mistaken. The internalist seems committed to holding that the
beliefs of Mary and Jane have equal epistemic status.

But clearly, to appeal to the objective validity of an inference is to appeal to a fact that is or may be external to an agent’s conscious awareness. Jane is not aware that there is an invalid step in her “proof”, correct though the conclusion may be. Mary may, though she perhaps need not, be consciously aware that the inferences she uses are objectively valid. The powerful intuition at work here is that, if a belief is not reliably produced, then merely taking oneself to have evidence that one’s belief is likely to be true is not sufficient to confer justification on a belief. Jane’s belief is not reliably produced because it is not the product of an objectively good inference.

This issue need not arise only with respect to cases of inferentially produced beliefs. The following case suggests that a similar problem can arise with respect to non-inferentially produced beliefs. Suppose that there is a speckled hen in an enclosed box. The box is opened for a few seconds and then closed, allowing you only a brief opportunity to study the speckled hen. You are then asked, “How many speckles does the hen have?” Presumably, your conscious visual experience includes a particular number $n$ of speckles on the hen, and you may come to believe that the hen has $n$ speckles. The mechanism of introspection that produces beliefs about the content of visual experiences might well be reliable when the number of speckles on the hen is below some certain threshold, but it plausibly would not be reliable when the number of speckles is above that threshold. In both cases, the internalist seems committed to holding that the content of the visual experience is sufficient to justify the belief that the hen has $n$ speckles, whether the number of speckles is four or ninety-four. However, in the first case we might say that the belief is produced by a reliable belief-independent mechanism, whereas in the second case it is produced by an unreliable belief-independent mechanism (or perhaps that the belief is
produced by a reliable mechanism but in inappropriate circumstances). The externalist can thus account for the justificatory difference between the belief that the hen has four speckles and the belief that the hen has ninety-four speckles.

If the internalist nevertheless insists that Jane’s belief is justified or that the belief that the hen has ninety-four speckles is justified, then he risks allowing that things such as fallacious reasoning (even from justified premises) or unreliable introspection can confer justification on a belief. The internalist risks divorcing justification from truth-conduciveness. As long as we want to preserve an intimate link between justification and truth-conduciveness, we should reject the idea that an agent’s belief can be justified in the absence of being the product of a reliable belief-forming mechanism. And thus, the externalist argues, reliably produced belief is necessary for epistemic justification.

III. Clairvoyance

In section one, we found a case that suggests that conscious access to the factors that justify a belief is unnecessary and that reliably produced belief is sufficient for epistemic justification. However, there seem to be cases in which a belief is reliably produced, but the agent nonetheless seems intuitively to be unjustified in holding that belief.75 This suggests that the fact that a belief is reliably produced is not sufficient to confer justification on a belief. As Bonjour notes, the externalist’s rejection of the requirement of conscious access to the justifying factors of a belief seems to be the Achilles’ heel of externalism.76 He describes the problem for the externalist as follows:

Why, the internalist will ask, should a reason that is outside the conscious grasp of a particular agent nonetheless be taken to confer justification on his belief? Is this

75 BonJour, “A Version of Internalist Foundationalism,” in Epistemic Justification, 28.
76 Ibid., 26.
not indeed contrary to the whole idea of justification, which surely has something to do with selecting one’s beliefs responsibly and critically and above all rationally in relation to the conscious goal of truth? How can the fact that a belief is reliably produced (or indeed any sort of fact that makes a belief likely to be true) make my acceptance of that belief rational and responsible when that fact itself is entirely unavailable to me?77

BonJour asks us to suppose, for the sake of argument, that clairvoyance is a reliable, belief-independent cognitive mechanism in Goldman’s sense. That is to say that beliefs arrived at via this mechanism tend to be true beliefs and that beliefs produced by this mechanism would thereby be justified on an externalist account. BonJour has framed his examples in terms of clairvoyance because, unlike more standard perceptual mechanisms, we do not already have internally available reasons for thinking that clairvoyance, if it exists at all, is a reliable mechanism.

First, suppose we have an agent, named Samantha, who believes herself to have the power of clairvoyance.78 She is in fact a reliable clairvoyant in certain conditions and with respect to a certain subject matter. She has no conscious access to reasons for thinking that she has this ability, nor does she have conscious access to any reasons against the belief that she has this ability. One day, she comes to believe, through this power, that the President is in New York City. On an externalist view, her belief is produced by a reliable mechanism in appropriate circumstances, and therefore her belief would be justified. But suppose further that she has conscious access to very strong counter-evidence to her belief regarding the President’s whereabouts. All of the evidence consciously accessible to her leads her to believe that the President is in fact in Washington, DC. In spite of this, she maintains her belief that the President

77 Ibid., 27. Emphasis in original.
is in New York City. But it seems that Samantha is acting irrationally and irresponsibly in failing to recognize the defeating counter-evidence that is consciously available to her. If there is in fact an intimate link between rationality, epistemic responsibility and justification, then such consciously accessible counter-evidence and Samantha’s failure to heed that counter-evidence is sufficient to rob her of her justification for her belief. For this reason, we should conclude that her belief is not justified.

