INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
AN ACCOUNT OF EPISTEMIC GOODNESS:
AN ALTERNATIVE TO WARRANT AND JUSTIFICATION

DISSERTATION

Presented in Partial Fulfillment of the Requirement for
The Degree of Philosophy in the Graduate
School of The Ohio State University

By

Haewan Lee, B.A., M.A.

The Ohio State University
2002

Dissertation Committee:
Professor Marshall Swain, Adviser
Professor George Pappas
Professor Diana Raffman

Approved by

M. Marshall Swain
Adviser

Department of Philosophy

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ABSTRACT

What is the property that makes knowledge when added to a true belief? This is one of the fundamental questions in epistemology. In this dissertation, I explore this property. Traditionally, epistemic justification, and, after Gettier, indefeasible epistemic justification has been considered as such a property. However, the discussions in the last four decades show that it is not at all easy to come to an agreement concerning what epistemic justification really amounts to. Answers given by internalists often conflict with those of externalists and reliabilists.

My general position is that reliabilism is correct as an account of the property that is needed for knowledge but has some difficulties as an account of epistemic justification. Thus, alternatively, I suggest a new term, “epistemic goodness,” in order to isolate the substantial pre-theoretic property required for knowledge without identifying it as epistemic justification. Then, I propose my own account of epistemic goodness. I focus on the fact that a belief has two aspects to be assessed, namely the aspect related to belief acquisition and to belief maintenance. Using two separate criteria for measuring the
goodness in each aspect, such as truth-conducivity and subjective correctness, I argue that an epistemically good belief is a belief that is reliably produced and is correct from the agent’s point of view. Finally, after considering the cases where epistemic goodness can be undermined, I conclude that indefeasible epistemic goodness is the property that makes knowledge when added to a true belief.

On the way to this conclusion, I critically investigate Alvin Plantinga’s theory of warrant and Ernest Sosa’s theory of intellectual virtue. They are chosen because they are claimed as improved adaptations of the basic reliabilist idea, and, for that reason, they are competing theories to my account. Although I accommodate many of the merits that I found in their theories in my account, I eventually reject both Plantinga’s proper functionalsim and Sosa’s version of virtue epistemology due to the various theoretic difficulties.
Dedicated to my parents
ACKNOWLEDGMENTS

I would like to thank Diana Raffman for her constant support and encouragement from the very beginning of my graduate education at The Ohio State University. I am also deeply indebted to George Pappas for his invaluable comments and discussions both about this dissertation and about epistemology overall. On numerous occasions, he has shown me how good philosophy is done, and I hope that I can follow this example. Most of all, though, I owe my deepest gratitude to my advisor, Marshall Swain, for his guidance, patience, help and encouragement not only through the many drafts of this dissertation, but also through my entire graduate education. Without his support and confidence in me, I would never have finished this dissertation. Finally, I would like to thank my wife, Minja, for her support.
VITA

March 15, 1963 ..................................................... Born – Seoul, Korea

1986 ............................................................... B. A. International Economics, Seoul National University, Seoul, Korea

1989 ............................................................... M.A. Aesthetics, Seoul National University, Seoul, Korea

1994 - present ................................................. Graduate Student, Philosophy, The Ohio State University, Columbus, Ohio

1997 – present ................................................ Graduate Teaching Associate, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Philosophy
# TABLE OF CONTENTS

Abstract ................................................................. ii
Dedication ................................................................. iv
Acknowledgments ....................................................... v
Vita ........................................................................... vi

Chapters:
1. Justification And Epistemic Goodness: A Historical Survey .............. 1
   1. Justified True Belief and Gettier ................................................... 1
   2. Influence of Gettier ................................................................. 5
      2.1. Defeasibility Approach ....................................................... 6
      2.2. Externalist Conditions ....................................................... 12
      2.3. Gettier Generalized ....................................................... 15
3. "Epistemic Support" and a Problem of Externalism ............................... 22
4. Epistemic Justification and "Epistemic Goodness" .................................. 26
   4.1. Reliabilism .......................................................................... 26
   4.2. Epistemic Justification ....................................................... 31
   4.3. Epistemic Goodness ............................................................ 41

2. A Criticism of Plantinga’s Proper Functionalism .................................. 47
   1. Warrant And Proper Functionalism ........................................... 47
      1.1. Conditions for Warrant ................................................... 47
      1.2. Overview of My Argument Against Proper Functionalism .... 52
   2. Is a Proper Function Necessary for Warrant? ................................ 56
1. Two Phases of a Belief and Epistemic Goodness in Each Phase 172
   1.1. Two Phases of a Belief 172
   1.2. Epistemic Goodness in C-Phase 175
   1.3. Epistemic Goodness in A-Phase 178
      1.3.1. What to Evaluate in A-Phase 178
      1.3.2. A "Correct" Belief 187
      1.3.3. Intellectual Virtue 191
   1.4. Application to Cases 197
2. Defeater Conditions for Epistemic Goodness 201
   2.1. Defeater in C-Phase 201
   2.2. A Global Deception: The New Evil Demon Case 208
   2.3. Defeater in A-Phase 217
3. Conclusive Summary 224

List of References 226
CHAPTER 1

JUSTIFICATION AND EPISTEMIC GOODNESS: A HISTORICAL SURVEY

1. Justified True Belief and Gettier

The status of *propositional knowledge* [S knows that p] is conferred upon a kind of mental state usually called a state of believing that proposition [S believes that p]. The term "belief" here is stipulative to some extent. No doubt, a belief state would be a very complicated psychological state. However, for epistemological purposes, it can be treated as if it is a simple state of mind that takes the contents of various propositions as its objects. This unique state of mind is usually distinguishable from other mental states, especially states where the agent merely entertains the proposition.\(^1\) When an agent believes a proposition p, his mental state is characterized by taking an assertive attitude toward the content of p. Although it is customary to distinguish occurrent beliefs from

\(^1\) Entertaining a proposition might be a presupposition for having a belief, thus a part of belief state. However, if so, many other kinds of propositional attitude such as desire, fear and hope also require the "entertainment" of a proposition. What constitutes the unique characteristic of belief is what is more than "entertaining" a proposition.
dispositional ones, what seems to be more typical about having a belief that \( p \) is having a stable disposition to assent to \( p \). Then, the occurrent belief could be treated in terms of the manifestation of such a disposition. Of course, the exact nature of this mental state, either occurrent or dispositional, needs more philosophical discussion and scientific investigation. However, it is at least safe to say that propositional knowledge is a specially qualified mental state of this sort in general. And in this sense, knowledge implies belief.\(^2\) Furthermore, knowledge requires that the proposition that is believed be true. This requirement is based on our ordinary usage of the term “knowledge.” It is impossible to know something that is not true. Then, together with the belief condition, knowledge implies true belief.

However, it is obvious that not all doxastic mental states of believing a true proposition amount to the epistemic state of having propositional knowledge. Suppose that, when I drive to school in the morning, I hear on the radio that a local bank was robbed yesterday. The report is brief and says nothing about the suspect. However, while I listen to the news, I come to believe that the robber is my annoying neighbor Sparky. Obviously, I lack any relevant evidence to make such an assertion. Although it is true that he is not a pleasant man to get along with, I do not have any particular suspicion to

\(^2\) I am open to the possibility that the nature of this mental state can be better captured by some other notions. Usually, the term “acceptance” is suggested as a replacement for “belief” on the ground that the agent needs to be more committed to the proposition than what is normally required for having a belief. On a similar ground, the distinction between belief (or acceptance) and mere possession of information might be proposed. I think these are important distinctions. However, it is unclear how much the agent should be committed (that is, how strongly he should be assertive) toward \( p \) in order to be considered as knowing. There probably is a spectrum in terms of the strength of the commitment, and possibly, acceptance. Belief and mere possession of information may take up their own space there. Here again, relying on the
connect him with the criminal activity of that caliber. However, his name and image happen to pop up into my mind at that particular time and, maybe because I have such an unpleasant feeling toward him, it just so happens that I come to believe that he is the robber. In other words, I have acquired the disposition to take an assertive psychological reaction toward the proposition that Sparky is the bank robber. Now, to everyone's surprise, suppose that Sparky is in fact the bank robber. He has a secret life not known to me. He occasionally robs banks for money and excitement, and he did it again yesterday. So, my belief that Sparky robbed the bank is true. However, there is no chance that this belief would be considered as knowledge. (Of course, I would say that I knew it all along when I see the arrested bank robber's mug shot in the front page of the newspaper a few days later. However, despite my claim, again, this would not really count as knowledge.)

It is in fact not at all hard to come up with a case in which believing a true proposition does not amount to knowing that proposition in one way or the other. As in Russell's original example of the stopped clock, I could have a true belief about the current time (it is noon) by looking at the stopped clock (it happened to stop at midnight). Again, this true belief cannot be knowledge. Even more commonly, we are all familiar with how guesswork can provide a true belief more often than not. However, since guessing is exactly what we do when we do not know, "a correct guess" (true belief as the result of guessing) is not knowledge. Then, it is obvious that in order to have knowledge,

---

3) According to Shope, "Russell only presents [this example] in order to show that there can be true belief without knowledge," and later Israel Scheffler suggests this as another Gettier type example. Shope, Robert.
there has to be more than just having a true belief.\textsuperscript{4} Identification and characterization of such an additional property (or properties) that fills the gap between a true belief and knowledge has been one of the major projects in contemporary epistemology.

In the long tradition of western philosophy, the term "justification" has been selected in order to represent this additional property. The outcome is the allegedly inherited JTB analysis of knowledge, in which knowledge is analyzed as a justified true belief. Here, justification is understood in a generic sense. However, the JTB analysis, relying on a then-not-clearly-articulated notion of justification, reveals its difficulty when Gettier questions it in his 1963 article.\textsuperscript{5} Gettier creates a pair of examples in which a true belief justified in a normally acceptable sense fails to be a case of knowledge. One of the famous Gettier example goes as the following. Smith believes the proposition that (a) the man who will get the job has ten coins in his pocket because he believes that (b) Jones is the man who will get the job and Jones has ten coins in his pocket. Smith's evidence for believing (b) is impeccable, including hearing the company president's assurance of Jones' promotion and watching Jones count the ten coins from his pocket right before the moment. As long as Smith is well aware of the entailment from (b) to (a), Smith's belief (a) is justified in our normally acceptable sense. However, it turns out that it is Smith himself, not Jones, who gets the job. Despite this major twist of fate, Smith still ends up

\textsuperscript{4) An interesting perspective that denies the requirement of the further condition (typically, the justification condition) for knowledge is found in Sartwell, Crispin, "Knowledge is Merely True Belief," \textit{American Philosophical Quarterly} 28 (1991) pp. 157 - 165

with a true belief because he himself happens to have ten coins in his pocket. However, Smith’s true and justified belief (a) cannot be a case of knowledge. Smith does not know that the man who will get the job has ten coins because it is accidental that he has a true belief.

2. Influence of Gettier

Various diagnoses have been given of Gettier examples together with solutions, i.e., revised analyses of knowledge. Without going into the details of those revised analyses, I will discuss here the notable main reactions to Gettier problem. From the perspective that these post-Gettier debates have shaped the current state of epistemology, I would like to point out the following three as the major influences of Gettier: (a) It makes epistemologists to look for an additional condition in order to correct the problem of JTB analysis. (b) Also it makes some epistemologists replace the justification requirement in JTB analysis with something else that can successfully fill the gap between true belief and knowledge without being caught by the Gettier problem. (a) is a direct response to taking the Gettier problem as a challenge to the sufficiency of JTB conditions, while (b) is the result of taking it as pointing out the un-necessity of the justification condition for knowledge. In any case, both (a) and (b) force epistemologists to reflect on the role of justification in an account of knowledge. Thus, the influence of the Gettier problem and post-Gettier debates should include the following: (c) It makes epistemologists investigate the notion of justification thoroughly.

In this section, I will first compare position (a) [JTB + Gettier solution] with position (b) [QTB, where Q is a replacement of J, the justification requirement] together
with a brief historical survey of each position. To state my own conclusion up front, I prefer the structure of an account of knowledge in general to be XTB + Gettier solution, where X is the replacement of the traditional justification requirement. This would be different from the position (b) because it needs a specific Gettier solution. I do not believe there would be any single property, such as Q in (b), that is a part of the analysis of knowledge, i.e., a necessary condition for knowledge, and at the same time also provides a complete Gettier-antidote when added to true belief. My understanding also partially denies the position (a) because what I think of as X would be eventually different from J in (a). Especially, if J in (a) is understood in a traditional way, typically as a deontic notion, then what I think of as X would not imply J. This point will be made in the following sections when Gettier influence of (c) is considered.

2.1. Defeasibility Approach

Naturally, (a) would be the most direct response to Gettier. Once Gettier examples are accepted as a genuine problem, many epistemologists think that, if we add one or more condition(s) on the basis of the lesson we have learned from the diagnosis of Gettier examples, this will take care of at least the issue of what else is needed for knowledge. As I mentioned, for such epistemologists, Gettier cases are understood as the problem of whether a justified true belief is sufficient for knowledge rather than a question of whether any of the elements in JTB analysis (typically, justification) is genuinely necessary for knowledge. However, concerning what exactly should be added, there has been some trial and error.
The most direct attempt to counter the Gettier examples would be the “No False Lemmas” requirement. The idea is that, since each Gettier example contains at least one false intermediate conclusion in its way to reach the final true belief (“Jones will get the job” is such a false lemma in the above mentioned Gettier example), simply asking that a true belief should not be acquired in such a way would avoid the problem. However, the variety of Gettier type examples, which come after the original Gettier cases, includes one that has nothing to do with a false lemma. The well-known barn façade case is also one of them. In this example, an epistemic agent is looking at a real barn under normal perceptual conditions, and correctly believes that there is a barn. We may say that this belief is not only true but also sufficiently justified because it is based on his perceptual evidence. However, according to the example, the agent’s environment contains many barn facades propped up by sticks, of which he has no awareness. In fact, what this agent happens to look at is the only real barn around. Furthermore, it is stipulated that he has no

6) Harman’s principle P is that “reasoning that essentially involves false conclusions, intermediate or final, cannot give one knowledge.” Harman, Gilbert, Thought (Princeton, NJ: Princeton University Press, 1973) p. 120.

7) Harman is not the only one who suggests this type of solution. In fact, many epistemologists including Lehrer and Armstrong thought that Gettier problem comes from the fact that the agent actually believes a false proposition. Thus, for example, Meyers and Stern proposed the following “principle” as the solution of Gettier example: a proposition can justify another proposition only if it is true. As long as this principle is true, they think that the JTB analysis is still acceptable. Meyers, R. and K. Stern, “Knowledge without Paradox,” Journal of Philosophy 70 (1973): 147 - 160

8) One such example is found in Feldman. Feldman, Richard, “An Alleged Defect in Gettier Counterexamples,” in Empirical Knowledge 2nd edition, ed. Paul Moser (Lanham MA: Rowman & Littlefield, 1996): 241 - 243 (originally, Australasian Journal of Philosophy 52 (1974): 68 -69) Here, Feldman shows that if the Gettierized person [Smith] deduces the proposition in his Gettierized belief [the person who will get the job has ten coins] from a true existential generalization [there is someone in this office who will get the job and has ten coins] instead of a false proposition [Jones will get the job], then we will have the a justified true belief that fails to be knowledge yet does not rely on false lemma.

ability to distinguish a real barn from a barn façade from his standpoint. So, he could easily see a barn façade instead of the real barn, and he would falsely believe that it is a barn even if what he sees is a facade. Under this circumstance, the example achieves what it intends to show, i.e., the agent does not know that there is a barn. However, note that in this process the agent does not rely on any false intermediate reasoning step or any of this sort. He infers it—if he infers it at all—from the perceptual experience of the barn, which is a genuine perceptual experience of the agent. So, it shows that a justified true belief without using any false lemma could still fall short to be knowledge. Thus, the strategy of not allowing “justification relying on falsity” such as “no false lemma” approach is seriously limited in handling a certain type of post-Gettier examples.

Dretske’s “Conclusive Reason” proposal\(^{10}\) can handle cases that the “no false lemma” requirement cannot. According to Dretske, a reason \(R\) is a conclusive reason for believing \(p\) if and only if, given \(R\), not-\(p\) is [ceteris paribus] impossible. Then, in the barn façade case, the \(R\) (maybe, the perceptual experience of a barn, which is the same as the perceptual experience of a barn façade) is not a conclusive reason for believing that there is a barn because, even with this \(R\), it is possible that the proposition that there is a barn is false. Or, putting it differently, even if the proposition were false (in the case where the agent sees a barn façade), the agent would have the same \(R\). Since knowledge is possible only with a conclusive reason, the agent in the barn façade case does not know that there is a barn. Judging from the way in which it handles the barn façade case, Dretske’s

diagnosis of Gettier examples is closer to the heart of the problem. However, eventually, Dretske’s proposal requirement generates a similar result as Harman’s. There are simply some other Gettier type examples that this proposal cannot properly handle.11 For example, suppose there is a cup on the table. This cup cannot be seen directly by an agent, S, because it is hidden behind a very complex system of mirrors. However, through the system of mirrors, S can see the reflected image of the cup. Now, given that S has no idea about the existence of the mirrors (thus, he thinks that he is directly seeing a real cup), it is doubtful that his belief that there is a cup on the table is a case of knowledge, although it is a true belief. However, the reason R (the perceptual experience of the cup) seems to satisfy Dretske’s conclusive reason. That is, if there were no cup on the table, S would not [ceteris paribus] have R.

Not all attempts are refuted, however. After a series of new analyses followed by new counterexamples, eventually the “Indefeasibility” requirement emerges as the fourth condition of knowledge.12 The defeasibility approach maintains the structure of the JTB analysis but requires its justification to be undefeated. The basic idea of the indefeasibility requirement is that, in order for S’s justified belief to be knowledge, the body of evidence that provides justification to the belief must be complete enough that no additional fact available to S overrides the prima facie justification of the belief. According to this, the


available additional fact that the environment contains many barn facades that the agent cannot tell apart from the real one overrides the *prima facie* justification of the original perceptual belief that there is a barn. Such a belief with defeated justification cannot be knowledge, which is the desirable outcome of the case. In the same way, the indefeasibility condition can effectively handle all of the known Gettier problems so far.