Suppose next that we have an agent, named Amanda, who, like Samantha before her, believes herself to have the power of clairvoyance. She is in fact a reliable clairvoyant in certain conditions and with respect to a certain subject matter. One day, Amanda comes to believe, via this reliable mechanism, that the President is in New York City. Again, her belief is produced by a reliable mechanism in appropriate circumstances. Though she generally accepts the beliefs that come to her on the basis of this mechanism, she generally does not have any conscious access to evidence in support of those beliefs. Neither does she have conscious access to evidence against these beliefs. She does, however, have conscious access to strong evidence against her belief that she is clairvoyant. It may be that she (incorrectly) believes that her beliefs of this sort are generally false or it may be that she has reason to question whether the power of clairvoyance is even possible. Either way, she has conscious access to reasons for thinking that clairvoyance is not a reliable belief-producing mechanism, if such a power exists at all. In spite of this, she continues to maintain her belief. Amanda is clearly being highly irrational and epistemically irresponsible in maintaining her belief about the President in the face of cogent evidence.

---

79 BonJour, “Externalist Accounts of Justification,” in *Epistemic Justification*, 28-29. BonJour presents two similar cases in “Externalist Theories of Empirical Knowledge,” in *Epistemology: An Anthology*, 368-369; and *The Structure of Empirical Knowledge*, 39-40. As he frames the case of Amanda in *Epistemic Justification*, it includes the case of Samantha discussed in the previous paragraph. However, as I have framed it, the Samantha case is concerned with object-level beliefs whereas the Amanda case is primarily concerned with meta-beliefs. For that reason, I have teased apart the two cases.
evidence that her beliefs produced by her putative clairvoyant powers are unlikely to be true. BonJour holds that, due to the irrationality and epistemic irresponsibility involved, we should conclude that Amanda’s belief is not justified, despite the fact that the belief was produced by a reliable mechanism. Yet the externalist seems committed to saying that Amanda’s belief is justified.

Finally, suppose that Bertha is a reliable clairvoyant in certain conditions and with respect to a certain subject matter.80 As before, the beliefs formed by this mechanism are highly reliable indicators of truth. One day, her clairvoyant powers produce in her, spontaneously and involuntarily, the belief that the President is in New York City. As before, the belief is produced by a reliable belief-forming mechanism in appropriate circumstances. She has no positive evidence in support of the belief regarding the President’s whereabouts, nor does she have any counter-evidence to that belief. Moreover, she does not possess any evidence for or against the existence or reliability of clairvoyance as a belief-producing mechanism. The externalist is committed to saying that any of Bertha’s beliefs produced via the mechanism of clairvoyance that are on the appropriate subject matter and produced in appropriate conditions are justified.

Consider a first variant of this case where, in addition to being clairvoyant, Bertha also believes herself to be clairvoyant and she attributes her acceptance of the belief that the President is in New York City to this power. But Bertha’s belief that she is clairvoyant is irrational and epistemically irresponsible, since she has no reason for thinking that it is true or for thinking that such a power is even possible. If her belief that she is clairvoyant is so clearly irrational, then we ought to say the same thing about any belief that she takes to be a product of this power.81

80 Ibid., 31-32. BonJour models the case of Bertha on one of his earlier cases, the case of Norman, which can be found in his article “Externalist Theories of Empirical Knowledge,” 369-372; and his book The Structure of Empirical Knowledge, 41-42.
Next, consider the variant of the case where Bertha does not believe herself to be clairvoyant. This would seem to be a case that is analogous to the various unreflective agents that were presented in the first section. Just as the unreflective thinker might not believe himself to be a reliable perceiver, neither does Bertha believe herself to be a reliable clairvoyant. There would appear to be no basis for justification other than the externalist one: the fact that a belief was produced by a reliable mechanism. But then it becomes all the more puzzling how Bertha’s belief regarding the President’s whereabouts could be at all rational or epistemically responsible. From Bertha’s perspective, this belief simply popped into her head one day. Its genesis is a complete mystery to her. It seems as though Bertha is being thoroughly irrational and irresponsible in maintaining a belief that is, from her perspective, no different from a stray hunch or guess. The epistemic irrationality involved here is in no way diminished by the fact that, entirely unbeknownst to her, her clairvoyant beliefs are reliably caused. Yet the externalist seems committed to the seemingly counterintuitive position that Bertha’s belief is justified.

Thus, the externalist who insists that the clairvoyant beliefs of Samantha, Amanda, and Bertha are nevertheless epistemically justified must apparently sever the link between rationality and epistemic responsibility and justification. The cases of Samantha, Amanda, and Bertha suggest that reliably produced belief is not sufficient for justification, and this supports the internalist’s emphasis on conscious access.