There are several ramified versions of this basic idea, especially concerning how each version deals with the case where the overriding additional fact is "merely misleading." An example made by Lehrer and Paxson illustrates one such case. In this example, suppose that I eye-witness my friend, Tom Grabit, stealing a book from the library. Since I am well aware of what Tom looks like, I come to believe that Tom stole a book on the basis of what I have correctly observed. Let's say that this belief is a justified true belief. However, justification of this belief would be overridden by the following facts: that Tom's mother told the police that Tom was not in the town at the time I saw him at the library, that Tom has an identical twin brother Tim, and finally, that Tim is suffering from kleptomania. Even if the testimony by Tom's mother is made unbeknownst to me, the fact that Tom's mother said such things would be the defeating evidence that, when added to my original evidence, overrides the *prima facie* justification of my belief that Tom stole a book. As the result, I do not know that Tom stole a book in this scenario, which is correct just like in the barn façade case. However, suppose now

---

that the testimony of Tom's mother is fabricated. In this scenario, it seems incorrect that my belief would lose justification because of the testimony of Tom's mother. So, what used to be a defeater in the first scenario now turns into "merely misleading" evidence in the second scenario. The question is how to capture the difference between them.

Of course, the obvious difference is that, in the first scenario, the testimony of Tom's mother is true while in the second scenario it is false. Because of that, Klein suggests that the genuine defeater should not "essentially" depend on a false proposition as an explanation of the merely misleading evidence. Another way of handling the notion of merely misleading evidence is suggested by Swain, who focuses on the defeasible nature of the defeater itself. That is, in a case like the second scenario of Tom Grabit, although there is a potential defeater (the mother's testimony), this defeater is not genuine but merely misleading because it is defeated by additional facts (i.e., she is lying). So, in this proposal, whether a given justified true belief is knowledge or not is determined after considering all the relevant facts. Namely, its justification should be ultimately undefeated.

By taking one of these versions of the indefeasibility condition as a necessary part of the "gap-filler" between a true belief and knowledge, we are admitting that what is required for a true belief to be knowledge cannot be completely internalistic. Roughly, an epistemic element is internalistic if it is what the agent is in a position to be aware of by reflecting on it. Thus, internalistic elements are usually mental states of the agent, and a kind of special access by the agent to those elements is presupposed. Now, since the

---

satisfaction of the indefeasibility condition does not depend on what is in the mental state of the agent (in other words, a justification-defeating circumstance, or a defeater, is obtained regardless of whether the agent is aware of it or not), what makes a true belief knowledge is not wholly constituted by internalistic elements. However, as the term "defeasibility" suggests, what it does presuppose is the existence of a certain epistemically good quality, such as prima facie justification, that can be overridden when there is a defeater. So, the indefeasibility requirement would be a proviso provided a true belief satisfies the other primary condition for knowledge. What then is such a primary qualification that should not be defeated in order to have knowledge? Would the generic notion of justification still be enough? The significant influence of Gettier is that it brings out the realization that a thorough investigation of the third condition for knowledge is in order if there is to be any proposal for the fourth condition. As a consequence, the notion of justification becomes the subject of many sophisticated analyses. Although debates concerning the nature of justification are not settled, many epistemologists do not doubt that knowledge implies justification. Here, the term "justificationist" is appropriate in order to refer to those who do not deny the necessity of justification as the third condition of knowledge.

2.2. Externalist Conditions

Not every reaction to Gettier problem assumes the necessity of justification. At this point, it is worth noting that the source of Gettier problem can be examined from a different perspective. What prevents a true belief from being knowledge in most of the cases is some kind of "accidentality." That is, an accidentally true belief is the kind of
true belief that is disqualified as knowledge. In fact, our notion of knowledge is very sensitive to the notion represented by a group of terms such as chance, accident, coincidence, luck and fluke. Although it is unclear exactly to what extent each of these should not be involved in knowledge or how to draw a line between what is possibly an acceptable kind of luck for knowledge and what is not, it is generally true that, in order to have knowledge, the existence of a true belief in the agent’s cognitive system should not be accidental. In this sense, the Gettier cases are disqualified as knowledge because some kind of accidentality is involved. According to this perspective, since knowledge is roughly a “non-accidental” true belief, the third requirement for knowledge—whatever it is—is supposed to deal with this accidentality in order to make any true belief knowledge.

Then, in this view, the fact that Gettier examples are all justified beliefs only shows that the generic sense of justification requirement cannot eliminate this kind of accidentality. If justification cannot completely block accidental true beliefs from being knowledge, why do we have to keep it as a condition for knowledge?

In this sense, Gettier cases can be understood as attacking the justification requirement. For those who think this is the real lesson from the Gettier problem, what is needed for a solution is a plausible interpretation of “non-accidentality” so that they can use it for the necessary and sufficient condition for knowledge. Then, knowledge would be defined eventually as a non-accidental true belief. One thought is that if there is an epistemically relevant connection between S’s belief that p (Bp) and the states of affairs that p, that is, if Bp is somehow dependent on the truth of p, then S’s being right about p
wouldn’t be accidental. Nozick implements this idea straightforwardly, and suggests the following condition in addition to the truth and the belief condition: if p were not true, S would not believe that p, and if, in changed circumstances, p were still true, S would still believe that p. Armstrong, on similar grounds, suggests the so-called “Thermometer View” of knowledge. According to this view, S’s true belief that p is knowledge if and only if there is a certain specification H such that, if S is so specified, then, if S further believes that p, then p is the case. In other words, for knowledge, there should be a law-like connection between S’s being in H and S’s having a true belief that p. Also in the same spirit, other epistemologists including Goldman propose an account that requires the connection between p and Bp to be causal.

These accounts of knowledge are characterized as externalist because, if we follow them, none of the requirements for knowledge has to be internalistic in the sense I mentioned earlier. In internalist theories, what is necessary for making a true belief knowledge—and this can be called the “justifier” given that most of the internalists are also justificationist—is typically reasons or evidence that the agent has. And, if how the fact that p is connected to Bp is supposed be the essential part of the third condition for knowledge, as externalists claim, then internalists would say that such an internal state of S (the reason state, for example) is what mediates p and S’s Bp. Indeed, such an internal

15) “This something more, I think, is not simply an additional fact, but a way that 1 [the truth condition] and 2 [the belief condition] are linked.” Nozick, Robert, Philosophical Explanations (Cambridge MA: Harvard University Press, 1981) p. 170.
state may be able to play this role. In an ideal case of a justified belief in internalism, the states of affairs p is connected to S’s evidential or reason state, and then this internal state is in turn connected to S’s belief. However, the question is whether the alleged task of the elimination of accidentality for knowledge should always be done in this way. Externalists think that this mediating internal state is not necessary. By satisfying a causal relation, a nomological relation or a subjunctive condition between p and Bp (if it weren’t p, it wouldn’t be Bp), the “accidental connection” between p and Bp can be effectively eliminated, and none of these accounts necessarily requires internal states of the agent.

2.3. Gettier Generalized

However, an important question for the externalists is whether these external conditions are also successful as a Gettier solution. Before answering this question, let’s first see the general structure of the Gettier examples. Here, I will provide a general “recipe” of how to Gettierize a belief, by following the lead of Zagzebski. According to her diagnosis, any candidate for the additional condition(s) for knowledge, if it is supposed to deal with Gettier problem, should imply the truth of what is believed. Otherwise, there always exists room for “Gettierizing.”

Suppose that the third component of knowledge, which is required for knowledge in addition to true belief, is called Q. Does having a Q-belief entail having a true belief?

---

Justification, although it is the traditional candidate for Q, is not usually taken to imply truth. That is, being justified does not imply having a true belief, and thus, a justified false belief is not a contradiction at all. And this seems to be intuitively right. Intuitively, it looks as if Q should refer to an independent epistemic property that even a false belief sometimes can have. However, as long as this is the case, Zagzebski points out that any definition of knowledge as true belief plus Q would be caught by a type of Gettier example. She thinks that Gettier examples are cases of "double luck" where bad luck is compensated by good luck. Thus, the general recipe for Gettierizing would go as follows: First, suppose a false belief that has enough of Q. Usually, if a belief has Q but fails to be true, the falsity of the belief is often due to some element of (bad) luck. In the earlier example, Smith's belief that Jones will get the job is such a belief. The falsity of this belief is due to a kind of bad luck, as when Smith becomes a victim of the practical joke of the company's president. Given that the first hand information from the president is normally a reliable source for such a matter, if there were no such "bad luck," Jones would have been the person who gets the job, and Smith would have ended up with a true belief that can be knowledge. Now, add another kind of (good) luck that just makes the belief true without changing Q. The fact that Smith also has ten coins in his pocket is this kind of luck. This second element of luck is independent of Q. As a result, we have a true belief which has Q but is not considered as knowledge. The source of the problem is

20) The term "luck" is chosen to represent a group of words that are related with the notion of chance, including accident and coincident.
21) A typical example would be a case where the conclusion of a strong inductive argument turns out to be false despite its high probability to be true than to be false.
that a Q-belief does not imply a true belief. The lesson should be this: In order to prevent any QTB analysis of knowledge from being victimized by Gettier type counterexample, Q must imply truth.

Now, how successful are the above-mentioned externalist conditions on this score? It looks that a nomological or a causal relation between p and Bp may well guarantee the truth of Bp. Can we then accept that any one of them would be the Gettier-proof analysis of knowledge? That would be too quick because, although Zagzebski’s point that the property Q must imply truth is correct, it is only a necessary condition for Q to avoid Gettier. It is not sufficient to do so. Even if the truth of Bp is connected to p by satisfying any of the above externalist conditions (thus, even if Q implies truth), the belief can still be Gettierized. Consider the example of the barn façade again. Here, such a causal condition is satisfied. The agent has a true Bp, which is straightforwardly caused by the state of affairs p.\textsuperscript{22} However, it is not a case of knowledge. Presumably, the reason is that we can spot some kind of luck or accidentality at a more general level, which cannot be compatible with knowledge, although we do not have a firm understanding of what exactly is the nature of such accidentality. The same point can be made by the example of the broken thermometers. Let’s suppose that a prepared mother has a dozen thermometers in her medicine cabinet. Unfortunately, none except for one is in a proper

\textsuperscript{22} When the state of affair p causes the true belief p, whether there also exists a nomological connection between them seems to depend on the scope of this alleged law-like connection. If the law-likeness of this causal connection is understood as a necessary and sufficient condition, then the very existence of the perceptual equivalence such as a barn façade, or maybe a holographic image, undermines the existence of a “law” between the state of affair p and the true belief p. However, even so, at least the “one way” law-like connection, i.e., when other conditions are met (including that the agent is appropriately presented by the
working order, though which one is not known to her. She randomly picks one without any suspicion, measures her child’s temperature and believes the reading. Luckily, she picks the only properly functioning one, and as a result, comes to have a true belief. Again, its causal chain or the law-like connection is the right kind but we are surely ready to challenge the claim that the mother knows her child’s temperature.

What these examples show is that Q, if it has to be sufficient for avoiding Gettier, would be a much more complex property than externalists initially think. Their initial approach to the problem, that is, eliminating the “accidental” elements from a true belief, is promising. However, establishing a truth-conducive relationship between p and Bp seems to be too weak to be a Gettier-proof analysis of knowledge. Even 100% truth-conducivity, as it is required by Zagzebski, cannot deal with all possible sources of accidentality. We will need a supplementary condition for this. That’s why I think that we should take advantage of the fourth condition such as indefeasibility condition in order to complete the analysis of knowledge even if we take the non-justificationist position. Given that Q can hardly be a single property, it can be divided into two parts. One is a part that is conceptually required for knowledge in addition to true belief (let’s call this X), and the other part is to take care of the “defeater” of it (let’s call this G). The justificationists and non-justificationists debate is about whether the property X should be the property of justification. However, regardless of who eventually persuades us on this issue, my point here is that we cannot, and need not, expect X to block the Gettier state of affair that p), if there is the state of affair p then there is true belief p, would be considered to be there.
problem completely. The property $X$ may contribute to removing a certain type of accidentality depending on how it is characterized, and in this sense, the above-mentioned externalist conditions may well be the candidate for $X$. However, accidentality such as the one involved in the barn façade case and the broken thermometer case would need additional conditions because these examples raise the issue of a special kind of defeater for $X$.

Again, the indefeasibility condition seems to fit into this scheme very nicely. The indefeasibility approach can be neutral concerning $X$. Also, it can be constructed in such a way that there should be ultimately no truth (defeater) such that, when combined with the source of $X$, it stops providing $X$ to the belief, regardless of the nature of the $X$. If so, setting aside the details of how to deal with true but misleading evidence, it allows the basic structure that we need, that is, $X + G$ (indefeasibility) implies truth of the belief. The indefeasibility approach is also general enough not to be considered as an ad hoc solution of Gettier. It generates a satisfactory result not just in original Gettier cases but also in many variations of it including the ones that those externalist alternatives could not handle.

This is not to acknowledge any one version of indefeasibility account already in the field as the general solution to the Gettier problem (that is, the $G$ part of any $XTB + G$ scheme) no matter what the other requirement for knowledge is. Naturally, we can expect that the exact shape of the Gettier solution would be different depending on the nature of $X$ although the operating idea and the goal is the same. This is because the nature of the defeater would be differently characterized depending on $X$. A defeater could either defeat $X$ (a belief had $X$ but in a new circumstance it no longer has $X$, thus fails to be
knowledge) or defeat knowledge (a belief's X is intact but it falls short of knowledge in a new circumstance). For instance, consider the barn façade example. Concerning the justificatory status of the agent's "accidentally" true belief, a defeater condition can explain why this justified belief fails to be knowledge. That is, this belief is justified but such justification is insufficient for knowledge due to the available defeater. What is normally observed in this example is that the defeater prevents the so-called effect of the *prima facie* justification (namely, the effect that justification generates knowledge when added to a true belief) rather than canceling the existing justification. In other words, the agent maintains the justification for the belief but such a justification is not enough to generate knowledge, thus it is appropriately called a defeated justification. This is a perfectly legitimate understanding of the case. It assumes that what provides justification to the agent's barn belief is intact even under the barn façade circumstance. Again, this assumption is unproblematic because in most of cases, we think justification is independent from the factors in the environment, over which an epistemic agent usually has no control. This is the case where a defeater does not attack X itself in the X + G scheme.

As for what provides justification (or the X), we can think of various candidates such as "an evidential support," "proper functioning of a belief-forming faculty" or "reliability of the belief-forming faculty." Then, no matter how the circumstance is changed, a belief that has X remains as such. It is highly "accidental" that the agent

---

23) Here, the defeater could be understood either as propositions (existence of a true proposition such as "there are many barn facades that the agent cannot distinguish from the real barns") or as the state of affairs.

---
happens to spot a real barn among many fake barns. It is this additional accidentality that prevents the belief from being knowledge, not the defect of the X. Then G is to deal with such additional accidentality. However, if the notion of justification (or the X) is sensitive to, for example, the local reliability of a belief-formation, then it is possible to have a defeater which directly attacks the X. It may be that Sosa’s justification, which will be discussed in details in chapter 3, reacts to a defeater in this way. Sosa’s justification presupposes “aptness,” which is the equivalent to the reliability requirement. Then, it appears that, in a barn facade circumstance, the barn belief that the agent has is not apt. As a consequence, the belief is not justified even though the same belief is apt and justified in circumstances where there are no barn facades. Here, the independency of G is weakened. It is absorbed by the conditions for the X. Thus, depending on how we understand X, the nature of the defeater would be construed differently, and in turn, the condition G, which is designed to deal with such a special circumstance, would be constructed differently.

In any case, the basic idea of the indefeasibility approach is readily adaptable, which I believe is another advantage of the indefeasibility approach. If X refers to any kind of positive contribution for a true belief to be knowledge, G would take care of special circumstances where either the X itself is deteriorated or the effect of the X is canceled. The idea of defeasibility shows that this can be done at least without denying the existence of the X, the necessary condition for knowledge. More often than not, the purported solutions of Gettier problem lose sight of what is necessary for knowledge in

(existence of many barn facades that the agent cannot distinguish).

21
the attempt to make a true belief Gettier-proof. Although it is still debatable whether such a necessary condition should be justification, the above externalists' candidates for Q (the property that can be the third condition for knowledge and the Gettier solution at the same time) seem to fail to consider some elements that are essentially required for the explanation of knowledge, which I will discuss in the next section. Thus, I will now proceed with the assumption that the indefeasibility condition or a very similar one would be the valid solution of Gettier problem. By doing so, my discussion will be able to focus on the property X rather than X + G. Concerning what would be my specific proposal for the condition G, I will add a sketchy proposal in the chapter 4, where I will mainly discuss my own understanding of the property X.

3. "Epistemic Support" and a Problem of Externalism

Consider a belief that an agent comes to have after reasoning through a strong inductive argument. Suppose that I believe that Sparky loves pasta, and also believe that almost all pasta lovers in this town dine at the restaurant Casa di Pasta. Then, given that I don't have any particular evidence to think otherwise, my new belief that Sparky has eaten at Casa di Pasta has a certain epistemic quality. It is acquired by virtue of seeing the inductive evidential relationship between the two propositions that I already believe and this new proposition. It's not because, as a matter of fact Sparky did have a dining experience at this restaurant, which makes the new proposition that I come to believe

---

24) Thus, from now on, my locution that 'something is what makes a true belief knowledge' usually refers to X only although, strictly speaking X cannot make a true belief knowledge. What is assumed is that G is
true. Of course, truth is the ultimate value when we believe something but we also see some additional epistemic property that is independent of truth. That is, this belief has an epistemic quality that wouldn’t be lost even if the proposition believed turns out to be false (when, to my surprise, Sparky has never been to the restaurant). This epistemic quality, I presume, has something to do with the characteristics of the particular conditions surrounding the existence of the belief in my “cognitive system” and/or the particular way in which I come to believe the proposition. Approximately, what we see as the quality in question is that my believing the proposition [Sparky has eaten at Casa di Pasta] is being “supported” in a generic sense of the term (by the other beliefs and evidence that I have and/or by non-fallacious inductive reasoning that I made based on them) in such a way as we think we want to see in knowledge.

Consider a belief that an agent comes to have using non-defective perception. If I believe that something gray is in front of me when I have a visual experience of a gray thing, my belief is being “supported” by my experience in a way in which we think appropriate for knowledge. Although it is a different kind of support that we find in the reasoning through an inductive argument, we find a certain property related with a kind of “support” that this belief gets from the perceptual experience. Similarly, in most of the cases we would intuitively agree as candidates for knowledge, we find this type of extra property other than being true, which can only be loosely circumscribed as “a certain kind of support” at this point. Let’s tentatively call this clearly existing property “epistemic

---

25) Concerning this notion, see chapter 4, p. 182.
support" for the convenience of further discussion (yet we cannot tell exactly what it is). By calling whatever property we find in true beliefs qualify as knowledge (in inductive reasoning, in perception and in many other occasions) “epistemic support,” I am not introducing any theoretical background here. It is only to choose a name for the unknown property based on its first and rough impression. I do not claim that we know what this property named “epistemic support” is. At this point, my claim for the departure of my project is that we can confirm this loose sense of “a support from something else” as an extra property that a true belief has in all non-controversial cases of knowledge. Also, since we named this extra property (whatever it exactly turn out to be eventually) as “epistemic support,” if this extra property is what is necessary for knowledge, then trivially “epistemic support” is necessary for knowledge. Explaining the nature of this “epistemic support” is the project that I will eventually undertake in this dissertation.

Externalists understand the nature of this “epistemic support” in terms of externalistic elements. In other word, externalists understand that the kind of support that the target belief enjoys in order to be knowledge in terms of natural and external relation. As I mentioned earlier, internalistic elements are what the agent is in a position to be aware of. Internalism is characterized as a position by the claim that only the internalistic elements can provide the legitimate epistemic support that is required for making a true belief knowledge. Externalism is the denial of this position, and claims that externalistic elements such as nomological or causal connection can explain the nature of the epistemic support that is necessary for knowledge. However, in addition to the problem of these external conditions not being the solution of Gettier problem contrary to the intention of their advocators, which I indicated above, there has been a doubt if these
conditions correctly capture the nature of the epistemic support that we think exists. This point has been raised using the following type of example.

Suppose a person named Mr. Truetemp, who has some kind of electronic chip implanted in between his skull and brain.\(^{26}\) This chip registers the outside temperature accurately as a normal thermometer does. However, adding some more futuristic idea, this electronic chip "informs" him of the correct temperature through a certain mechanism. On the basis of this "information," Mr. Truetemp forms beliefs about the current temperature, which are always true. There is a nomological and/or causal connection between the current actual temperature and Mr. Truetemp's true belief. The externalists' requirement for knowledge seems to be satisfied. However, the tricky part is that he has absolutely no idea of the existence of such a chip in his body. Furthermore, he has no reason or evidence concerning his temperature belief that he is in the position to be aware of. Apparently, he believes "what comes into his mind," if that's the correct description of how the electronic chip "informs" him of the current temperature. Although there is a sense in which his beliefs are supported as externalists might insist, the dominant feature of this example is that his beliefs are groundless. Epistemically relevant grounds are usually understood as accessible mental states of the agent as internalists normally assume. Consequently, Mr. Truetemp's true beliefs are not considered as knowledge unless one attempts to appeal to some extremely externalistic intuition. If Mr. Truetemp does not have knowledge, it shows that externalism is missing

some ingredients in its analysis of knowledge, ingredients that we normally consider essential for knowledge. What would it be? Would it be “justification”?

4. Epistemic Justification and “Epistemic Goodness”

4.1. Reliabilism

So far, I have compared the justificationists’ reaction to Gettier (indefeasibility approach to knowledge) with the non-justificationists’ (externalism on knowledge). I think the indefeasibility account is promising. But I do not have to agree that “justification” is necessarily the property that, if undefeated, generates knowledge. In a similar way, even though I have pointed out the unsatisfactory nature of the externalism on knowledge, I have to acknowledge that what they attempt to put stage in epistemology is nevertheless very significant. When the post-Gettier debates eventually ignite the full-scale investigation of the notion of justification, the view that the externalists have about the qualification of “the third condition” for knowledge helps to form a new tradition even in the theory of justification. Through the nomological, counter-factual and/or causal relationship, what externalists see is that “truth conducivity” is relevant for our knowledge attribution. Truth-conducivity is in fact a very plausible candidate when we attempt to explain the nature of the epistemic support (or the property X) that knowledge should have. For example, Goldman thinks that this is essential for epistemic

27) Goldman’s account of justification can be called “externalism on justification,” while non-justificationists such as Armstrong’s position can be called “externalism on knowledge.”
justification.28 For him, if a belief is justified, it is truth-conducive. This is the motivation for a reliabilist theory of epistemic justification.

We need to distinguish a proposition's chance of being true from a belief's chance of being true. For example, a proposition expressing a mathematical truth such as “7 is a prime number” has a 100% chance of being true. Even a proposition like “Sparky did not dress up like a circus clown when he robbed the bank” has a great chance of being true given how things are in this world. However, this does not mean that, whenever anybody believes such a proposition, it is justified or has the epistemic support needed for knowledge. What matters here is the characteristic of a propositional belief as a belief, not as the propositional content.

Each belief in one's cognitive system has certain characteristics by virtue of being a particular belief of a particular agent. Most notably, each belief has its genealogical history. According to reliabilism, especially Goldman's version of process reliabilism, justification of a belief depends on the characteristics of the belief as the product of certain belief-forming processes. A belief produced by a particular belief-forming process, namely a reliable process, would have a high chance of being true. According to him, this is what we find commonly in the case of having a belief via strong inductive reasoning, or via perception, and in all the other beliefs for which we recognize the epistemic support. Since the property of "being produced by a reliable belief-forming process" is an objective relation that the belief (or the agent) has to the world, and this

---

relation is not required to be grasped by the agent, Goldman's account on justification inherits the idea of externalism on knowledge. This view, namely justification as truth-conducivity, provides a different perspective on the nature of justification and generates many new discussions. Before reliabilism, the nature of justification has been debated mainly between foundationalists and coherentists. Reliabilism is accepted by many, at least partially if not fully, as an alternative, especially because neither foundationalism nor coherentism could successfully solidify their positions as viable. Given that some basic assumptions and consequences of each position are challenged, reliabilism becomes one of the most discussed positions in epistemology over the last two decades.

However, reliabilism as an account of justification also faces several serious challenges of its own. First, there is a difficulty called the Generality problem, officially spotted by Feldman. Feldman points out the nature of the problem in the following way. A belief-forming process should be understood as a type rather than a token in order to measure the reliability. Obviously, there is in fact one unique process that is responsible for the existence of the belief in question. However, if this unique belief-forming process is unrepeatable, it does not make sense to talk about its reliability. The outcome of this unique belief-forming process is the belief in question, and that would be the only outcome. Then, any true belief would be considered as the outcome of a 100% reliable process, while any false belief would be the outcome of a 0% reliable process. Reliabilism cannot be maintained with this understanding. This is what Feldman calls the

29) In addition to Goldman, the notables include Schmitt, Alston, Swain, Sosa and Plantinga.
Single Case problem. In order to avoid the Single Case problem, a belief-forming process should be repeatable, and that means it should be considered as a type. However, once we consider a belief-forming process as a type, any given belief could be the result of several different belief-forming process types. For example, the belief that is just formed in me, say the belief that there is a computer screen in front of me, could be the result of belief-forming process type “perception.” However, it is equally plausible to think that my belief is the outcome of the process type “daytime visual perception,” “visual perception of an immobile thing,” “visual perception after staring at the same object more than an hour,” and so on. Moreover, it is possible that all those process types have a different reliability. It is imaginable that my perception in general is unreliable although my daytime visual perception is fairly mistake-free. This is the other side of the Single Case problem, called the No-Distinction problem. Then, in order for reliabilist to say that a given belief is a product of reliable or unreliable belief-forming process, there should be a principled way to determine which process type is the relevant one. According to Feldman, there is no non-arbitrary way to avoid the No-Distinction problem without being caught by the Single Case problem. This criticism, if it stands as Feldman claims, threatens process reliabilism at a fundamental level.

As for the replies from reliabilists, there have been several including Goldman’s. More recently, Alston provides a criterion for identifying a unique type, which might

---


be considered as an acceptable solution of the problem, or at least an acceptable beginning of the solution. Furthermore, it is not certain that other versions of reliabilism such as reliable indicator theory or reliable "faculty (mechanism)" theory (as in Plantinga's proper functionalism) would suffer from the exact same problem as Goldman's process reliabilism. Despite Feldman and Conee's conclusion, this problem is still an on-going issue.

The other criticism against a reliabilist theory of justification is based on the evil demon scenario. The criticism presented by Cohen is this. If an epistemic agent S is justified in having, for example, a perceptual belief that p in this (normal) world on the basis of whatever reason R, his counterpart S* in the demon world (or S who now becomes a victim of demon's deception unbeknownst to him), who maintains exactly the same internal features as S including R should be equally justified in believing that p. The problem is that, due to a demon's manipulation, all of S*'s perceptual beliefs would be false. Thus, the process that S* uses to acquire his beliefs is unreliable. This is considered as a counterexample against the main claim of reliabilism that objective reliability is necessary for epistemic justification.

Yet another criticism is raised regarding the sufficiency of reliability for justification. Bonjour has made up several different cases in order to show this point. Through these examples, Bonjour asks whether we want to say the each agent's true belief is justified if the agent acquires it using reliable clairvoyance about which the agent has no awareness. For one reason or the other, Bonjour expects that our intuition would judge these beliefs as not being justified. If so, this would cause a problem in reliabilist theory of justification because their beliefs are the result of reliable clairvoyance. The example of Mr. Truetemp can be used to point out the same problem. Although the Mr. Truetemp example is about knowledge, it is equally true that his belief is not justified. Except for the difference that Mr. Truetemp’s process involves an implanted “external” device, while characters in Bonjour’s examples have clairvoyance as an innate ability, a difference which might be ignored for the purpose of this type of criticism, these examples are making the same point against reliabilism and against externalism in general.

4.2. Epistemic Justification

The force of these two counterexamples against reliabilist analyses of justification could be differently evaluated depending on how we approach the notion of justification in the epistemic context. Two different approaches are these. First, we may approach the

36) See details in Chapter 3
37) See page 25 of this chapter.
examples with a thought that there is a property called *epistemic justification*, and that we have some idea of what that property should be. Second, we may think that whatever is additionally required for knowledge (other than truth, belief and a possible un-Gettierizer) makes up the property of epistemic justification. That is, in this understanding, what I have named "epistemic support" previously, i.e., the unspecified necessary condition for knowledge, is identified (and re-named) as "epistemic justification." Whatever our understanding of epistemic justification would be, those counterexamples would raise a question against a Goldmanian analysis of epistemic justification. If we read them under the first understanding of epistemic justification, the result of the criticism using the counterexamples would be that Goldman’s analysis of justification is incorrect as an analysis of epistemic justification. The reason would be that epistemic justification that we know of is something other than being produced by a reliable process. With the second understanding of epistemic justification, critics can show that Goldmanian analysis of justification is at least insufficient. The reason would be that there could be something more than being produced by a reliable process in that property required additionally for knowledge ("epistemic support"), and thus in epistemic justification. In my opinion, if critics want us to read the counterexamples with the first understanding of epistemic justification, they are trying to achieve more than they can legitimately show with those examples, although I presume both Cohen and Bonjour want us to read their examples with the first understanding.

Let’s see how the first understanding of epistemic justification affects the reading of those counterexamples. Here, it is assumed that we the readers have some preconceived idea of what an epistemic justification amounts to. Critics might say this in
two ways. Either we already have an idea of epistemic justification closely related with a
deontic concept (the demon victim is doing his “duty” in the given circumstance, for
example) or with a concept of non-culpability (we cannot “blame” the victim),38 or we
have at least some workable pre-theoretic grasp of what epistemic justification is or
should be. The reason we might want to think that we have the former kind of concept of
epistemic justification is that the idea of doing duty and responsibility, thus being free
from a blame, seems to be a common denominator in various kinds of justification. After
all, we use this term in many other contexts. In fact, those “other contexts,” such as our
moral discourses, everyday practical life, our conscious reasoning and decision making,
our act of debating and persuading others, and so on, are considered as typical situations
where justification occurs. And the deontic and non-culpable sense of justification is used
there. Thus, the idea is that epistemic justification, as being a kind of justification, should
respect this sense of justification. If it is too much of a stretch to think that we have a
concrete concept of deontic justification for epistemic justification, it might be claimed
that at least we have some intuition concerning what the property of epistemic
justification is, which is affected by all those “typical” cases where justification is
required.

Now, if we continue to read the examples with this understanding, it explains the
source of the alleged “general intuitive agreement” that we have concerning the
justificatory status of the demon victim beliefs and the clairvoyance beliefs. That is,
critics would claim that we have at least a pre-theoretic grasp of epistemic justification,

38) Some special sense of “duty” and “blame,” namely epistemic sense, would be required.
which is unspecified but nevertheless certain enough to make us say 'justified' to the
demon victim's beliefs, and 'unjustified' to clairvoyance beliefs. Or, in case they
advocate the existence of deontic justification, they would say that we recognize the
deontic justification in the demon victim's beliefs (thus, we say, "justified"), and we
cannot find the same property in the various characters in Bonjour's clairvoyance
examples (thus, we say "unjustified"). Then, the next move from the critics is to show
that all these things are happening without any relation to the reliability of the belief-
forming process. That is, no reliability is present in demon victim case, and reliability is
in fact present in "unjustified" clairvoyance beliefs. What the anti-reliabilists need is the
examples that show that our intuition concerning epistemic justification has nothing to do
with the reliability of the belief-forming process. And this is what the counterexamples
are supposed to show. Our unspecified and unsophisticated intuition that the demon
victim is justified, or the intuition that those clairvoyance beliefs are not justified, would
be considered as the confirmation of the correctness of our preconception concerning
justification and therefore the confirmation of the incorrectness of Goldmanian analysis
of justification. That is, Goldman's account is "wrong" in the sense that it is saying
something (being produced by a reliable process) that is not relevant to the epistemic
justification as if it is epistemic justification. It might be that he is explaining some other
epistemically positive element but not epistemic justification.

It is possible to read the above counterexamples in this way. After all, I would
think that the intuition that exists there favoring the critics of reliabilism is hard to deny
(thus, I am open to the possibility that epistemic justification might be deontic at the level
of our intuition, regardless of its source). However, if this is indeed the punch line of
those counterexamples, there is something clearly unsatisfactory in debates between reliabilists and the critics over those examples. In this reading, critics like Cohen and Bonjour would have to rely on at least the existence of pre-theoretic yet clearly non-reliabilistic intuition concerning justification (since asserting blatantly that justification is deontic seems to be too much). If it is not begging the question against reliabilists who might have the exact opposite intuition as their sincere responses to those examples, it is at least forcing a handicapped race upon reliabilists. Furthermore, there seems to be no room for a compromise if the whole point of the example is to affirm our intuition. And, most crucially, the fact that we have the intuition that favors anti-reliabilists concerning the demon victim's justificatory status does not prove that the preconceived idea of non-reliabilistic justification (deontic or otherwise) is the correct analysis of the epistemic justification.

If the two major counterexamples against Goldmanian account of justification are not merely asking a confirmation of non-reliabilist conception of epistemic justification, they can be read in the following way. Let's not assume anything about the property of justification. Instead, let's agree on rather an uncontroversial fact that something else other than truth and belief is needed for knowledge. Furthermore, stretch this idea a little more so that we can assume that we "see" this property presented in various legitimate cases of knowledge although we don't have any organized answer of what this property exactly is. This is in fact the idea behind making up a term "epistemic support" in order to refer to this yet-to-be-analyzed yet certainly necessary condition for knowledge. Now, according to the tradition, knowledge implies justification. Then, it is possible to identify the role of epistemic support as what epistemic justification is supposed to do. That is,
“epistemic justification” is (and should be) the term referring to the property of “epistemic support” in terms of its role in knowledge. However, we don’t know what the common property of “epistemic support” is, and therefore, we don’t know the property of “epistemic justification” either. If an acceptable analysis of epistemic support is given, we will consider that as an acceptable analysis of epistemic justification, too. This is the second understanding of “epistemic justification” that I mentioned above. In short, epistemic justification is whatever property makes true belief knowledge. And I believe that it is surely possible, if not intended by Goldman himself, to think that Goldman’s analysis of justification is an analysis of justification as the epistemic support. Now, my contention is that the above counterexamples should be read with this understanding in order to avoid the problems of the first understanding. The counterexamples can still show the insufficiency of Goldmanian account of epistemic justification.

What I call “epistemic support” is another pre-theoretic, yet-to-be-analyzed concept although we know that this property is necessary for knowledge. This concept doesn’t have to be limited to being internalistic (non-reliabilistic). That is, we really don’t have much commitment on whether this property should be internal or external, or even any combination of both. So, by saying that Goldman’s account is not about the concept of epistemic justification that some internalists might have in their mind, but about the necessary condition for knowledge (whatever it may be), we can first deflect the objection based on our intuition concerning justification. And, in this understanding, the force of demon victim’s having justification is weakened because it is unclear whether the
justification of the demon victim is the justification as a necessary condition for knowledge. However, Bonjour’s examples are still doing their share of the work. According to Goldman, justification, thus epistemic support, is being produced by a reliable belief-forming process. However, in clairvoyance cases, even if there is Goldman’s alleged property of epistemic support (reliability), yet we don’t think that those clairvoyance beliefs have a property that makes the beliefs to be knowledge. Of course, it does not show that reliability has nothing to do with the epistemic support. However, it does show that reliability is not everything in epistemic support if it is there at all.

In my opinion, treating epistemic justification as a different kind of justification from moral and practical justification is plausible. What is also plausible is that, if a belief is justified, it should be getting closer to the epistemic status of knowledge. Epistemic justification has to contribute to knowledge. Our preconception related with justification in general does not regulate what the epistemic justification should be. Rather, our concept of knowledge, that is, our preconception of what is needed for knowledge, regulates what the epistemic justification should be. Goldman’s notion of epistemic justification is surely of this sort. Knowledge needs truth-conducivity, thus truth-conducivity is the nature of epistemic justification. And the counterexamples are to show that this is at least not the whole story of what is needed for knowledge, thus not the whole story of what epistemic justification is.