This argument presents us with an agent, Bertha, who initially seems to be analogous to

---

82 BonJour, “Externalist Accounts of Justification,” in *Epistemic Justification*, 29.
83 Ibid., 32.
84 In response to the first two cases, an externalist might modify his position to require that an agent not possess any counter-evidence to the belief in question or to the reliability of the mechanism that produced the belief. But the Bertha case suggests that it is not enough that an agent possess no cogent counter-evidence. In response to this last case, one possible move is to simply bite the bullet and reject the idea that justification is tightly connected to epistemic rationality and responsibility. This is precisely the line that Alvin Goldman takes. See Alvin Goldman, “Internalism Exposed,” in *Epistemology: An Anthology*, 379-393.
the unreflective agents that were discussed in section one. But there also seems to be an important difference between them. Bertha’s irrationality stems from the fact that she is a fully functioning rational agent who, in most cases even if not in this one, generally engages in sophisticated and complex reasoning with respect to what beliefs she ought to hold. Young children, cognitively disabled adults, and non-human animals, in contrast, often do not and generally cannot engage in this sort of sophisticated reflection. Since Bertha can engage in this sort of reflective reasoning, she plausibly ought to do so. Since unreflective agents cannot engage in this sort of reasoning, there does not appear to be any sense in which they ought to do so. As a result, Bertha’s belief appears to be irrational and epistemically irresponsible in a way that the beliefs of the unreflective agents may not be. Bertha’s belief is thus a result of an epistemic vice, whereas the belief of the unreflective agent does not seem to be the product of an epistemic vice.

IV. The Brain in a Vat

In section II, I discussed a case that suggests that reliably produced belief is necessary for epistemic justification. In this section, I will discuss a case that challenges this idea. In arguing that reliably produced belief is not necessary, it would need to be shown that an agent can nonetheless be epistemically justified in holding a belief that is produced by an unreliable mechanism. For example, in a world in which we are merely disembodied brains in vats, the suggestion is that an agent can still be epistemically justified in holding a belief about the external world, even though the processes that give rise to this belief are unreliable.

Suppose that, unbeknownst to you, you are merely a disembodied brain in a vat. Your brain is wired to a very advanced computer that is capable of perfectly simulating the sensory experiences you would have if you were embodied. The computer simply sends an electrical signal to your brain, thereby producing the conscious experience of, and consequently a belief
about, say, a red book on the desk in front of you. It can readily be conceded that brains in vats generally do not have knowledge about the external world. Any internalist should be perfectly willing to admit that many of the beliefs of a brain in a vat fall short of knowledge, simply because many of the beliefs of an evatted brain will not be true. There may be no book. There may be no desk. But, it may be argued that such beliefs are still justified.\textsuperscript{85} There is nothing implausible or incoherent about a justified belief that falls short of knowledge simply because it is not true, and it seems as though the beliefs of a brain in a vat may be of this kind.

If we were brains in vats, then our conscious perceptual experiences would not be reliable indicators of truth. On this point, both internalists and externalists would agree. But what the brain in a vat argument can suggest is that there can nevertheless be something epistemically valuable about an agent’s beliefs in spite of the fact that his belief-producing mechanisms are not reliable. By hypothesis, the perceptual experiences of an evatted brain are just the same as those that are caused in an embodied reliable perceiver of the external world. Even if an agent’s state is caused by the masterful simulation of an advanced computer, he would have the exact same reasons for thinking that his beliefs are true, no more and no less, as we do (assuming that we are not brains in vats). And since we do have good reason to believe that at least some of our perceptual mechanisms are reliable, then if we are in fact merely evatted brains, we would have the exact same reasons to believe that those same perceptual mechanisms were in fact reliable, regardless of the fact that, unbeknownst to us, they were not.\textsuperscript{86} Since, if we were evatted brains, our perceptual mechanisms would not in fact be reliable, an externalist is committed to holding that few (if any) of our own empirical beliefs about the world would ever be justified in such a


\textsuperscript{86} Cohen, “Justification and Truth,” 281.
scenario. This, however, seems counter-intuitive. A brain in a vat might still use impeccable reasoning from the conscious experience of a red book on the desk to the belief that there is a red book on the desk. This suggests that the beliefs of the brain in a vat can still have some sort of epistemic value. Although the use of such reasoning would not be a reliable indicator of truth in this case, it nevertheless seems to be valuable because the agent is not acting irrationally or irresponsibly in any obvious way. Such beliefs may seem to be justified, the internalist may insist, even if they do not count as instances of knowledge. The agent not only has conscious access to evidence that suggests that his belief is true and forms his belief in accordance with that evidence, but he also uses reasoning that, were he normally embodied, would be a reliable indicator of truth.