39) It is possible that Cohen might be claiming that the property we identify as justification in demon case is a part of epistemic support.
The situation that surrounds the nature of epistemic justification could be summarized in the following way. We might have some intuition concerning the concept of justification. We might be able to tell whether a given analysis is correct on the basis of this intuition. However, even if a given analysis of justification is judged correct by this intuition, this justification is not applicable in epistemic context because it is not the one that provides the epistemic support required for knowledge. However, if we equate epistemic justification with the epistemic support, a yet to be analyzed property that is needed for knowledge, then it is not anchored in our intuition. That is, we still have to grasp the nature of the epistemic support and shape the concept of epistemic justification according to our needs. However, because there is no common intuition other than that it is required for knowledge, the nature of epistemic justification based on epistemic needs is hard to draw agreement. Wayne Riggs reports the situation as the following.

There are internalist justification, externalist justification, coherentist justification, foundationalist justification, deontic justification, consequentialist justification, propositional justification, doxastic justification, personal justification, situational justification, objective justification, subjective justification, cognitive justification, and structural justification. None of these is quite equivalent to another, yet each is proposed as a way in which something can be epistemically justified. Alston also expresses a similar thought in his “Epistemic Desiderata,” and claims that “we should abandon the idea that there is a unique something or other properly called ‘epistemic justification.’”

---

A seemingly dilemmatic situation concerning the concept of epistemic justification may find two different way-outs. On one side, one may stick to the idea that there is a common denominator of justification that we can intuitively agree on and claim that this kind of justification is necessary for knowledge. We can expect that this kind of justification would be strongly internalistic and likely to be deontic. Although it would never be sufficient for making a true belief knowledge, it could be argued that the additional fourth condition for knowledge, which we need anyway, would handle this problem. Thus, together with an externalistic fourth condition, a strong internalistic notion of justification would be still accepted as a necessary condition for knowledge here. If we consider variety of justification, it seems to be a plausible assumption that the traditional sense of justification is internalistic. However, although some epistemologists recently wanted to take this option,42 the success of this position, especially its argument that this internalist justification is necessary for knowledge, is not clear.

I prefer the solution that lies at the opposite end, which also is taken by some other epistemologists. It is to treat the epistemic justification as a mere placeholder. According to this view, whatever property required for knowledge is the content of epistemic justification. So, it's not that there first is the property called epistemic justification to be analyzed, but rather that, at first, there is a special property needed for knowledge, and then, that property is considered as epistemic justification.43 (So, in this


39

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
view, knowledge necessarily implies epistemic justification.) This idea includes the case of abandoning the notion of epistemic justification entirely and adopting some other term for that property. Given that epistemic justification is simply a name for the property that we are looking for, it would be indeed beneficial if we can investigate this property without worrying about our intuition concerning justification. Plantinga's strategy in his warrant theory is one of this kind. Plantinga would agree with the internalist-justificationist idea that justification is essentially internal (and deontic). However, exactly for this reason, Plantinga thinks that justification is not necessary for knowledge, and what is necessary for knowledge should be captured by some other notion, namely warrant. Plantinga's warrant is also a placeholder. That is, it is merely a name for the property that makes knowledge when added to a true belief. And because of that, it is a highly artificial concept, which means that there seems to be no pre-analytic intuition of warrant that puts a limitation on how to shape up the conditions for warrant. Thus, Plantinga can freely put forward a variety of conditions in accordance with what he thinks as necessary for knowledge such as following a design plan, which has no essential relation with warrant. This is one reason why I prefer a Plantinga-type approach. It gives more flexible ways to explore what is necessary for knowledge than holding on to the traditional notion of justification does.

The project of finding the additional conditions for knowledge can be done without being obliged to keep the concept of epistemic justification. Surely, there would

---

4) Concerning Plantinga's theory, see chapter 3 of this dissertation. Also, Plantinga, Alvin, *Warrant and proper Function* (New York: Oxford University Press, 1993)
be some epistemically valuable elements that internalists would not want to give up in their analysis of knowledge, but that does not have to be justification. For the same reason, the epistemic justification or warrant does not have to be purely externalistic given that knowledge seems to allow a certain degree of subjective elements. Then, treating epistemic justification as a mere placeholder gives such flexibility as can accommodate both internalist and externalist insights in order to explain what we think of as a necessary condition for knowledge.

4.3. Epistemic Goodness

If the concept of epistemic justification is a placeholder whose substantial content can be stipulated according to what is really required for knowledge just like Plantinga’s warrant, continuing to use the term “justification” may only bring an unnecessary confusion to the analysis of knowledge due to the generic meaning embedded in the term “justification.” That is, while the idea of flexible coverage of many different but epistemically desirable features is the merit of treating epistemic justification as merely a placeholder or an artificial concept, we also have to consider that we cannot be totally free from what normally considered as the generic sense of justification as long as we continue to use the term “epistemic justification” in order to designate the necessary property for knowledge. For example, can we call Armstrong’s nomological connection, which is claimed to be required for knowledge, a kind of (or a part of) epistemic

\[45\] It is hard to think that Plantinga’s conditions for warrant constitute a “meaning” analysis of the concept.

41
justification? Having Armstrong's nomological connection between truth and what is believed may well be an epistemically desirable property for a belief to have, thus it could be the content of epistemic support. It would upgrade the epistemic status of a mere true belief to something much more epistemically positive, if not yet to knowledge. Even so, I am sympathetic to the point that we would not want to describe it as having (a part of) epistemic justification. Surely, it can be called truth-conducivity. However, truth-conducivity is not equivalent to what is normally thought of as being justified, which is the point that critics of reliabilism want us to see through their counterexamples. Again, reliabilists may claim that such a property is in fact epistemic justification. However, this exactly is the conflict that I think unnecessary. This conflict can be avoided by simply choosing another term, preferably a more general one than "epistemic justification." This is the reason I suggest a different term, "epistemic goodness."

So far, what we have is this. We see that there is an unknown property tentatively called "epistemic support," which is the gap-filler between a true belief and knowledge (in a non-Gettierized situation), and thus a necessary element for knowledge. This property can be called "epistemic justification" no matter how its content is eventually analyzed. However, by identifying this property as a kind of justification, a theory of epistemic justification has the unnecessary burden of satisfying our intuition concerning justification. At least a part of our intuition concerning justification is not a necessary element for knowledge. This situation can be avoided simply by choosing another term. From the fact that "justification" is overused, and from the need of a better term that would replace "epistemic justification," I suggest "epistemic goodness." Thus, in my understanding, "epistemic goodness" and "epistemic justification" refer to the same
property that we tentatively agree to call "epistemic support." The suggested replacement of "epistemic justification" with "epistemic goodness" is only to save the concept from unnecessary confusion so that it can capture the diversity of epistemically positive elements we may find in the property of epistemic support.

However, let me remind you that such a broad coverage of various epistemically positive features does not mean that this concept would cover everything in between a true belief and knowledge. In this sense, my concept of epistemic goodness is not equivalent to Plantinga's warrant. As I mentioned above, Plantinga's main concept of warrant, which he claims as a necessary condition for knowledge, is a placeholder just like the notion of epistemic goodness. It refers to whatever property makes true belief knowledge. However, the problem of Plantinga's warrant is that it intends to take up all the space between true belief and knowledge. In other words, it intends to fill the gap between true belief and knowledge completely. The criticism has been made that it cannot be so. For example, despite Plantinga's confidence in how his conditions for warrant can handle typical Gettier situations, a warranted belief can still be vulnerable to similar types of Gettier examples.\(^{47}\) That is, there are cases in which warranted beliefs fail to be knowledge, which shouldn't be possible if his theory of warrant is correct. In fact, warrant does not seem to imply truth,\(^{48}\) which is against Zagzebski's conclusion.


\(^{48}\) At least, Plantinga does not claim explicitly that this should be the case.
The difference between Plantinga’s warrant and my concept of epistemic
goodness is that epistemic goodness does not intend to fill the gap between a true belief
and knowledge completely, especially regarding Gettier cases. In this sense, we may say
that epistemic goodness is not as artificial as warrant. For example, warrant includes a
condition for an appropriate environment. Therefore, even if the other conditions related
to the agent are met, sometimes the belief cannot get warrant just because the agent is in a
wrong environment. Given that many Gettier cases rely on some factors in the
environment, Gettierized beliefs do not have warrant because of the environment
according to Plantinga. However, I think this kind of treatment of the property of
epistemic support is unnatural. The scope of epistemic support that we see in knowledge
does not include an environment or circumstances that exist outside of the agent.
Epistemic support is what we find strictly in beliefs of the agent. Therefore, Plantinga’s
warrant is not really the same as the property of epistemic support. Thus, the concept of
epistemic goodness, which in my opinion, should capture the property of epistemic
support, does not have to have the antidote for all possible Gettier cases in it. Instead, I
suggest a distinction of *prima facie* epistemic goodness and ultimate epistemic goodness,
which allows for the possibility that epistemic goodness can be undermined or defeated.
This is not possible in Plantinga’s warrant. Warrant cannot be undermined depending on
circumstance. Under such defeating circumstances, warrant just does not exist.

Thus, epistemic goodness takes up a necessary condition, rather than a necessary
and sufficient condition, for a true belief to become knowledge, which is related with how
a belief is formed, grounded, supported, maintained, and so on. One clear advantage of
taking a concept, such as epistemic goodness, that does not borrow too much from its pre-
theoretic meaning is that this concept is specifically neutral in the debate of whether internalism or externalism is correct concerning the nature of what is required for knowledge. Some epistemically relevant elements are not always under the control of the agent, or not always within the awareness of the agent. (For example, success/failure ratio [tendency or history] of one’s belief-forming mechanism in getting true beliefs.) The agent may not be able to access to those elements in the typical way of reflecting one’s “inside.” According to internalists, such elements are not qualified to be the source of justification. However, epistemically positive features that characterize the difference between mere true beliefs and knowledge might be found in those non-intemalistic elements too. In fact, this would be the main reason for externalists to reject internalism. And this is also why some of externalists claim that epistemic justification is in fact external (under the assumption that knowledge implies justification). Now, the notion of epistemic goodness can provide a better ground where compromise is promising. Since my claim would be that achieving a positive epistemic status for knowledge needs both internalistic and externalistic requirements, the notion of epistemic goodness would serve my purpose better.

So far, I have compared five different notions: (i) generic sense of justification that may well be intemalistic and/or deontic, (ii) epistemic justification under a reliabilist interpretation, (iii) Plantinga’s warrant, (iv) the idea of capturing the unknown property of epistemic support using the term “epistemic justification” and using it as a placeholder (because the property of epistemic support is yet to be analyzed) and finally, (v) my suggestion of the term “epistemic goodness,” also as a placeholder for the property of epistemic support. What is measured is whether any of these notions provide a better
structural ground to capture the property of epistemic support, the observed necessary condition for knowledge. The first two notions are too narrow so they can only be a partial explanation of the epistemic support. Plantinga's warrant is too broad because it intends to include a Gettier solution within the concept of warrant. (Moreover, Plantinga's warrant fails to be a Gettier solution.) None of these three is equivalent to the property of epistemic support. Thus, to pick the term "epistemic justification" and to use it as a placeholder is the right way to deal with this issue. However, this approach still suffers unnecessarily because of the influence and limitation tied to the overused concept of justification. So, "epistemic goodness" is my choice in order to refer to the property of epistemic support, a necessary condition for knowledge. I will provide my account epistemic goodness in the last chapter. Before that, I will first examine Alvin Plantinga's account of warrant and Ernest Sosa's account of justification in the next two chapters in order to see whether the property of epistemic goodness can be explained better in either one of these accounts.
CHAPTER 2

A CRITICISM OF PLANTINGA’S PROPER FUNCTIONALISM

1. Warrant And Proper Functionalism

1.1. Conditions for Warrant

In his two books, *Warrant: the Current Debate* and *Warrant and Proper Function*, Alvin Plantinga argues for his theory of knowledge. It is called a theory of warrant, and the basic idea can be titled proper functionalism. According to this theory, knowledge is defined as a warranted true belief. Here, the term “warrant” does not attempt to pick out a particular epistemic property such as justification. Rather, it is introduced as a placeholder. That is, the term “warrant” refers to whatever property makes a case of knowledge when added to a true belief. He then provides a positive account of warrant. In order to be warranted, hence to qualify as knowledge, a belief must be produced by a properly functioning cognitive faculty. Cognitive malfunctioning is a

2) “whatever precisely it is, which ... makes the difference between knowledge and mere true belief.” *WCD*, p. 3.
recurring theme which arises in many of the examples that Plantinga uses to criticizes other accounts. For example, Plantinga argues that his example of the Case of Epistemically Serendipitous Lesion would cause a difficulty for reliabilist accounts such as Dretske's and Goldman's. The upshot of this example is that an epistemic agent who is suffering from an abnormal brain lesion helplessly believes whatever the lesion causes him to believe regardless of the evidence he has. While the majority of those beliefs are false, one of them is a true belief, namely the belief that the agent has a brain lesion. Given that this brain lesion would always cause this belief, this true belief would be considered a belief produced by a reliable belief-forming process, thus satisfying the reliabilists' requirement. However, it's hard to say, at least according to Plantinga, that the agent knows that he has a brain lesion. While reliabilism suffers from this counterexample, Plantinga claims that his account can deal with this type of example easily because this is a case of cognitive malfunctioning, and his account features the proper functioning faculty requirement. A similar diagnosis and treatment is given for the example of the Coherent but Inflexible Climber. The epistemic agent in this example, while he is enjoying rock climbing, becomes non-responsive to experience due to a burst of high-energy radiation. His beliefs are fixed as they were at the moment of the burst, and let's stipulate that his system of beliefs is coherent. Given that none of the beliefs in the system is close to knowledge, Plantinga uses this example to show that coherence is not sufficient for warrant. Again, his claim is that the problem posed by this example is that the belief-forming faculty is not functioning properly.

3) WCD, p. 195 and p. 207. Also WPF, p. 29.
What then amounts to a proper function of any given faculty? Here, Plantinga introduces the concept of a design plan as the criterion for measuring the appropriateness of a function of a faculty. Belief-production is the result of a properly functioning faculty if and only if the faculty is functioning in accordance with its design plan. Asking a faculty to function properly in this sense (following its design plan) is Plantinga's first necessary condition for warrant. However, Plantinga claims that proper function alone is not sufficient for warrant. There are other requirements for a belief to have warrant which, together with the idea of a properly functioning faculty, eventually constitute a set of necessary and jointly sufficient conditions for warrant. According to Plantinga's first approximation, S's belief, B, has warrant if and only if

the relevant segments (the segments involved in the production of B) are functioning properly in a cognitive environment sufficiently similar to that for which S's faculties are designed; and the modules of the design plan governing the production of B are (1) aimed at truth, and (2) such that there is a high objective probability that a belief formed in accordance with those modules (in that sort of cognitive environment) is true.5

Plantinga thinks that sometimes a faculty functioning merely in accordance with its design plan does not produce a warranted belief. One reason is that the faculty may be in a "hostile" environment, that is, a kind of environment in which the designer does not intend it to function. There is some ambiguity here concerning what kind of environmental glitch does or does not affect the sufficient similarity of the given environment to the originally intended one. However, presumably Plantinga has in mind

4) WCD, p. 82 and p. 110.  
5) WPF, p. 19.
environmental glitches like those found in some Gettier cases. For example, if believing the testimony of others is a designed faculty, it would not be designed for an environment in which people frequently make practical jokes on others. Thus, the problem of Smith in the original Gettier example, who comes to believe that the person who will get the job has ten coins, is not that his faculty is not functioning properly, but that it is functioning in a hostile environment. Similarly, we may say that the barn-façade environment, i.e., the environment that contains many perceptual equivalents, is not sufficiently similar to the environment for which our faculty, i.e., perception, is designed.

However, it is possible that a belief-forming faculty is not aiming at truth but aiming for something else, such as survival. Plantinga uses the example of "optimism" that could produce beliefs vital for the survival of the being, a belief such that one's hard-to-cure illness is eventually conquerable. If this kind of "optimism" or "wishful thinking" is a designed feature of the being, then the belief would be produced by a properly functioning faculty. However, even when it is indeed true that the illness is conquerable, such a belief does not have warrant because the segment of the design plan that the faculty is following is not aimed at producing true beliefs. Finally, Plantinga worries about the possibility that a design plan itself is not "good." For example,

---

6) Plantinga explains Gettier cases relying on his environmental requirement in WPF, p. 31–37.

7) For the details of the original Gettier example, see chapter 1, p. 4.

8) However, it seems possible to push Plantinga here by pointing out that the belief formed in a hostile environment (i.e., the environment where his boss is telling a lie) is Smith's false belief that Jones would get the job, not his true belief that the person who will get the job has ten coins. Unless Plantinga wants to claim that any belief formed "in" that environment is unwarranted, citing the environmental problem, at least in this original Gettier case, seems to be awkward and not really on the essence of the problem. I think Plantinga should have mentioned something about transferring warrant from one to the other here. That is, assuming that Smith's intermediate belief [Jones will get the job] is unwarranted, the final belief does not have warrant because it is an inference from an unwarranted belief.

9) For the barn-façade example, see chapter 1, p. 7.
Plantinga considers that "a well-meaning but incompetent angel" may come up with a design plan aimed at the production of true beliefs. Suppose this design plan fails such that most of the time when a faculty is working in accordance with this design plan even within the intended environment, it produces false beliefs rather than true ones. This case satisfies three previous conditions for warrant but we don’t think the occasional true belief produced by this faculty qualifies as knowledge. So, Plantinga requires that the design plan should be good, or reliable. That is, "the objective probability of a belief’s being true, given that it is produced by cognitive faculties functioning in accordance with the relevant module of the design plan, is high."\textsuperscript{11}

These are Plantinga’s four conditions for having warrant. For the convenience of further discussion, I will summarize them in the following.

(1) Properly Functioning Faculty Requirement: A belief is produced by S’s properly functioning cognitive faculty, which in turn means a faculty which functions in accordance with its design plan.

(2) Favorable Environment Requirement: S’s faculty is functioning in an environment sufficiently similar to the one for which S’s faculties are designed.

(3) Truth-aiming Design Plan Requirement: The segment of the design plan that governs the production of a belief is aiming at truth.

(4) Good (Reliable) Design Plan Requirement: The segment of the design plan which governs the production of a belief is reliable.

\textsuperscript{10) WPF, p. 16.} 
\textsuperscript{11) WPF, p. 17.
1.2. Overview of My Argument Against Proper Functionalism

Various criticisms have been raised against Plantinga's theory of warrant. Many have doubted whether his construction of warrant is necessary for knowledge. Using counterexamples, critics attempt to show that some true beliefs qualify as knowledge even though the faculty which generates the belief is not functioning properly in Plantinga's sense. Critics claim that, for example, Pinocchio, the Swamp Man, or some other creatures that are created accidentally (thus neither with design plan nor with the intention of a designer), could have knowledge even though they do not satisfy Plantinga's conditions for warrant. However, the force of these counterexamples is limited. Since Plantinga understands proper function as being equivalent to following a design plan, no faculty would be able to function properly unless it has a design plan to follow. However, is there really no ground that makes sense of the cognitive behavior of Pinocchio or the Swamp Man? Or, in other words, if the existence of Pinocchio is purely unintentional, should we conclude that there is no way to understand "appropriateness" of Pinocchio's various cognitive functions? I think a distinction could be made here on behalf of Plantinga, and that the known counterexamples of this kind need to be re-evaluated. According to the distinction I have in mind, even if a particular faculty does not owe its existence to a particular design plan (or to a designer's particular intention), it may still be the case that there is a general sense of design plan for the faculty to follow. If so, then at least we could discover a ground for making sense of whether Pinocchio or the Swamp Man is "properly" functioning without worrying about the fact that there has been no design plan for him. To be sure, the ground that serves as a criterion for
evaluating Pinocchio’s function is not the design plan exclusively prepared for Pinocchio. However, we may still be able to evaluate why and how Pinocchio is properly functioning. And, if this is the case, it will weaken the impact of the alleged counterexamples because now there is a reason to think that the critics’ point that Pinocchio has knowledge without satisfying the proper functioning requirement may not necessarily be the case.

However, the reason I point out the somewhat unsatisfactory nature of this type of counterexample is not to defend Plantinga's theory. Rather, it is for launching my own counterexample, which I think serves the same purpose better. In my example, in order to show that functioning in accordance with a design plan is not necessary for knowledge, I utilize a designed but malfunctioning faculty rather than using an un-designed faculty. I will show that even a clearly malfunctioning faculty (by not following the assumed design plan of its own) can generate knowledge. My general argument strategy in this chapter will have the following structure.

I will begin with my counterexample against the necessity of a proper function for knowledge. I will then consider some possible replies from Plantinga to restore the necessity between a proper function and knowledge, and to avoid my example. I will argue that they are unacceptable. I will specifically focus on one type of response that seems to be intuitively appealing. It could be true that the success of my example depends on what Plantinga’s idea of following a design plan exactly is. And, given that Plantinga’s explanation on this issue is not so clear, it may be possible that there is a way of understanding the notion of following a design plan and proper functioning that makes
my example unsuccessful. However, I will then discuss why such an understanding that
disarms my criticism is unacceptable. In short, I find that this response is not compatible
with Plantinga’s other essential claim, namely that his four conditions for warrant are
sufficient for a true belief to be knowledge. Here again, I admit that there is a possible
reply to this charge. However, I will show that this reply only makes my initial
counterexample against the necessity of proper function for knowledge more plausible.
So, according to my criticism, Plantinga cannot consistently maintain both the necessity
of proper function for warrant and knowledge, and sufficiency of warrant for knowledge.

In the place where I argue against the sufficiency of warrant for knowledge, I will
note that some epistemologists also have raised similar questions. For example, according
to Klein, some “Gettierized” true beliefs which satisfy all of Plantinga’s conditions for
warrant could fail to be knowledge.\(^{12}\) This would be another interesting case against
Plantinga’s project, particularly concerning the issue of whether a warranted belief could
be false, just as a justified belief could be false. Intuitively, it looks as though there
shouldn’t be any problem for allowing the existence of warranted false belief even in
Plantinga’s theory. However, if so, it seems to be possible to construct an example in
which a warranted true belief fails to be knowledge simply by “Gettierizing” it.\(^{13}\)
Plantinga eschews this in his theory. I will discuss this problem in some detail in the
second part of this chapter. My claim there will be that the difficulty spotted by Klein and
others is not just a matter of one or two counterexamples but a more systematic defect of

\(^{12}\) Klein, Peter, “Warrant, Proper Function, Reliabilism, and Defeasibility,” in Warrant in Contemporary
Epistemology: Essays in Honor of Plantinga’s Theory of Knowledge, ed. Jonathan Kvanvig (Lanham MA:

\(^{13}\) See the general “recipe” for making Gettier example in chapter 1.
Plantinga’s theory. In my opinion, Plantinga does not really consider the reliability of the belief-forming faculty, which I think he should, although he does consider the reliability of the design plan and following such a design plan. I spot a gap between the reliability of a faculty and the idea of following a reliable design plan. My argument will utilize a case of an occasionally successful faculty. My example will show either that there is a true belief which satisfies all four of Plantinga’s conditions for warrant but intuitively fails to be knowledge, or that there is a false belief which satisfies all four of his conditions for warrant depending on what the idea of following a design plan amounts to. In either way, Plantinga’s account would get into trouble.

After that, my attention will shift to Plantinga’s favorable environment requirement, which says that in order for a belief to have warrant, it should be produced in an environment close to the one that the given belief-forming faculty is designed for. This condition also draws a lot of critical attention from many epistemologists including Klein, Feldman and Swain. What is interesting for my purpose is Plantinga's actual reply to the criticism. His reply is based on the distinction between a locally and globally favorable environment, and is intended to answer the question of why two beliefs generated in the same environment could have two different epistemic statuses in terms of warrant. Whether this reply can handle the criticism is not my main interest here. What I think is that this rejoinder, which is supposed to meet the criticism of the environment condition (the charge of arbitrariness), is doing much more (or even something else) than

that. For me, this reply seems to be an active accommodation of the idea that a faculty should function "reliably" in a given circumstance rather than (or in addition to) just following its reliable design plan. This is closely related with the defect that I have pointed out previously. In this sense, I consider this rejoinder as a possible reply to my criticism, too. However, regardless of whether this is Plantinga's real intention or not, I think that the reply overshoots the target. This rejoinder would eventually make Plantinga's theory an unattractive version of reliabilism. My overall conclusion for this chapter would be that Plantinga's proper functionalism is not a successful account of warrant, i.e., the property that is necessary for knowledge, and sufficient for making a true belief knowledge.

2. Is a Proper Function Necessary for Warrant?

2.1. Design Plan and The Swamp Man

Plantinga's first necessary condition for warrant requires a properly functioning cognitive faculty, where a cognitive faculty functions properly if and only if it is working in accordance with its design plan. However, since this implies that the properly functioning faculty has its design plan, we may understand the necessary and sufficient condition for a properly functioning faculty as (A) its being designed and (B) its working in accordance with its design plan. Many have objected that this condition is not necessary for warrant and knowledge. However, what they are mainly questioning is (A). Many counterexamples attempt to show that a faculty which is not designed (thus, the faculty is not working in accordance with its design plan because there is no design plan
to follow) may nonetheless produce warranted beliefs. One of them is the case of Geoffrey whose faculty of detecting magnetic fields is acquired purely by accident to compensate for his loss of vision.\textsuperscript{16} Although visual perception may well be a designed faculty according to Plantinga's point of view, neither the newly acquired faculty (ability of detecting magnetic fields) nor the substitution process can be said to be designed. However, our intuition favors, at least generally, that Geoffrey can have warranted belief (thus, knowledge) about where he is, whether he is going toward an obstacle, and so on, despite the fact that his belief generating faculty is not a designed one.

The same point is made by the example of the Swamp Man\textsuperscript{17} who is created by a miraculous accident, or by the example of Pinocchio\textsuperscript{18} whose having a belief itself is clear evidence of its not following the design plan of Geppetto, the designer. These examples make it clear that the cognitive activities of the Swamp Man and Pinocchio are not designed and not intended. However, as critics attempt to show, from an epistemic point of view, it is hard to see why the fact that they are not designed \textit{per se} would necessarily be a negative property which nullifies any positive epistemic status of the outcome beliefs. After all, according to critics, the Swamp Man and Pinocchio can have warranted beliefs as well as knowledge. I believe that many instances of this type of counterexample provide opposing cases against Plantinga's project.

\textsuperscript{16} WPF p. 29 - 31. Plantinga cites the example from William Hasker's unpublished paper "Plantinga's Reliabilism."
\textsuperscript{17} Sosa, Ernest, "Proper Functionalism and Virtue Epistemology," in Kvanvig, \textit{Op. Cit.}: 253 - 270. Here, Sosa creates a case against Plantinga using Donald Davidson's example from Davidson's "Knowing One's Own Mind" (\textit{Proceedings and Addresses of the APA}, 1986, pp. 443 - 44).
However, as I mentioned at the beginning, it should be pointed out that what the critics are really criticizing here is not the necessity of a proper function for knowledge but rather the necessity of a design plan for proper function. Furthermore, when critics attempt to show that a design plan is unnecessary for a function, they take the following strategy: For the sake of an argument, suppose human beings’ faculties are designed. Critics have to suppose this because, if they don’t, there is no point of looking for a Swamp Man or Pinocchio. If critics cannot tolerate Plantinga’s idea that there are design plans behind human beings’ various cognitive faculties, then just by pointing out that human beings have knowledge, critics would have their counterexample, i.e., a case of a warranted belief without following a design plan. That is, human beings’ warranted true belief would provide enough reason to reject Plantinga’s account. However, I don’t think critics should do that. So, they at least pretend to agree that human beings are designed for the sake of an argument. Then, they move on to find a clear case of un-designed epistemic agent, and show that this un-designed thing (the Swamp Man or Pinocchio) can do whatever the designed human being can do. However, in doing this, the critics have to make sure that what the Swamp Man and Pinocchio lacks is exactly the same sense of design plan they assume that a human being might have. What I will explore here is the possibility that they are not the same. That is, if there is more than one sense of design plan in the way that anything can be considered as being designed, then it might be possible to say that, although the Swamp Man clearly lacks a design plan in one sense, it is unclear whether he lacks every sense of design plan, one of which possibly makes up the idea that human beings are designed.
To see Plantinga’s idea of “following a design plan” as if it must presuppose a theistic understanding of a design plan seems to be unnecessary although Plantinga is eventually committed to this view. Instead of a deity, evolution might be considered as playing the role of “designer.” In this way, the idea of proper function may have the potential to be a general “normative” notion for evaluating the positive elements in belief and cognition, having Plantinga’s theistic understanding of a design plan as one version. However, what I see in the issue of whether the Swamp Man kind of example is a legitimate counterexample that forces Plantinga to reconsider his basic idea of “proper function as following the design plan” is not the distinction of a theistic and non-theistic approach to the design plan. Regardless of whether we are under the assumption of theism, or whether we understand the design plan only as an analogy of much more secular cases, I can detect confusion, or an unnecessary assumption, in the way the critics understand the idea of design plan. By pointing this out, I can develop an argument on behalf of Plantinga that the Swamp Man is not an uncontroversial example of “no design plan to follow.” And if so, this example falls short of a case that shows proper function is not necessary for knowledge. Thus, what I am going to do here is to see whether there could be a less painful way to defend Plantinga's theory from at least the Swamp Man type examples without forcing him to sacrifice too much of what he has to say. I think there is such a way, and this can be done by clarifying an ambiguity that lies in the idea of design plan. Roughly, if there is a certain sense (maybe a generic sense) of design plan

19) For just a possibility, suppose that we could measure the properness of a faculty’s function not by comparing it with its design plan but with the ideal or normal performance of the same kind of faculty. And also suppose that the ideal or normal performance can be set by science. Although I am not sure whether it
other than what the Swamp Man clearly does not have, then the Swamp Man would cease to be a threat to the essential idea of proper functionalism. After that, I will launch my own example to show that Plantinga's account still has a serious defect.

Now, concerning the example of the Swamp Man and similar examples, first note that the critics do not deny that the Swamp Man or Pinocchio have cognitive faculties that are working. For example, the Swamp Man has a faculty that works just like our perception. Their point is that faculties are functioning but there is no design plan which governs (guides) the way they function. Moreover, critics are taking advantage of the fact that the Swamp Man's faculties work just like those in a normal human being. The critics claim that the Swamp Man has knowledge. Intuitively, this is agreeable. However, if our intuition tells us that the Swamp Man can have knowledge, the major influence on our intuition comes from the fact that his cognitive faculties function just like a normal human being's faculties. Unlike the original example of the Swamp Man, for the purpose of this example, critics do not need to claim that the material components of Swamp Man's cognitive faculties are exactly the same as in a person. Donald Davidson, as described in Davidson's original story. Given that he is made out of carbonized tree trunk, his faculties may have a radically different material structure from a normal human being. However, I think they have to maintain at least the functional identity claim between a human being's cognitive faculties and the Swamp Man's counterpart faculties. It is required because, as soon as they admit that the Swamp Man's faculties could be

would be viable, the elimination of a deity (the designer) and the concept of design plan from Plantinga's theory needs to be attempted for those who want to introduce evolution as the other possible designer.
functionally different from ours, they may lose their ground for claiming that the Swamp Man has knowledge. At least, it seems that they would have to go through a long argument for why the Swamp Man knows although he has functionally different faculties. Recall that, for their objection, critics need both that the Swamp Man's faculties are not designed and that he acquires knowledge using these faculties. So, the easy and efficient way for their purpose is to present the Swamp Man example as a case where an undesigned faculty happens to be functionally identical to a designed faculty. (Of course, recall that, for the sake of the argument, we are ready to assume that human beings' faculties are designed.) Now, if the Swamp Man acquires knowledge using this undesigned faculty, would it be a counterexample against Plantinga's theory?

2.2. Two Kinds of Design Plan and Escaping the Swamp Man

Let's see what one can say about this on behalf of Plantinga. In my opinion, it seems that more than one kind of design plan is involved when we normally talk about whether a given faculty is designed or not. And to say my conclusion first, the Swamp Man lacks only one kind of design plan that is relatively less important for maintaining Plantinga's theory. In other words, the Swamp Man who, for example, perceives things just like us can still be considered as functioning in accordance with the design plan of the faculty called perception although every physical aspect of the Swamp Man, namely the Swamp Man as the host (the container) of this faculty, hasn't been designed by anybody.

20) Although Geoffrey's case is little different, we can still say that the way in which Geoffrey's magnetic field detecting eyes work is functionally identical to the way in which some of familiar detecting devices
Let me first try an example that may illustrate my point. Suppose I am conceiving an imaginary device called H-63 which functions in the following way. This device is basically a machine that responds to the object in front of it by playing music. If it detects a dog, it plays a tune of the song “Bingo,” if it detects a girl with a Lamb, then it plays “Mary has a little Lamb,” and the list goes on. Here, what I am doing is designing various “proper” functions of H-63 by assigning my intended correct response to each conceivable circumstance. According to Plantinga, a faculty is a set of ordered triples (or pairs) which regulates a particular response in a given particular circumstance. Then, I think what I am doing concerning H-63 here is equivalent to the “design planning” of the faculty of H-63. However, note that, at this point, I may have no idea of the shape, size and the physical structure of H-63. So, the design of the faculty of H-63 is different from the design of the machine, the physical thing, called H-63. I can design the faculty (by intending a certain function in a certain circumstance) even though I have absolutely no idea of how to actually realize all these functions in a certain physical entity. Making H-63 requires a different level of design plan, maybe a blue print of a sort that we are familiar with. However, the design plan for the physical structure of H-63 is not the same as the design plan for the faculty of H-63. This is roughly the distinction that I want to make here, and I suspect that the Swamp Man counterexample confuses these two.

Suppose that, due to my lack of relevant knowledge, I cannot come up with any design plan for how to make H-63. Suddenly the well-known Q-radiation hits my furnace and other items in my basement, melts them and restructures them. Eventually, as the Q-radiation usually does, there stands a thing that works exactly like my idea of how H-63 (such as metal detector or sonar) do.
works. I have my dog stay in front of it, then it plays the song “Bingo” with sparkling piano solo, and so on. It looks like the spontaneous Q-radiation provides the physical base to actualize the faculty of H-63. It is true that nobody (including me) intends this thing in front of me to work in this way. However, it surely is working in accordance with my design plan of H-63. The question is whether we can make sense of the idea that this thing is working properly. My claim is that we can.

Consider perception. In Plantinga, perception would be a faculty which functions in an X-way, where X-way is the sum of all the proper functions prescribed in its design plan. Functioning X-way is in fact the functional definition of perception. And if Plantinga is correct, it is all intended by the designer just like the way I design (and thus define) the faculty of H-63. If so, I think that anything that functions X-way deserves to be called perception regardless of how the thing comes to function X-way. Also, any functioning X-way should be considered as following the design plan of the faculty of perception.

I am not completely certain that Plantinga means his design plan to be understood exactly in this way. However, this is surely a plausible option to take, and the merit of taking this is to parry the attack from the Swamp Man. Indeed, Plantinga suggests that a design plan is a list of ordered triples of <circumstance, response, function>. This is a list of commands or recommendations from the designer to the faculty. Since it is the designer’s intention, it is what the faculty ought to do (so, here is the source of

21) Still, one might want to say that this thing has no design plan of its own although it may follow a design plan, namely mine. I agree, and this would be in fact my point. That is, a proper functionalism, which is based on the appropriate functioning of a given faculty (provided that we identify a faculty), may not necessarily require a design plan of its own.
normativity). However, at the same time, this list of recommended responses defines the faculty functionally. It defines the faculty as an entity which functions in a certain way in a certain circumstance. A faculty is what it is by doing what it ought to do as intended by a designer in its design plan.