Another problem presented by the brain in a vat hypothesis is that, in such a scenario, externalism has difficulty accounting for the relative epistemic value of different belief-forming mechanisms. Suppose that we have two brains in vats, whom we will call Smith and Jones. Suppose that Smith and Jones both believe that there is a red book on the desk. Smith forms his belief in part because he has the conscious visual experience of a red book on the desk. We may suppose that he reasons impeccably from a belief about the content of his conscious visual experience and a belief that the conscious visual experience is likely caused by there being a red book on the desk to his belief about the book. Jones, on the other hand, forms his beliefs quite independently of his visual experience based on wishful thinking (perhaps Jones wants the book to be on the desk because he needs to return it to a friend). Since both mechanisms will tend to yield false beliefs in an envatted brain scenario, and are thus unreliable, the externalist is committed to rendering identical judgments with regard to the justification of both beliefs.

---

87 Ibid., 283.
But there seems to be a clear epistemic difference between the beliefs of Smith and the beliefs of Jones. Smith’s belief is rational, whereas Jones’ belief is not. Had Smith been a regularly embodied agent, his belief that his conscious visual experience was likely caused by there being a red book on the desk plausibly would have resulted from truth-conducive reasoning. Neither is Smith being irrational or epistemically irresponsible in believing as he does. Jones’ belief, on the other hand, is both irrational and irresponsible regardless of whether the agent was envatted or embodied. Thus the externalist is again forced to sever epistemic justification from rationality and epistemic responsibility. In the clairvoyance case, the externalist seemed committed to holding that Bertha’s belief was justified despite her rampant irrationality and epistemic irresponsibility. In this case, the externalist seems committed to holding that the beliefs of a brain in a vat would be unjustified despite being wholly rational and epistemically responsible. And so the externalist is forced to say, counter-intuitively, that rationality and epistemic responsibility are neither necessary nor sufficient for epistemic justification.

V. Conclusion

The analysis so far has left us with conflicting intuitions. The case of the unreflective agent suggests that reliably produced belief is sufficient for epistemic justification and that conscious access to evidence is unnecessary. Sosa’s case of the two logicians, Mary and Jane, suggests that reliably produced belief is necessary and that conscious access to what one takes as evidence is not sufficient to confer justification on a belief. BonJour’s clairvoyance cases suggest that reliably produced belief is not sufficient for justification and that conscious access to the factors that justify a belief is necessary. The brain in a vat scenario suggests that reliably produced belief is not necessary and that conscious access to evidence is sufficient for
justification. We would appear to be at a standoff, with no clear conclusion one way or the other. In the next chapter, I will attempt to offer a sort of hybrid theory in response to these cases. However, I also intend to argue that there are other considerations that tip the balance in favor of internalist theories of epistemic justification.
The Indispensability of Conscious Access

The four arguments presented in the previous chapter, taken together, seem to leave us with contradictory intuitions. The unreflective agents argument suggests that reliably produced belief is sufficient for epistemic justification and that conscious access to evidence in support of a belief is unnecessary. BonJour’s clairvoyance cases, on the other hand, suggest that reliably produced belief is not sufficient for epistemic justification, and that conscious access to the factors that justify a belief is necessary. Sosa’s case of the two logicians, and the related case of the belief about the speckled hen, suggest that reliably produced belief is necessary and that conscious access to (what one takes as) evidence is not sufficient to confer justification on a belief. The brain in a vat scenario suggests that reliably produced belief is not necessary for epistemic justification and that conscious access to evidence in support of a belief is sufficient. Whether one finds the unreflective agents argument to be convincing will largely depend on one’s response to BonJour’s case of Bertha the clairvoyant and vice versa. Similarly, whether one finds the brain in a vat argument convincing will largely depend on one’s response to Sosa’s case of the two logicians and the related case of the speckled hen and vice versa.

In what follows, I will first explore what an internalist might say in response to the arguments that seem to favor externalism. However, I will not conclude that the internalist has a decisive response to the externalist or that the intuitions that favor internalism decisively outweigh the intuitions that favor externalism. This leads me to believe that some sort of hybrid theory might be the best theory. However, even if that is the case, I will attempt to argue that
internalism is more fundamental in one very important way by expanding upon a suggestion that Laurence BonJour has briefly sketched.

I. Internalist Responses to the Externalist

What might the internalist say in response to the case of the unreflective agent? Recall that the case of the unreflective agent suggests that internalism sets the bar too high for epistemic justification and entails that few, if any, agents ever have justified beliefs. In response to the unreflective agent argument, the most plausible response open to the internalist seems to be to simply bite the bullet and concede that young children and other unreflective agents do not have justified beliefs. The internalist might say that we treat such unreflective agents as having justified beliefs, but none of their beliefs are, strictly speaking, actually justified. Recall that BonJour’s case of Bertha the clairvoyant suggests that reliably produced belief is not sufficient for epistemic justification. In that case, we had an agent who has a reliably produced belief and no conscious access to reasons either for or against the belief in question or for or against the reliability of the mechanism that produced it. Bertha seems in many ways to be analogous to the unreflective thinker who has a reliably produced beliefs but is unable to actively reflect on any reasons for his beliefs. If Bertha is indeed analogous to these unreflective agents, then we ought to draw the same conclusion about the beliefs of these unreflective agents as we did about Bertha’s beliefs.