On the other hand, the relatively more familiar idea of design plan is to treat it as if it is a blueprint of a house or an assembly direction for a bicycle. However, this would not really be the relevant design plan for the purpose of proper functionalism. Consider this in the following way. Suppose that there are (1) a particular bed and its actual functions (a physical thing which functions X-way, thus is considered as having the faculty in question), (2) a carpenter’s design plan for making this particular bed (a design plan for making the container of the faculty), and (3) an ultimate design plan of bed which may be equivalent to the functional definition of a bed (a design plan for being that faculty). I suggest that Plantinga's design plan, if it is understood as (3), does not also need to be understood as (2). Through his theory, what Plantinga crucially needs is the design plan by which we can measure the appropriateness of the given performance of the faculty. He does not need to identify this design plan as the design plan required for the faculty-making.

Now, it is obvious that the Swamp Man does not have the design plan for its faculty in this sense. The Swamp Man comes to exist accidentally with all those faculties, that is, without anybody's intention to bring about him in the way as he is. However, according to the distinction that I draw, this only means that the design plan for the physical aspect of the faculty is absent. This is like the case in which a bed pops into
existence in a physical form with all its functions, not being mediated by a carpenter's design plan. If so, at least, the design plan that governs how the faculty ought to perform could still be there. For this, all that Plantinga has to claim is that all proper functions are intended and designed. So, the proper function of the human faculty is designed. The Swamp Man's faculty is functionally identical to our human faculty, and that's why we think that his beliefs qualify as knowledge. Then, the design plan as a functional definition of human faculties could also be the design plan that governs the Swamp Man's faculties. In this way, Plantinga can have room to say that the function of the Swamp Man's faculty is designed while admitting that placing this faculty in the Swamp Man is completely accidental, not intended by anybody.

Thus, the point is that it is possible to think that the faculty of Swamp Man (and Geoffrey, too) is actually following a certain, well-established design plan of that faculty. It is not designed for the Swamp Man, to be sure, but it is functionally identical to the one which is designed for us. If the problem is merely that such a faculty is accidentally located in that creature, then it is possible to say that the Swamp Man is functioning properly, and functioning in accordance with its faculty's design plan. And if Plantinga can say this, the criticism based on the fact that the Swamp Man's faculty has no design plan to follow will not stand any more.

2.3. Plantinga on Reliable but Malfunctioning Faculty

So far, I have explained the unsatisfactory nature of the Swamp Man objection against Plantinga. This objection unsuccessfultly proves that a properly functioning faculty
is not necessary for warrant. What we need is an example which grants that the faculty in question is designed for the sake of an argument. If we can show that a designed but malfunctioning faculty (by failing to work in accordance with the design plan we just assumed to be there) still can produce a warranted belief, this would be a much stronger case against Plantinga's theory than any other criticism. For example, consider color-blindness. Regardless of the issue of whether our visual perception is designed or has been evolved, color-blindness would be a typical case of malfunctioning of our visual perception, and we can expect that Plantinga would agree. If so, according to the properly functioning faculty requirement, a color belief produced by this malfunctioning faculty (color blinded perception) would not have a chance to be a warranted belief. It is important to note the reason why outcome beliefs do not have warrant in Plantinga's account. It is not because color-blind perception produces more false beliefs than true ones concerning certain colors. Unreliability of the faculty is not a primary factor here. According to him, the primary factor is the fact that the faculty in question is not working properly, not following the design plan, and as a consequence, the agent would have more false beliefs than true ones, which in turn makes the faculty unreliable.

Thus, for Plantinga, when there is a malfunctioning but reliable faculty, the verdict from his theory is always that the outcome belief cannot have warrant. We can see this in his way of handling the Case of the Epistemically Serendipitous Lesion, which I mentioned earlier. In this example, a certain kind of brain lesion reliably provides the patient a true belief to the effect that the patient has a brain lesion. As long as the tumor and its ability of belief-production is not a design feature of a human cognitive faculty,
Plantinga’s verdict would be consistent with his theory; this belief does not have warrant because it violates the properly functioning faculty requirement. He thinks that this is the right result and it is a reason to think that his proper functionalism is a better theory than generic reliabilism because reliabilists have to accept this reliably produced true belief about the lesion as knowledge, which is counter-intuitive.

However, it is unclear whether reliabilists really face a difficulty even if they affirm this as a case of knowledge. The key feature which, I think, guides our intuition concerning this kind of example is whether this malfunctioning faculty (a cognitive system contaminated by the brain lesion) produces only the true belief (that the patient has a brain lesion) or whether it also produces a lot of wacky, mostly false beliefs and a true lesion-belief is just one of them. It would be the latter scenario which makes a rather obvious case of unwarranted belief, and Plantinga does set up the case in this way in order to induce more intuitive agreement that it is not a case of knowledge. However, if it is the former, our intuition does not obviously favor the same verdict, at least concerning knowledge. Suppose that the following is a truth discovered by medical science; people have a belief about the brain lesion type A if and only if their cognitive system is malfunctioning due to the intervention of brain lesion type A. That is, having such a belief is a sure indicator of having such a lesion, and having such a lesion is most likely to be the only way for having such a belief in this world.22 Then, it becomes unclear why we should deny that the agent knows that he has the brain lesion type A, even if the brain lesion is still considered as a malfunctioning of the faculty. It might be expected that the

67
reliabilist theory of justification would probably have difficulty saying that such a belief is justified. However, for those who have externalist intuitions concerning knowledge, the knowledge claim wouldn't be really counter-intuitive. Indeed, in this scenario, the brain lesion belief is a reliable indicator of truth. For externalists like Armstrong, this would be enough reason to admit that the victim has knowledge because it works just like the reliable thermometer reading in his analogy.

Even in the cases where a malfunctioning faculty is supposed to produce many false beliefs, sometimes this negative consequence of malfunctioning (having false beliefs) can be systematically removed due to other epistemically relevant features. In this way, suppose that we can come up with reliable malfunctioning. Then, concerning a belief generated by such a reliable but malfunctioning faculty, Plantinga would insist that we should evaluate this belief negatively (unwarranted) on the basis of the given faculty's malfunctioning, rather than evaluating it positively on the basis of its being reliable under the given environment. However, this is not so obvious.

The case of a reliable malfunctioning faculty seems to be one way to see if proper function is really necessary for warrant. What I am going to explore in the following is based on this idea. Specifically, I want to consider a possibility that, if some extremely favorable environment systematically compensates for the epistemically negative aspect of a malfunctioning faculty, a true belief produced by a clearly malfunctioning faculty may qualify as knowledge.

22) Let's suppose that, due to our other cognitive features teamed up with the extremely complicated nature of type A lesion belief, it is impossible to have the belief of the same content via other belief-forming
2.4. Proper Function is Not Necessary for Knowledge: My Example

Let us consider a fictional type of defective color perception. Suppose that Kim is exactly the same as a normal epistemic agent in every aspect except that she cannot perceive colors as normal people do. Kim sees everything as black-and-white, thus colored things appear to her as various shades of gray. For example, a certain shade of red appears to her as if it is dark gray, a certain shade of yellow appears to her as if it is bright gray, and so on. As a result, Kim cannot tell, for example, an actual dark gray thing from a red colored thing that appears to her as the exact same shade of dark gray of the actual dark gray. Although it might be hard to tell exactly which faculty is malfunctioning there, it is certain that some part of her faculty is not working properly. At least, we can say that her visual perception as a whole is malfunctioning, if that is a proper way to individuate the faculty. According to the design plan, she ought to have two different percepts depending on whether she perceives red or gray. By having just one kind of percept in both cases, her visual perception is working against her design plan. She believes every red thing is gray and she also believes that every gray thing is gray. Our intuition tells that she does not know the color of what she is seeing when she believes it is gray. Even if it really is gray, it cannot be knowledge given that there are red things around her. It would be just like the famous barn facade case. If Kim does not have a discriminative power between gray and red, and if she has neither meta-level awareness of this fact nor a method to compensate for the lack of such a discriminative power, then none of her beliefs about the color gray are warranted even when they are true. Concerning this, Plantinga would say that the lack of a discriminative power between red and gray is a mechanism.
case of cognitive malfunctioning. There seems to be no particular reason to doubt this. Then, this would be Plantinga's reason to think Kim's gray-belief has no warrant and she does not know.

This is a typical case of unreliable malfunctioning. However, what if there is no such thing as red within the environment in which Kim is living? Suppose she is a citizen of a fictional place called "Monochromatic World," where every thing is in fact black-and-white. At least in this environment, Kim's unreliable and malfunctioning faculty always produces true beliefs. This faculty is no longer unreliable in this environment. However, the fact that it is malfunctioning has not been changed. Now, does she know when she believes that what she is seeing is gray?

Since she has a malfunctioning faculty, and that faculty produces the belief in question, in Plantinga's account, there wouldn't be any chance to make this belief a warranted one. However, the truth is that, although she does have a malfunctioning faculty, this faculty never has produced a false belief due to the favorable environment. Suppose that Monochromatic World is her whole world. In this world, it is only theoretically possible that there are colors. Nobody has ever experienced them. Still, other people's faculties are working properly (as they are designed) exactly because, if they would see red, they would react differently than when they see gray. (This is the design

---

23) There would be several different ways to make sense of this monochromatic environment depending on how we understand colors. We can say that everything appears to be black-and-white to the normal people in this world due to some particles in their air that disturbs sunrays in a special way. Then, the difference between Kim and other people is that Kim would never be appeared "color-ly" even without the disturbance by the particles, while other people would if there were no such disturbance. Or maybe we can imagine the involvement of a deity. In any case, Monochromatic World is where the grass is gray is true just in the same way that grass is green is true in this world.
Kim’s faculty is not a properly functioning faculty because she would not notice the difference.

Before we evaluate this case, let's consider one more analogous scenario. Suppose that our visual perception is indeed all designed by a designer. Since the designer knows all about the epistemic havoc presented by the evil demon scenario, he equipped us with a kind of screening function as a part of our system of visual perception. Suppose that, somewhere in our perceptual process, we all have a screening function for detecting a demon-generated visual image. In normal cases, this screening function allows what is perceived to pass through. (That is, 'do nothing' is what it ought to do in normal situations, and as a result, we have what we normally call perception and perceptual beliefs.) However, if an evil demon-generated percept attempts to pass this screening process, this detector would warn the perceiver by seriously distorting the percept, thus indicating that what she is perceiving is a demon’s trick. Meanwhile, the powerful and benevolent designer also patrols the environment so that no evil demon can sneak into this world and play tricks on us. As a consequence, no demon has attempted to deceive us (therefore, none of our detector screening function has been activated yet in the way that it distorts the image) and it is very unlikely to be otherwise in the future. Now, a person in this world named Victoria also has this screening function, but only a defective one. It would not distort the percept even if it is generated by an evil demon. So far, Victoria's visual perception which includes this defective detector has provided true beliefs due to the demon-free environment. Still, it is certain that Victoria's visual perception, unlike those who have a properly functioning detective screen, is not working properly because,
as I stipulate it, distinguishing demon-generated percepts from normal ones is a part of the design plan of our visual perception. Now, the question is again whether a true belief generated by this faculty in this environment could be a case of knowledge.

I think, in both cases (Kim’s true gray-belief in Monochromatic World and Victoria’s true belief generated by a defective demon-proof visual perception in a demon-free environment), there is a sense in which they know. One might disagree by saying that this is just a matter of intuition. Those who have a more rigorous standard for knowledge might disagree with this intuition. They might say that it is only a case of incredible coincidence, similar to Gettier cases, which does not sit comfortably with the notion of knowledge. However, for them, I think we could add the following features in this picture as a gesture to remove this element of accidentality. Suppose Kim thinks in the following way: “Doctors find out that I have a malfunctioning faculty. They say that, according to my design plan, I am supposed to distinguish a certain hypothetical color, “red,” if I see it. However, they also say I would be just fine as long as I am living in here. So, I just keep believing as I see.” Then, in the next instance, Kim sees a gray cat and believes that the cat is gray. In this case, I think we are much more comfortable saying that Kim knows the color of the cat. There no longer is a sense of coincidence that happens outside of Kim and Kim has no awareness. However, I think that two facts haven’t been changed. First, the belief in question is produced by a visual perception. Even if Kim is an ultra self-conscious person concerning what she believes, it would still be the visual perception, rather than inductive reasoning for example, which causally generates the belief that the cat is gray. Even if this belief generating process essentially includes the endorsement by
inductive reasoning, still at least a portion of this process is visual perception. Second, Kim is unable to distinguish gray and red contrary to what she is supposed to do according to her design plan. Thus, the segment of her visual perception that is either entirely or partially involved in the production of the belief is malfunctioning. Now, if we can agree on the steps and assumptions so far, it looks like we have a case in which a belief produced by a malfunctioning faculty is a case of knowledge.

2.5. Possible Replies

Several replies seem to be possible on behalf of Plantinga. First, one might question the assumption that the belief in question is produced by Kim’s visual perception. There in fact is a similar type of reply when Plantinga considers the case of Geoffrey.24 He might want to give us the same reply in the case of Kim or Victoria. That is, Plantinga may insist that, in fact, the correct intuition is that Kim and Victoria do not know and their true beliefs do not have warrant due to the malfunctioning faculty. (It is due to the un-designed faculty in the case of Geoffrey.) Plantinga would continue that, despite this fact, the reason why we would like to think of these cases as knowledge is because of the induction or the additional reflective thoughts for the truth of what they believe. However, this type of reply is inadequate at least on two different grounds.

First, it is surely questionable whether our intuition tells that they do not know if there is no help from induction. In Geoffrey’s case, we could be reluctant to say that Geoffrey’s first belief produced by the magnetic field detecting eyes has warrant given

24) WPF p. 29 - 31.
that it is a gradual replacement of normal visual perception. However, concerning Kim’s or Victoria’s belief, it would be hard to deny that they know even if the extra help they get from induction or reasoning is not there. Their cognitive system has been working in the same way all the time. What is being added by the reflective thought is the understanding or the background knowledge concerning how their cognitive system works. So, in this way, it may contribute to eliminate a sense of accidentality by providing a meta-level awareness of what is going on. However, it does not change the party responsible for the belief production. In this sense, the original, non-reflective Kim and Victoria are similar to children who have no reflective thoughts about why their perceptual beliefs are true. And if so, it is hard to deny that they know.

Second, in this reply, the status of the beliefs which are produced later with assistance from induction is unclear even if we admit that the additional thoughts play the role of justifier. As I mentioned above, it wouldn’t be induction which produces these beliefs. Even Plantinga cannot claim that induction is solely responsible for producing these beliefs. It is Geoffrey’s magnetic field detecting faculty which produces the beliefs in question regardless of whether they are produced earlier or later. And, it is not a designed faculty. Then, according to Plantinga’s account, even with the help of induction or whatever additional reflective processes, these beliefs cannot have warrant. However, it is the whole point of this example that this verdict does not fit with our intuition. Concerning Geoffrey’s beliefs produced later, we want to say that they are cases of knowledge, and even Plantinga finds that this intuition is extremely hard to challenge. The same would be applied to the cases of Kim and Victoria. The fact that their beliefs
are produced by improperly functioning perception is the same. Then what would be the explanation for the fact that they suddenly know when induction is added? Unless Plantinga can explain this within his account, this type of reply may not be able to save him.25

The other possible reply is to deny the other assumption that Kim or Victoria’s beliefs are produced by a malfunctioning faculty. Then, what could be a ground for saying that Kim’s faculty is properly functioning? Here, the reply would be based on the intuition that Kim’s color perception is not defective at least in Monochromatic World. Since her gray-perceiving function is doing exactly what it is supposed to do while she is in a Monochromatic World, what produces her gray-belief in the Monochromatic World is a properly functioning faculty. This is surely a possible reply. One might even think that this reply is so obvious that it would immediately block my examples to be a legitimate case against Plantinga. However, I think this reply, although it looks promising, leads to an unacceptable result.

First, it is suspicious whether this reply is taking advantage of the conveniently evasive nature of a proper function of any given faculty. This reply also relies on the somewhat arbitrary nature of faculty individuation. In short, if visual perception is the unit of a faculty, and if it is more or less a job of science to tell whether the given faculty is working properly or not, then there hardly would be room to claim that Kim’s faculty is working properly. Suppose that we describe how Kim’s faculty works using a set of <circumstance, response> pairs and compare it with its design plan (following Plantinga’s

25) Plantinga actually says that his theory needs a chisolming on this matter. I am not sure whether it would
rough guide line of what that is). Here, she indeed has a different ability or capacity for perceiving colors. The description of Kim’s faculty does not contain a pair like <red, red response (or at least something distinguishable from gray response)> while the design plan prescribes this pair as the way she ought to respond. If showing ‘red’ response when the ‘red’ circumstance obtains is a part of the design plan of the faculty in question, the faculty that does not (and cannot) respond in this way is not a properly working faculty. Of course, this fact could be shown only where some objects are red. It is just that, in Monochromatic World, the circumstance ‘red’ cannot be obtained due to the environment, so the way in which Kim’s gray-perceiving faculty works is indistinguishable from the way in which other people’s faculties work. However, whatever the nature of the faculty is, if two faculties are different in the sense that one of them is unable to perform what the other can do, it can’t be that both are following a design plan (thus functioning properly) in Plantinga’s sense.26

Putting aside the murky nature of the faculty, we may see this issue in terms of the environment that the designer has in mind. By hypothesis, the intended environment for Kim and other citizens is the environment that contains colors as well as black, white and gray (call it E). An environment that contains only black-and-white could be a sub-environment of E (call it e). Then any faculty that works properly in E also works

be a little chiseling that he needs here.

26) This is in fact another place to question the nature of Plantinga’s proper functioning faculty requirement. We can surely suppose a faculty that is doing more than what it ought to do, thus violating the “working in accordance with its design plan” condition. For example, an autistic savant may have extra-ordinary ability of memorizing telephone numbers. Given that such a high power memory in a limited subject is not the part of the design plan of human beings (after all, almost none of us seems to have such ability), the autistic savant’s telephone number beliefs are the result of a malfunctioning faculty, i.e., a faculty that does not follow it’s design plan. However, it is unclear how we can treat those beliefs automatically as unwarranted ones because of that.
properly in \( e \). However, if a faculty is not working properly in the environment that the designer has in mind (\( E \)) like Kim’s faculty, there is no reason in Plantinga’s account to think that bringing it to one of its sub-environments could change it into a properly functioning faculty. By confining a faculty’s working environment to one of the sub-environments, we may be able to avoid the consequence of malfunctioning (producing false beliefs). However, it wouldn’t be right that whether a faculty produces true or false beliefs is the criterion for the proper functioning. What seems to be the right description here is that a malfunctioning faculty (by some other independent criteria) is avoiding some of its typical consequences (producing false beliefs) in \( e \). It is not that a faculty which is working improperly in \( E \) begins to work properly in \( e \).