As noted in chapter four, however, there is one point at which the analogy seems to break down. Bertha is a fully rational epistemic agent who is capable of actively reflecting on the justification for her beliefs. She merely fails to do so in this case. The unreflective agent, in contrast, is unable to engage in such reflection. Since they are unable to engage in such reflection, it seems inappropriate to fault them for failing to do so. But even if Bertha is open to
the charge of irrationality and epistemic irresponsibility in a way that these unreflective thinkers are not, none of them are exercising rationality or epistemic responsibility in holding their respective beliefs. And so the internalist will want to insist that rationality and epistemic responsibility is a necessary feature of epistemic justification. The internalist can acknowledge that there may be epistemic value in having a reliably produced belief and so he need not deny that the beliefs of the unreflective agent lack any value whatsoever. The internalist will, however, want to insist that such epistemic value is insufficient for epistemic justification.

As Bonjour notes, if we adopt this approach, then the internalist notion of justification at play clearly becomes stricter and more idealized than the externalist notion. The vast majority of the beliefs of ordinary human agents are formed without the sort of reflection that would be required by the internalist. Furthermore, it is not even clear what would count as an internallistically acceptable reason in favor of basic perceptual beliefs. If the internalist is correct about epistemic justification, then few (if any) of the beliefs of ordinary agents will count as justified. Epistemic justification on the internalist model is thus an ideal that is rarely, if ever, obtained. But there appears to be no reason, BonJour insists, why we should regard our ordinary, intuitive notions of justification to be so ironclad that we ought to reject this idealized standard out of hand.88 One of the purposes of epistemology is, after all, to devise a more accurate account of the various conditions for knowledge, in this case justification, because it is recognized that our ordinary, commonsense accounts are overly vague. If, through the dialectical process of philosophical debate, we come to the conclusion that our ordinary working notions of justification are not sufficiently precise, then so much the worse for our ordinary working notions of justification.

The externalist will likely object to this response on the grounds that the skeptical price that the internalist is willing to pay is too high. The externalist will want to insist that ordinary agents are often, and perhaps even usually, justified in their perceptual beliefs about the world, even though they rarely if ever engage in the sort of active reflection required by the internalist. When an agent’s belief, e.g. that there is a red book on the desk, is caused by the reliable functioning of the agent’s visual mechanism in appropriate circumstances, the agent is fully justified in her belief that there is a red book on the table. If the externalist is correct about the justification for the perceptual beliefs of ordinary agents, then even young children, non-human animals, and cognitively impaired adults may have justified perceptual beliefs. Further, there appears to be little or no reason to hold that Bertha’s clairvoyant beliefs are unjustified. The upshot of this is that both the internalist and the externalist have a way of responding to these sorts of cases and so the arguments that appeal to these cases seem to be indecisive.

How might an internalist respond to the argument that reliably produced belief is necessary for epistemic justification? This argument appeals to the distinction between objectively good reasons and subjectively good reasons for a belief. There seem to be instances in which an agent might take herself to have conscious access to good reasons in favor of a belief, but that evidence might not in fact be truth conducive. Recall that in the case of the two logicians, the inferences that Mary uses in her proof are objectively good inferences, whereas Jane merely takes herself to have made good inferences. The powerful intuition at work here is that Mary is justified while Jane is not and this is due to the fact that only Mary’s inferences are objectively good, while Jane’s inferences are not. The speckled hen case raises similar worries with regard to the reliability of our beliefs about the content of our conscious sensory experiences. That case suggests that having conscious access to an experience of being appeared
to x-ly can serve to justify a belief about the content of that conscious experiences only if one is able to reliably assess that content. These cases together suggest that reliably produced belief is at least necessary for epistemic justification. No internalist will want to insist that faulty reasoning or faulty assessment of conscious experience (even if an agent takes that reasoning or assessment to be objectively good) should confer justification on a belief. To avoid this, the internalist must first of all insist that an agent’s evidence must in fact be truth-conducive. But the internalist must also require that an agent have conscious access to the features of the inference that make it an objectively truth-conducive one. Otherwise, the internalist risks allowing an external factor to make a decisive contribution in determining whether a belief is justified. While it may be granted that many epistemic agents do in fact make use of objectively good reasons for their beliefs, it is questionable whether they actually have conscious access to those features of a conscious experience or inference that make it truth-conducive when forming many of their beliefs. It not clear that the internalist has an adequate response to this problem.

But the internalist will insist that the intuition appealed to in the previous case conflicts with the internalist intuition that an envatted brain might be justified in holding at least some perceptual beliefs about the world despite the fact that none of those beliefs are reliably produced. There seems to be a clear epistemic difference between Smith the envatted brain who uses impeccable reasoning from a belief about the content of his sense experience and a belief that his experience was caused in the right sort of way to a belief about a red book on the table in the external world and Jones the envatted brain who uses fallacious reasoning to support a belief with the same content. The internalist will want to claim that, since the externalist is committed to rendering identical judgments with regard to the justification of both beliefs, externalism cannot account for the epistemic difference between Smith and Jones. But there does appear to
be a way that the externalist can respond to this case. The externalist can maintain that neither Smith’s nor Jones’ respective beliefs are justified, but perhaps Smith’s belief is epistemically virtuous in a way that Jones’ belief is not. Smith’s inference from his belief about the content of his visual experience and his belief that that experience was likely caused by there being a red book on the table is truth-conducive, but one of his premise beliefs is not justified. In Goldman’s terms, the belief producing mechanism that Smith uses is conditionally reliable. What goes wrong with Smith’s belief is that one of the input beliefs, namely his belief that his visual experience was likely caused in the right sort of way by the presence of a red book on the table, is not itself reliably caused. It is because one of the input beliefs is not itself reliably caused that Smith is not justified in believing that there is a red book on the table. Jones, on the other hand, uses a conditionally unreliable mechanism (wishful thinking). Jones’ belief is not justified because the mechanism used is simply unreliable. The externalist can explain the extra epistemic value of Smith’s belief on the grounds that Smith but not Jones uses a conditionally reliable inference based on premises that are accepted as true. As before, both the internalist and externalist have ways of responding to these cases and so the arguments that appeal to them seem to be ultimately indecisive.

II. Two Perspectives for Assessing Epistemic Justification

It would seem that the internalist and the externalist both have responses to the arguments presented in chapter four against their respective views. Whether one finds any of these responses plausible will largely depend on one’s prior intuitions about the cases. BonJour suggests that space can be created for these conflicting intuitions if we allow ourselves to make a distinction between the first-person perspective of the believing agent and the perspective of a third-person assessor. Whenever we ask whether or not a belief is justified, there are actually two
questions that we might ask. One question is whether or not an agent himself has any reason for thinking that any one of his beliefs is true. Such a question is relative to the agent’s own first-person perspective. Alternatively, we may inquire into the justification of an agent’s belief from the perspective of some third-person assessor, as when we ask whether there is a reason why an agent’s belief is likely to be true more generally.

When an agent reflects on the justification for his own beliefs from his own first-person perspective, there appears to be no justifying feature to which he can appeal that is not at least potentially (given an appropriate period of time for reflection) within his conscious grasp. The internalist’s emphasis on conscious access to the justifying features of a belief is ultimately rooted in this first-person perspective. Traditionally, this first-person perspective has been the implicit, if not explicit, concern of many epistemologists. This was the perspective that Descartes, for example, was explicitly concerned with when he engaged in his method of radical doubt. Since it seems reasonable to regard the first-person perspective as a legitimate one, then there is reason to think that there is a role for conscious access to play in questions of epistemic justification. When our concern is this particular, first-person perspective, it would seem that the internalist conception of justification is most appropriate.

When operating from the third-person perspective of some assessor, it may be the case that factors external to the agent’s own conscious awareness play some part in our determining that a belief is justified. If we take the beliefs of an unreflective agent or Bertha the clairvoyant to be justified, it seems to be by virtue of the fact that their beliefs are reliably produced. If we take Jane’s belief and the belief of Smith the envatted brain to be unjustified, it seems to be by virtue of the fact that their beliefs are not reliably produced. Thus, it may be the case that externalism is better able to account for our intuitions regarding our assessments of the
justification of other epistemic agents.

If we allow for a distinction between first-person questions of justification and third-person questions of justification, then there would seem to be some space opened up for both internalist and externalist conceptions of justification with regard to their respective questions about justification. When considering specific cases, we are urged to keep in mind the perspective from which we are engaging in our epistemic projects, for it is this perspective that determines whether we should adopt an internalist approach or may take an externalist approach.

Since it is reasonable to consider both the first-person perspective and the third-person perspective to yield equally legitimate, interesting, and informative answers with regard to questions of justification, BonJour suggests that one way to resolve the tension between internalism and externalism is to realize that they are not competing conceptions of justification; they are “simply different and incommensurable” conceptions, focused on different aspects of justification. The apparent tension between internalism and externalism seems to be a consequence of failing to keep distinct the first-person perspective and the third-person perspective. The suggestion is that the mistake made by proponents on both sides of the debate is in thinking that they are offering competing answers to one and the same question.

If this approach is correct, then it would seem plausible that an agent might be externally justified in holding a belief, but not internally justified (and vice versa). In response to the four arguments presented in the previous chapter, we could say that, if the agent himself or herself does not have conscious access to objectively good evidence for a belief, then he or she is not internally justified with respect to this first-person question. However, he may still be externally justified in holding a belief.