We can generalize this issue using the idea of multiple functions. It is natural to assume that the design plan of any moderately complex faculty or module prescribes more than one type of response to the faculty in changing circumstances. That is, according to the design plan of a faculty \( F \), \( F \) is supposed to respond in \( X \)-way under the circumstance \( A \), while responding in \( Y \)-way under the circumstance \( B \). Then surely it is possible that \( F \) is functioning in \( X \)-way under \( A \) while \( F \) is unable to function in \( Y \)-way under \( B \). Then, even while this faculty is functioning in \( X \)-way under \( A \), isn’t it nevertheless a malfunctioning faculty? When a faculty has more than one function, being unable to do one of them gives the reason to think that the faculty is malfunctioning. However, there are cases in which this malfunctioning has nothing to do with whether the agent has knowledge through this faculty, provided that the malfunctioning faculty is
reliable. Thus, my examples are constructed on the basis of this idea, and they show that having a properly functioning faculty is not necessary for knowledge.

However, the above generalization seems to give Plantinga at least one very tempting and seemingly clear way-out. He can just deny my essential claim and say that, when F is functioning in X-way under A, it is working properly as long as that is what the design plan says. F’s not functioning Y-way in B, which F is supposed to do, does not affect proper functioning of F in A. F would be malfunctioning when it functions in Z-way under A, where it is supposed to function in X-way. He might also want to remind us that his account only requires that "the segment" of the faculty that produces the belief be functioning properly. He could say that at least “the segment” of the faculty which is used for producing Kim's gray-belief in Monochromatic World is the one working in accordance with its design plan, and that would be a sufficient reason for saying that Kim’s faculty is working properly in Monochromatic World, just like any other citizen’s faculty.

However, in my opinion, this seemingly simple way-out of the problem generates another significant problem in Plantinga. This is the argument that I would like to develop through the next section. If I apply the basic idea of this argument to the case of Kim and see the conclusion in advance, it would go like this. If Plantinga's explanation for why Kim's faculty is functioning properly is because the “segment” (namely, gray perception as a segment of perception) is working as it is designed, it follows from this explanation that proper function is achieved whenever any <circumstance, response> pair in the design plan is satisfied. (It is true of Kim that <gray object, gray belief> is satisfied
whenever Kim has gray belief in the Monochromatic World.) However, for the same reason, gray belief of Kim in our normal world where there are colored things should also be considered as knowledge whenever it is in fact a correct response to gray. We know this cannot be right. If it were, even a true barn belief in the barn façade environment would have to be a warranted belief. Therefore, Kim’s lack of discriminative ability between gray and red should make all of her gray beliefs not a case of knowledge in a normal world. Thus, the understanding of proper function that makes Kim’s faculty a properly functioning one cannot be consistently maintained. Now, let’s see how this is so in detail.

3. Are Plantinga’s Four Conditions Sufficient for Warrant?

3.1. Are Occasional Successes Enough for Proper Functioning?

Imagine that the visual perception of Mr. Magoo, a wacky myopic, is defective in the following way: many things, which can be easily distinguished by a normal person’s point of view, appear to him exactly the same. For example, a golf ball, a light bulb, a piece of cotton ball and a cat toy (a white rubber mouse) all look exactly the same to him within a range of perceiving circumstances. If he were a normal myopic who is aware of his inability, he would have withheld believing anything. However, being unaware of his inability and being wacky, he just forms a belief on the basis of his percept. That is, even if the percepts appear to him more or less the same, he sometimes believes that it’s a golf ball and sometimes believes that it’s a rubber mouse. If so, using

---

27) Mr. Magoo here is not exactly like the well-known cartoon character.

79
this belief-forming mechanism, Mr. Magoo acquires many false beliefs. His faculty is unreliable. Now, consider one of Mr. Magoo's perceptual beliefs, a belief that there is a fly on the wall. Suppose it's true that there indeed is a fly on the wall. However, Mr. Magoo does not have a discriminative power to distinguish between a fly on the wall and a nail head on the wall. They look exactly the same. It's an epistemic fluke that he has a true belief, no matter how sure he is about his belief. (The lack of discriminative power in this domain is an objective fact of which he is unaware.) Then, this belief cannot be a candidate for knowledge. If so, given that Plantinga's warrant is what makes a true belief a case of knowledge, this belief also lacks warrant. Now, by applying Plantinga's account, let us evaluate how exactly this belief violates (if it violates at all) Plantinga's conditions for warrant.

If Mr. Magoo and normal persons like us have the same design plan in terms of our visual perception, it would be neither the third (truth-aiming design plan requirement) nor the fourth requirement (good design plan requirement) that Mr. Magoo's unwarranted perceptual belief is violating. It is implausible to think that Mr. Magoo's faculty is working in accordance with a unique but very unreliable design plan of his own. Rather, we should think that Mr. Magoo and those who have normal eyesight both share the same design plan for their faculty of visual perception, namely the design plan of human visual perception. It is natural to assume that what my visual perception ought to do is the same as what Mr. Magoo's visual perception ought to do in terms of its design. The difference is that my faculty actually does what it ought to do most of the time, while
Mr. Magoo's faculty very often fails to do what it ought to do. Both faculties have the same design plan. What is different is the way in which each faculty follows its design plan.

Also it would be implausible that all of Mr. Magoo's unwarranted true beliefs are produced under a hostile environment unless the environment is completely relativized to Mr. Magoo's point of view. We need not say that, since Mr. Magoo has no discriminative power between nail heads and flies, the environment which contains both of them is hostile to him. That would be like saying that the normal world is hostile to a color blind person. Then, the most plausible explanation would be that the belief is produced by a faculty which does not work in accordance with its design plan. That is, this belief violates the first requirement.

However, my point is that, with the understanding of a proper functioning from the last section, we cannot make sense of the conclusion that we have just reached. That is, even though it is tempting, Plantinga cannot say that Mr. Magoo's visual perception (as a whole) is malfunctioning. Taking this option returns us to the issue of multiple functions, and it supports my previous examples of Kim and Victoria, which were intended to show that proper function is unnecessary for knowledge.

In order to argue my point, let's begin with whether we can make more sense of the forced conclusion of the Mr. Magoo case, the conclusion that Mr. Magoo's true perceptual belief that there is a fly on the wall does not have warrant because it is not produced by a properly functioning faculty. Suppose that the difference between Mr.

---

28) If we really have to think like this, proper functioning would become an empty and arbitrary notion. It is because now any faculty, even a very unreliable one, can be made as if it is a properly functioning faculty
Magoo and me is the lack of such a discriminative power between a fly percept and a nail head percept. I have one but he doesn’t. According to the design plan that we have assumed, we all should have one. Thus, I am working in accordance with the design plan, and he is not. Therefore, even if he has a true belief that there is a fly on the wall, it’s the result of a malfunctioning faculty in just the sense that I mentioned. This is exactly the same explanation that I have offered in previous sections 2.4 and 2.5 of this chapter to explain why Kim’s faculty is malfunctioning even in a Monochromatic World. So, if this is a correct way of understanding Mr. Magoo’s malfunctioning faculty, it shows that the above conclusion is correct, but it opens the way to see how Kim’s faculty is malfunctioning. In short, if Mr. Magoo is malfunctioning, so is Kim. And if Kim is malfunctioning, then it shows that there is a case in which a malfunctioning faculty can generate a warranted belief if some favorable environment guarantees to compensate the (bad) consequence of the malfunctioning faculty.

In order to avoid this consequence, let’s follow the intuition that Kim’s gray-perceiving faculty is not malfunctioning but functioning properly in the Monochromatic World. As I have considered it, the ground for this is that the belief is produced by the activation of the function <gray, gray response>, which is a design-plan-following-segment of a faculty. We can generalize this idea in the following way: A faculty under a circumstance C and responding R-way at t is functioning properly (by following its design plan) at t iff responding R-way under C is prescribed in its design plan. However, in this case, Mr. Magoo can be considered to be functioning properly too whenever his true belief is the result of activation of a function <a fly on the wall, a fly on the wall following some unreliable design plan, and if so, warrant would be all about having reliable design plan.

82

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
response>, which is a design-plan-following-segment of a faculty. Then, contrary to the above conclusion, Mr. Magoo’s fly on the wall belief is produced by a properly functioning faculty whenever it is produced by a fly on the wall input. Since all four of Plantinga's conditions are satisfied, Mr. Magoo's belief should have warrant. However, it wouldn’t be right that every true fly on the wall belief of Mr. Magoo has warrant, given that a nail head on the wall input could also easily generate the fly on the wall belief in him.

The whole situation looks like a dilemma. And I think that a way of finding the source of this difficulty is to generalize Mr. Magoo’s case. I think that Mr. Magoo is one instance of cases where a well-designed faculty is only “occasionally successful” in following its design plan. Let me use another analogy. Consider the unreliable ignition system of my 89 Ford Festiva. The dilemma I find in Plantinga’s proper functionalism concerns what we should say if the ignition system of this car ignites the engine only one out of three cases. Again, it’s not the problem of the design plan because all 89 Ford Festiva’s share the same design plan, but only mine ignites the engine so infrequently. Maybe we can even say that all four cylinder combustion engines have the same design plan in terms of their function (or faculty) called the ignition system. It’s only my car’s ignition system that has such a problem. So, it is not the case that my car's ignition system is following an unreliable design plan. Rather, what matters would be how it follows this reliable design plan. If my car, a lemon, frequently fails to follow the design plan, one intuition would tell us that the ignition system of my car, the faculty, is malfunctioning as a whole because igniting the engine only one out of three times is not the design plan of
the system. However, by taking this option, we should realize that we are admitting that some successful ignitions (33.3%) are also the result of the malfunctioning ignition system. However, at this point, our intuition pulls in another direction. Since igniting the engine when you turn the key is the design plan of the ignition system of my car, why not say that it is functioning properly whenever it ignites the engine, no matter how rarely it happens? Isn’t it intuitively true that my car functions properly from time to time?

These are the two options that I think we can take when we talk about the proper function of a faculty using ‘whether it works in accordance with its design plan’ as its evaluative criterion. It could be either for evaluating a faculty as a whole (concerning its general performance and the general tendency) or for evaluating each and every individual performance of the faculty. Then, we have to either admit that there is a successful ignition by an improperly functioning faculty (or unsuccessful ignition by a properly functioning faculty) or claim that any successful ignition is the result of a properly functioning faculty. Neither case is problem-free. Based on the first understanding, we can generate a case in which proper functioning is not necessary for warrant. Kim’s belief is knowledge in a special circumstance even though it is produced by a malfunctioning faculty. Furthermore, on this understanding, Plantinga’s theory allows a warranted false belief, which is the subject of yet another criticism. Based on the second understanding, we can generate a case of unwarranted belief that satisfies all

---

29) There is an argument that warrant should entail truth. Merricks provides the following reason for why warrant entails truth, and therefore, why warranted false belief should not be allowed in Plantinga’s theory. If warranted false belief Bp is possible, it could transfer its warrant to some other belief Bq. (For example, a valid deduction from Bp to Bq done by a properly functioning faculty in a normal environment would be able to transfer warrant.) Then, just like the Gettier case, if Bq is accidentally true, we would have an accidentally true warranted belief. However, this concept is a contradiction. Accidentally true belief can
four of Plantinga’s conditions. Some of Mr. Magoo’s unwarranted beliefs can satisfy all four conditions. Especially, contrary to our expectation, his unwarranted true belief satisfies the properly functioning faculty requirement under this understanding.

It wouldn't be Plantinga's intention that a faculty's occasional success in following its design plan is considered as a proper function of the faculty. However, a faculty is not identical to its design plan. A faculty is an independent entity that may or may not follow the design plan (thus making sense of the idea of malfunctioning). Unless Plantinga puts a regulation on how well the faculty should follow its design plan, the idea of 'occasional success in following the design plan' would keep bothering his account.

3.2. Where is the 'reliable faculty' requirement?

A faculty that is occasionally successful in following a reliable design plan is not reliable. However, curiously, in Plantinga's account for warrant, there is no explicit requirement that the belief should be produced by a reliable faculty. Instead, he has the requirement that the belief-generating faculty should follow the reliable design plan. However, what I will argue is that these two are not the same. The reliable faculty requirement is the major requirement of other reliabilist theories such as Goldmanian process reliabilism. Of course, it is not that Plantinga should be understood as a version of process reliabilism. Even so, those who think of proper functionalism as a version of reliabilism seem to think that Plantinga’s 'properly functioning faculty in an appropriate

never be knowledge, while warranted true belief is knowledge by definition. Therefore warranted false belief is impossible. (Merricks, Trenton, "Warrant Entails Truth," PPR 55(1995): 841-855.)

85

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
environment’ amounts to a reliable process. We might even grant that if a faculty satisfies all four of Plantinga’s conditions for warrant, then the reliability of the faculty comes with it.

It is true that both Plantinga and other reliabilists are apparently aiming for the same thing. Surely, most of the beliefs produced by a proper functioning faculty in an appropriate environment, where proper function is functioning in accordance with a truth-aiming reliable design plan, will be true. This is Plantinga’s idea. And this is exactly what the reliable process theory wants to secure. However, the problem is that even an unreliable faculty (a faculty which produces more false beliefs than true ones) can be a faculty which functions in accordance with a truth-aiming reliable design plan in an appropriate environment.

In order to see this, consider what would be the best reason for thinking that Plantinga’s proper functionalism does require the reliability of the belief-producing faculty. No doubt, there are two requirements for the reliability of the design plan. Particularly, a design plan should be reliable (condition 4), as well as truth-aiming (condition 3). Plantinga’s motivation in adding these two conditions is to prevent a poor or unreliable design plan from being a design plan for our warrant-generating faculty. For example, as Plantinga mentions, an incompetent angel with a good intention (therefore, aiming for truth) may come up with an unreliable design plan. If so, following this design plan would generate more false beliefs than true ones. However, since

30) For example, see Steup’s understanding in his "Proper Functioning and Warrant after Seven Vodka Martinis," Philosophical Studies 72 (1993), p. 90.
31) See page 5 in this chapter.
32) WPF p.17
following the design plan is a definition of proper function for Plantinga, a belief acquired as a result of acting in accordance with such an unreliable design plan in an intended environment meets three conditions (proper functioning, environment and faculty aiming for truth) for warrant. Plantinga thinks that this belief does not have warrant due to the unreliability of the design plan. So he adds condition (4) in order to block this case. Plantinga says, it is necessary that "the design plan is a good one" and that "[T]he modules of the design plan governing the production of B are ..... such that there is a high objective probability that a belief formed in accordance with those modules (in that sort of cognitive environment) is true."34

However, if what Plantinga wants to secure with these conditions is the reliability of a faculty, then I think he fails. Asking a faculty to work in accordance with a truth-aiming reliable design plan is not the same as asking it to be reliable. The critical point is this: Whether or not a faculty is following its design plan is an evaluation of the performance of that faculty at one point in time.35 However, a faculty works through the time, and the reliability is an evaluative term for this overall performance. Even if a faculty works in accordance with a truth-aiming reliable design plan in some instance (thus, provides a true belief this time), if it fails to work the same way the next time (providing a false belief), then the actual reliability of this faculty would be 50%. If so, in

---

33) WPF p.17
34) WPF p. 19
35) It is not necessary to understand "following the design plan" in this way. One could say that a faculty is either following or not following its design plan even if it is not working at this moment (because no circumstance to respond is acquired) or even it hasn't produced any belief yet (because the faculty is just made). However, again, this understanding supports the footing of my previous counterexamples, the cases of Kim and Victoria. What brings us up to this point is the assumption (for the sake of an argument) that
order to have a reliable faculty requirement in Plantinga's theory, we should ask a faculty to work in accordance with a truth-aiming reliable design plan "most of the time," or at least more than half of the times the faculty is used. However, Plantinga's conditions only require that a faculty should follow the design plan. They do not say how often a faculty should follow the design plan. All he can say is that failure in following the design plan causes unwarranted beliefs while success in doing so would provide warranted beliefs. However, a faculty which usually fails to follow the design plan may not be able to provide warranted beliefs even when it occasionally succeeds in following the design plan.

Some might think that a design plan of a faculty includes the overall performance of a faculty. That is, a design plan of a faculty is not just “do X under A,” but more like “reliably, do X under A.” Therefore, the occasionally successful faculty is not working in accordance with its design plan. However, this is implausible. Even if “reliably, do X under A” could be a design plan, the reliability we are talking about here need not be 100%, as we can see in Plantinga’s good design plan requirement. The requirement only requires high reliability. Consider a design plan <gray, gray response except for every 10th case>. This design plan is highly reliable, and captures the spirit of “reliably, do X under A.” If S forms a belief using a faculty working in accordance with this design plan, then in 9 out of 10 cases S would have a true belief. Of course, once in a while, it produces a false belief, but that wouldn’t be a problem for being a reliable design plan. Thus, the false belief is still produced by a properly functioning faculty which follows a

Plantinga would be able to avoid Kim and Victoria cases by an alternative understanding of “following the design plan.” Here, I am working against such a possible alternative.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
reliable design plan. Then, this false belief has as much warrant as the other true beliefs produced by the same faculty have. Then, the conclusion is the existence of a warranted false belief in Plantinga's account, which leads to the existence of warranted true belief that falls short of knowledge.

To summarize, I claim that Plantinga's reliable design plan requirement does not secure the reliability of a faculty. As a result, it is possible that even a belief produced by an unreliable faculty could satisfy all four conditions for warrant. Then, Plantinga's theory has to allow either a warranted true belief which intuitively cannot be knowledge or a warranted false belief. If it is the former, the sufficiency of Plantinga's four conditions for warrant is questionable. Even if it is the latter, it is no better than the former, because it is one step away from the same result.