---

justified with respect to the third-person question. And so we can say, with regard to various unreflective agents, that they are externally justified with respect to the third-person question, but that they are not internally justified with respect to the first-person question. With regard to Bertha, the internalist can affirm that her belief about the President’s whereabouts is not justified from a first-person perspective because she lacks conscious access to evidence and good reasons for the belief. But there would seem to be nothing wrong with some hypothetical third-party assessor who attributes to Bertha an externally justified belief on the basis of the reliability over time of her clairvoyant powers. It might be the case that the agent’s belief is internally justified, but not externally justified.90 We can affirm that the perceptual beliefs of a brain in a vat may be justified from the agent’s own perspective, even if they may not be justified from the perspective of some external assessor.

It is not clear, however, that appealing to this distinction will account for the epistemic difference between Mary and Jane. BonJour’s distinction allows us to say that Mary is justified from both the first-person and third-person perspectives. She has conscious access to objectively good reasons for her belief, and the belief is produced by what is in fact a reliable cognitive mechanism (valid inference). This is a plausible response. But it is not as if Jane is justified from the first-person perspective but not justified from the third person perspective. She does not even seem to be justified from the first-person perspective. In Jane’s case, unless we want to collapse the distinction between objectively good and subjectively good reasons for a belief, it seems that appealing to conscious access to what Jane merely takes to be evidence (faulty as it is) will not be sufficient to confer epistemic justification on a belief, even from the first-person perspective.

90 Stewart Cohen has put forth an argument of this sort, though he does not explicitly frame it in these terms. See Stewart Cohen, “Justification and Truth,” Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 45, no. 3 (1984), 279-295.
It seems to me that the most plausible response to this worry is to hold that reliably
produced belief and conscious access to at least some of the features that justify a belief are
independently necessary and jointly sufficient for first-person epistemic justification. The case of
Jane plausibly shows that reliably produced belief is at least necessary for epistemic justification.
But the Bertha case plausibly shows that reliably produced belief is not by itself sufficient for
first-person epistemic justification. An agent must also have conscious access to evidence in
favor of the belief. A purely internalist theory of first-person justification cannot require that an
agent’s evidence in fact be truth-conducive (as opposed to requiring merely that the agent takes
the evidence to be truth-conducive), as this is to appeal to an external factor that is or may be
outside the agent’s conscious awareness. Consequently, a purely internalist theory of first-person
epistemic justification will have difficulty distinguishing between objectively good reasons for a
belief and subjectively good reasons for a belief at the first-person level. Thus, first-person
epistemic justification seems to require that an agent have conscious access to evidence in
support of the belief and that evidence must in fact be truth-conducive. Where the evidence that
produces or sustains a belief is in fact truth-conducive, the mechanism that produces belief will be a reliable mechanism.

This approach would allow us to affirm that Mary’s belief is justified, while denying that
Jane’s belief is justified because Jane’s belief is not reliably produced. We can also affirm that
the beliefs of an envatted brain might be epistemically valuable from the first-person perspective
in those cases where an agent has conscious access to what would be considered reliable
evidence for a regularly embodied agent, although the beliefs of an envatted brain might not after
all be first-person justified because the belief is not in fact reliably produced. Conversely, the
perceptual beliefs of the unreflective agent might be epistemically valuable because those beliefs
are reliably produced, although such beliefs will not be justified because such agents lack conscious access to the factors that justify those beliefs.

III. The Indispensability of Conscious Access

While I agree with BonJour that conscious access to what one takes to be good reasons in support a belief is necessary for justification from the first-person perspective, there are reasons to think that it is not sufficient. And so a purely internalist theory of justification may not be adequate to handle even first-person questions of justification. When it comes to third-person questions of justification, it may not be necessary that an agent have conscious access to what he takes to be good reasons for a belief and it may be the case that reliably produced belief is by itself sufficient for third-person epistemic justification. However, following an idea briefly suggested by BonJour,91 I will argue that the internalist criteria of conscious access nevertheless plays a fundamental role in motivating any claim that externalist theories of justification are adequate to handle third-person questions of justification. The idea is that externalism can only be motivated as the sort of justification that is both necessary and sufficient for third-person questions of justification by invoking the first-person perspective of a third-party assessor. The appeal of the externalist position is due to the fact that we, from our third-person perspective, have conscious access to information that the agent in question does not have access to. Consider the externalist arguments in the previous chapter or indeed any of the hypothetical cases present in the published literature that have been used to motivate externalism. By virtue of being the ones stipulating the facts of the case, we as the assessors are in an ideal position to judge that an

91 BonJour, “A Version of Internalist Foundationalism,” 39. See also, “Externalist Theories of Empirical Knowledge,” in Epistemology: An Anthology, eds. Ernest Sosa, Jaegwon Kim, Jeremy Fantl, and Matthew McGrath (Malden, MA: Blackwell Publishing, 2008), 371. In both places, however, he devotes only a paragraph to explaining this argument. It what follows, I will attempt to draw out and explain, in a more detailed manner, the point that I take BonJour to be making.
agent’s belief is reliably produced. Even in ordinary, real-world contexts, where our third-person perspective might be less than ideal, we are still likely to have conscious access to evidence that suggests that an agent’s belief is reliably produced. What drives the intuition that agents in such cases are justified is precisely the fact that we as the assessors have conscious access to relevant information, either because such information is stipulated as part of the thought experiment or because we have amassed evidence of our own.

How are we to understand the externalist conception of justification, and the third-person epistemic project for which it may be most suited, as inherently involving the first-person perspective? Suppose that an agent $S$ comes to believe that $P$ through some reliable cognitive mechanism such as sense perception. Even where $S$ lacks any conscious access to evidence that suggests that the belief is reliably produced, some external assessor $R$ may still judge that $S$ is justified in his belief that $P$ on the basis of $R$’s belief that $S$’s belief that $P$ is produced by a reliable cognitive mechanism. Let us call $R$’s belief here the belief that $Q$. It seems to be highly implausible that $R$ would hold the belief that $Q$ if $R$ lacked conscious access to reasons (objectively good or otherwise) that suggest that $S$’s belief that $P$ was reliably produced. If $R$ did not have conscious access to such evidence, then $R$ would not be inclined to judge that $S$’s belief is reliably produced. Moreover, if $R$ lacks conscious access to such evidence, then it would be epistemically inappropriate and irresponsible for $R$ to form her own belief about $S$’s belief that $P$ on the basis of $S$’s reliability. It seems that in this case the assessor could no better make an accurate determination about $S$’s justification than $S$ himself could.

Moreover, if $R$’s belief that $Q$ is justified in an internalist way, it would be entirely appropriate for $R$ to cite the belief that $Q$ in justifying some further belief of $R$’s own, namely her own belief that $P$. By attributing a reliably produced belief to $S$, $R$ is in effect treating $S$ as a
reliable cognitive instrument. Assessments of another agent’s reliability can thus serve as premises for \( R \)’s own inferences. The assessor \( R \) is making an inference from \( S \)’s belief that \( P \) as well as \( S \)’s reliability in this circumstance to her own belief that \( P \). In the cases purported to support externalism, it is the assessor \( R \), rather than the agent \( S \) himself, who has the conscious access to evidence and good reasons that is appealed to by the internalist, and this is why the assessor is able to make a determination about whether or not the agent’s belief is justified, even if the agent himself cannot.

It is the third-person perspective on the justification for \( S \)’s belief that \( P \) that allows for a purely externalist take on epistemic justification. But this third-person perspective seems to presuppose a first-person perspective on the assessor’s own justification for the belief that \( S \)’s belief that \( P \) is reliably produced. As we saw earlier, conscious access plays a necessary role in first-person perspectives on justification. As such, a purely externalist justification would seem to be a derivative sort of justification. It is derivative because those cases where it is plausible to hold that the fact that an agent’s belief is reliably produced is sufficient to confer justification on that belief are nevertheless cases in which some assessor has conscious access to what she takes to be evidence or good reasons for the fact that the belief is reliably produced. It is by virtue of having conscious access to reasons and evidence that a third-person assessor is able to attribute justification to an agent.

One reason that externalism is initially plausible is that the externalist can say that, if the external relation appealed to actually obtains, then the belief is justified. There need not be any third-person assessor who attributes justification in order for a belief to be externally justified. A belief can be externally justified even when an agent is alone in the world. But then no question of justification has arisen. And as BonJour notes, “the question of whether either the agent or
anyone else has any good reason to think that the [external relation actually obtains] can in the end only be dealt with in an internalist way." In any case where a question arises about the justification for S’s belief that P, no answer can be given unless some agent (either S or the third-person assessor) has conscious access to putative reasons for thinking that S’s belief is likely to be true. It may be the case that a belief is justified from the third-person perspective by features that are entirely external to the agent’s own conscious awareness, but then any third-party assessor who wants to assert that S’s belief that P is a justified belief must be able to put forth reasons for thinking that the belief is likely to be true, reasons that she, rather than the agent, has conscious access to. And this presupposes a first-person question of justification, and thus internalism. It is in this sense that conscious access is a fundamental and indispensable aspect of epistemic justification.

IV. Conclusion

In this thesis, I began by examining the criterion for knowledge and focusing specifically on epistemic justification. In the second chapter, I explored three versions of epistemic externalism and concluded that the most plausible version of externalism was process reliabilism. In the third chapter, I explored three versions of epistemic internalism and concluded that the most plausible version was accessibilism. In chapter four, I explored four arguments that result in seemingly conflicting intuitions about the roles of both reliably produced belief and conscious access to the features that justify a belief in epistemic justification. In this final chapter, I have argued that the most plausible response to these arguments is to hold that reliably produced belief and conscious access to at least some of the features that justify a belief are

---

92 BonJour, “Reply to Sosa,” in Epistemic Justification, 183.
independently necessary and jointly sufficient for epistemic justification from the first-person perspective. It may be the case that reliably produced belief is by itself sufficient for third-person questions of justification. But these third-person assessments of justification presuppose first-person assessments, and so conscious access to at least some of the features that justify a belief (on the part of the assessor) is fundamental to these third-person assessments as well.
Bibliography