3.3. The New "Favorable Environment" Requirement as Plantinga's Solution and Its Difficulty

Let's see how the existence of a warranted false belief causes a difficulty in Plantinga's account; specifically how it leads to a conclusion that Plantinga's four conditions are not sufficient for warrant. Earlier, I have mentioned Merricks' argument that Plantinga's warrant has to imply truth because, if not, using warranted true belief, we can produce a Gettier-like example concerning warrant.36 Klein uses the same formula, and creates the Generalized Gettier example as a counterexample against Plantinga.37 This is the case of the Lucky Ms. Jones and her Ford, where Ms. Johns believes that she

36) See previous ft. 29.
owns a Ford on the basis of the evidence that she parked it in the parking lot this morning, while her Ford is demolished unbeknownst her yet she still owns a Ford because she has just won one in the lottery. Ms. Jones’ belief that she owns a well-functioning Ford is true but it’s not a case of knowledge. However, Klein convincingly shows that none of Plantinga’s four conditions for warrant is violated in this example.

Facing this and similar kinds of charges, Plantinga begins to realize the problem of his account. He finds a way to solve this problem in his second condition, the favorable environment requirement. In fact, his second condition has been seriously criticized by many including Feldman and Swain. As pointed out by these critics, Plantinga’s initial solution of Gettier cases which heavily utilizes this environment condition is indeed problematic. Suppose that Plantinga attempts to meet the charge of Klein by saying that the environment described in the example (where one’s car is destroyed but a new car is replacing it unbeknownst the agent, and so on) is not a sufficiently similar environment to that for which the faculty is designed. (It is unclear how this can be so, but let us assume that the problem is at least in the environment.) The problem is that Ms. Jones believes many other things under the same allegedly problematic environment. For example, Ms. Jones may believe that Ms. Smith owns a Ford too, which could be a perfectly warranted belief even though it is produced in the same allegedly problematic environment. According to critics, there is no non ad-hoc explanation for why, between two beliefs formed in the same environment, only one lacks warrant due to the hostile environment.

---

38) See previous ft. 14.
Now, Plantinga makes a mini/maxi distinction of the environment and attempts to meet this criticism. The general idea is this: Against the criticism, Plantinga ultimately wants to claim that what the critics consider as "the same environments" in the counterexamples are in fact two different environments. That is, his critics' idea of the same environment is based on the similarity of two maxi (or at least larger than mini) environments. If we look into the mini environmental level, the two beliefs in question are produced in two different environments, namely, one is hostile and the other is favorable. So, if we apply this idea to the Ms. Jones case, a possible response would be that the mini environment in which Ms. Jones' belief that she owns a Ford is produced is different from the mini environment in which Ms. Jones' other belief that Ms. Smith also owns a Ford, although both of them are produced in the same maxi environment.

However, it is not entirely clear how exactly this mini/maxi distinction avoids the criticism. Let's call S's Gettierized true belief that there is a barn (acquired in a barn-facade field) B1. And call a perfectly normal true belief that there is a tree (there is no shade of Gettier in this belief because it is a tree-belief, not a barn-belief, although this belief, too, is acquired in the same barn-facade field) B2. Finally, call another presumably normal true belief that there is a barn (acquired in the same barn-facade field but when S has an ability to distinguish a real barn from a facade) B3. The criticism (by Feldman and by Swain) includes the claim that B1, B2 and B3 are all produced in the same, namely hostile, environment although there should be no problem to say that B2 and B3 are warranted. Now, in order to avoid this criticism, Plantinga should be able to say that, although they might be produced all in the same hostile maxi-environment, only the mini
environment in which B1 is generated is hostile while the mini environment in which B2 and B3 are generated is favorable. In order to make sense of this difference at the mini-environmental level, Plantinga takes the idea that a mini environment is "maximally specific."

For example, there is the cognitive mini-environment of the van case, the Gettier-Russell mini-environment in which I happen to look at a clock that has stopped, and the fake barn environment. On the other hand, there are also the cognitive mini-environments in which the clock I glance at is keeping proper time, the ones in which my van remains unmolested in the place I parked it, and the ones where there are only real barns. We can think of a cognitive mini-environment of a given exercise of cognitive powers E ["MBE," or the mini-environment M for E] as a state of affairs (or proposition)—one that includes all the relevant epistemic circumstances obtaining when that belief is formed. ... [The MBE] will include, for example, the presence or absence of fake barns, of my van’s being destroyed in unforeseen ways (if it is), of Paul’s brother Peter being in the neighborhood, and any other relevant epistemic circumstance. To be on the safe side, let MBE be as full as you please, as large a fragment of the actual world as you like.39

If a mini environment is understood as "maximally specific" in the sense that it includes "all the relevant epistemic circumstances," it would be correct to think that the mini-environment in which B1 is generated is different from where B2 or B3 is generated. In fact, given that B1, B2 and B3 are results of three different event tokens, we may have to admit that the "maximally specific" mini environments in which beliefs are produced are bound to be different in every case. However, affirming this fact cannot be the end of the solution. We need a standard for distinguishing which environment is hostile and which is favorable. Concerning this, Plantinga claims that


92
MBE [the mini-environment M for E] is favorable for E iff, if S were to form a belief by way of E, S would form a true belief.

and that

a belief produced by an exercise E of cognitive power has warrant only if MBE is favorable for E.40

This surely is a reinforcement of the previous ‘favorable environment requirement.’ According to the original condition (Condition (2)), in order for an environment to be favorable for exercising a cognitive faculty, all it needs is to be sufficiently similar to the one for which the faculty is designed. It says nothing about the reliability of the ways in which beliefs are produced. Now, the revised condition asks a favorable environment to be a reliable environment in exercising a given cognitive faculty. This amounts to a claim that, in a favorable environment, a faculty works reliably.41 In other words, if a favorable environment for a certain faculty is where S would form a true belief by exercising a given faculty, the condition that an environment must be favorable for exercising the faculty is equivalent to the condition that the faculty must work reliably in that particular environment. This is a new requirement. As I mentioned previously, in Plantinga’s original account for warrant, there was no requirement that a faculty should work in a reliable manner. Now, the revised condition specifically requires this. From this point of view, the revised condition looks more like a replacement, rather than a reinforcement, of the previous one.

40) Ibid., p.328
Indeed, unlike the previous one, this revised condition is no longer a requirement on environment in its literal sense. With a plausible understanding of “all the relevant epistemic circumstances,” this condition can be very versatile. For example, this condition now can be utilized in explaining some ambiguous situations like when the agent is drunken or hallucinating. In the previous account, it was ambiguous whether these are cases of malfunctioning faculties or cases of hostile environments.42 Now, we do not need to worry about such an ambiguity because intoxication and hallucination would be “relevant epistemic circumstances” anyway, so the second condition now handles them.

My point here is that Plantinga’s revision of the environmental condition, seemingly motivated by replying to the Gettier problem, seems to be in fact closer to a major repair of a deeper problem of his account, namely the problem caused by not having a reliability condition regulating directly how a faculty should work. However, if this is all I can say, that wouldn’t be a fair complaint against Plantinga—not if his revision can really solve the problem that it is supposed to solve. However, there seems to be a reason to doubt the success of his revised environmental requirement and the Gettier solution. Let’s consider one Gettierized case mentioned by Plantinga. My justified true belief that my neighbor Paul is standing in his doorway is not a case of knowledge if, unbeknownst to me, his identical twin brother Peter moved into Paul’s place last night.

---

41) “We must therefore say that a cognitive faculty—vision, say—can be unreliable in a given mini-environment M even though it is reliable in the maxi-environment including in M.” Ibid., p. 374 (End Note 12).

42) Steup criticizes the proper functioning faculty requirement with his example of a drunken man’s cognitive faculty in his article (Steup, Op. Cit.). However, I think he does not consider a possibility that Plantinga’s environment condition may handle his criticism.
would believe that it is Paul even if it were Peter. I am Gettierized. In order to explain this, Plantinga would appeal to the hostile environment.

I form a belief as to the identity of the person standing in Paul’s doorway by taking a quick look from across the street; this exercise of cognitive faculties can be counted on to produce a true belief in a mini-environment where Paul is the only person in the neighborhood that looks at all like Paul; not so for one in which Paul’s look-alike twin brother Peter is (unbeknownst to me) staying in Paul’s house.43

In short, the night when Peter moved into Paul's house, the cognitive environment under which my Paul-identifying faculty is working in accordance with whatever design plan it has, has turned hostile for exercising that faculty. However, according to the revised solution, we cannot and may not stop here. We could go on until we know whether the maximally specific mini environment is hostile.

Surely, some relevant epistemic circumstantial changes would result in different mini environments. For example, suppose that even though Paul and Peter both frequently came in and out yesterday, today, Peter decided to stay inside the house all day long. Then, interestingly, the mini environment of today (strictly, this is not yet reached to the maximally specific level that Plantinga allows us to stipulate) doesn't seem to be hostile to my Paul-identifying faculty, at least, not as hostile as yesterday's environment. Today, if I see a Paul-like person outside Paul's house, then that would be Paul because my chance to see Peter instead is zero. Peter did not come outside of the house today. My Paul-identifying faculty would produce a true belief in this mini environment. So, according to Plantinga, my true belief that Paul is standing in his doorway is a warranted
belief today, thus, a case of knowledge. However, I didn't know that it was Paul yesterday even though all I saw yesterday was in fact Paul, just like today. This is not very intuitive but I will leave the possibility open. That is, it might be right that I really didn't have knowledge yesterday but I do today, depending on what kind of mini environment I am in. Probably, the different objective probability of seeing Peter instead of Paul in each mini environment would play an important role in determining which mini environment is hostile and which is favorable. However, once we start to think of a mini environment as an actual maximally specific situation, the real problem seems to be this: the hostility of the environment which Plantinga hopes to find in many Gettier cases would not be found there because, in any maximally specific mini environment in which S has formed a true belief, S would form a true belief if S were to form a belief. In other words, in every maximally specific mini environment in which I form a true belief that Paul is there, the objective probability that I do see Paul, not Peter, is 100%. Then, whenever I have a true belief, it is produced by a favorable environment.

Let's suppose E, the faculty involved here, is this: <seeing a Paul-like person, believing that it is Paul, in order to get truth>. No doubt, if I keep exercising this cognitive power E under an MBE (a mini environment for E) which includes the fact that Paul's identical twin brother is living in the same house, then I would not always get true beliefs. So, this MBE does not satisfy Plantinga's definition of the favorable MBE. However, note that any maximally specific MBE concerning my true Paul-belief also includes the fact that what I see is indeed Paul and also that Peter is not in my visual field.


96
Under this MBE, if I were to form a belief by way of E, I would form a true belief. Then, this mini environment turns out to be favorable for E. For the same reason, as long as the MBE includes the fact that I look at a real barn, this MBE is favorable for E such as <seeing something looks like a barn, believe that there is a barn, for truth>.\textsuperscript{44} Even in the case of the dead watch, the E, namely <looking at a watch, telling time, for truth> would always produce true beliefs if it is used under the MBE including that it is 12 o'clock and that the watch is dead at 12 o'clock.\textsuperscript{45} Therefore, introducing maximally specific mini environment would make all Gettierized true beliefs into beliefs produced in a favorable environment.

In fact, Plantinga expects this kind of difficulty in his revision. So, he adds the following restriction to the way in which we describe the maximally specific mini environment:

MBE must therefore be specified in such a way that it doesn't include E's producing a true belief and also doesn't include E's producing a false belief. The proposition that S forms a true belief will be neither true nor false in MBE.\textsuperscript{46}

However, it is unclear how this restriction can defend Plantinga from the criticism that I have raised. If this restriction is understood literally so that Plantinga's mini environment is a maximally specific environment from which only the proposition "E produces a true (or false) belief" is removed, then the restriction is obviously insufficient to achieve the

\textsuperscript{44} The MBE also includes facts like that there are 300 barn facades near the real barn, that I have no awareness of such barn facades, that I have no discriminative ability between a barn and a well-made barn facade, and so on., but it does not matter. Whenever we use the E under this maximally specific MBE, we will get a true belief that there is a barn.

\textsuperscript{45} So, in this case, using the E under the MBE always generates a belief that it is 12 o'clock, and again, this will be always true because whenever it happens, it is 12 o'clock.
intended goal. As we can see, in order to make my point so far, I did not have to include the proposition that E produces a true belief or its denial in any of the MBE that I used.

Let's try differently that the mini environment is a maximally specific environment in which 'S forms a true belief' is neither true nor false. (That is, 'S forms a true belief' is not a part of the specification of the mini environment. 'S forms a false belief' is not a part of it, either.) Consider an environment in which Paul is standing in his doorway and Peter is around but not in my visual field at this moment. Is the proposition that 'I form a true belief (that Paul is standing in his doorway)' true or false in this environment? I think that this proposition is "neither true nor false in this environment" as Plantinga asks for. (That is, the truth of this proposition is not implied in this environment.) For one thing, this environment does not include anything about my forming a belief at all. (That is, we can stipulate the environment in this way.) So, it is qualified to be a kind of mini environment that Plantinga would endorse as having an appropriate degree of being "maximally specific." Now, in this environment, if I were to form a belief using E (say, Paul-identifying faculty or just a faculty of vision), would I form a true belief (that Paul is standing in his doorway)? There is no reason to think that I wouldn't as long as E is functioning properly. Thus, it is a favorable mini environment for exercising E. Given that E is functioning properly in accordance with a truth-aiming reliable design plan, this belief satisfies all the conditions to be a warranted belief. However, this belief is a Gettierized true belief which cannot have warrant.

However, one might suspect, as a way of defending Plantinga, that the mini environment that I have used above includes propositions from which the proposition that S forms a true belief is directly inferred. That is, contrary to my claim, it might be that the proposition that S forms a true belief is in fact true (or false) in this environment. Then, it might be said that Plantinga’s restriction can easily handle my challenge by just adding the following: MBE also should not include the proposition from which the proposition that S forms a true belief or its denial is directly inferred.

However, I do not think this would be a viable option. First, it would be very difficult to specify the nature of the “inferring” relationship here. If it is a rigorous logical notion of deduction, it wouldn’t be true that “S forms a true belief” is implied by some of the propositions which are already true in that mini environment. We have seen this above. Then, this “inference” relationship should be understood more like an induction with a high probability. However, this view is also hard to maintain. According to this view, one way to stipulate a mini environment would be to remove relevant true propositions from the maximally specific environment until it reaches to the stage in which the truth of “S forms a true belief” is neither highly probable nor highly improbable. First, this environment is far from being a maximally specific one, not to mention its chance of being a very arbitrary one. Second, we may not even be able to determine in what mini-environment the belief is produced. For example, suppose a maximally specific environment is constituted by a conjunction of propositions A, B and C. Suppose that, since A&B&C makes “S forms a true belief” highly probable, we need to remove a proposition in order to reach the stage in which the truth of “S forms a true belief” is
belief" is neither highly probable nor highly improbable. Now, it is quite possible that a removal of any proposition among A through C would result in the stage that we are looking for. Then, which environment (A&B, A&C and B&C) is the environment in which the target belief is produced?

To conclude, even if we are sympathetic to what Plantinga wants to do when he puts this restriction on his revised favorable environment requirement, it is hard to see how it could be done. It is my guess that the difficulty that Plantinga is struggling to avoid here, i.e., how to specify the relevant environment in which the given belief is produced, is similar to the generality problem which Feldman has raised against reliabilism and which has been considered a major difficulty of the theory. If Plantinga’s favorable environment requirement ultimately amounts to a requirement for a reliable working environment for a faculty, then stipulating a maximally specific mini environment without falling into the “Single Case Problem” is indeed the very question Feldman asks of reliabilism. So far, I do not think Plantinga has an answer better than the ones already tried, which have turned out to be more or less unsatisfactory.

4. Conclusive Summary

Let me summarize my argument developed in this chapter. I first focused on the argument that a properly functioning faculty is not necessary for warrant and knowledge. For this, I began by pointing out a popular criticism, namely the Swamp Man type example, because I think this type of counterexample can be answered by Plantinga.

---

Then, I proposed examples which I think serve better, namely, the case of Kim and the case of Victoria which show that a belief produced by a malfunctioning faculty can be knowledge: thus proper function in Plantinga's sense (following its design plan) is not necessary for knowledge. I didn't intend these examples to be a "no-way-out" criticism by themselves at that point. Instead, I considered a possibility that Plantinga might be able to dodge the difficulty by giving a different understanding of "working in accordance with design plan" than I used in those examples. In fact, I acknowledged that this option may give us a better perspective on how his idea of a properly functioning faculty (by following its design plan) should be understood. Thus, the purpose of my counterexamples of Kim and Victoria was either to show the non-necessity of proper function for knowledge or to force Plantinga to hold the following view in order to avoid the problem: A faculty under a circumstance C and responding R-way at t is functioning properly (by following its design plan) at t iff responding R-way under C is prescribed in its design plan. I then argued that there is another kind of difficulty waiting for him if he takes this option. With this understanding of following the design plan, we can generate a case where Plantinga's four conditions for warrant are met but knowledge is not acquired as in the case of Mr. Magoo. Then, I provided my diagnosis of this problem using the case of occasional success, and noted that Plantinga's original conditions for warrant do not have a requirement for a faculty to be reliable over a period of time. As a consequence, it would be either that his original four conditions are not sufficient for warrant, or that a warranted belief could be false. The former is a problem by itself, and the latter would lead to a problem pointed out by Merricks and by Klein, namely the

48) Concerning the Generality Problem, see chapter 1, p. 28.

101
problem of the Generalized Gettier concerning warrant. Finally, I considered Plantinga’s reply to avoid this difficulty by adding a mini/maxi distinction in his environment condition. I argued that this would not solve the problem due to the difficulty of specifying the relevant environment.
CHAPTER 3

A CRITICISM OF SOSA'S VIRTUE PERSPECTIVISM

In this chapter, I will critically review virtue perspectivism, Ernest Sosa's most recent epistemological position. His theory of knowledge and justification combines two major ideas. These are: intellectual virtue theory and perspectivism. The idea of intellectual virtue constitutes a version of externalist theory called 'virtue reliabilism.' The perspectivism is a version of internalism, which specifically requires the epistemic agent to be aware of the intellectual virtue (reliability of the belief production and/or the permissibility of the target belief). And according to the way in which Sosa develops his account, his perspectivism also has something to do with coherentism.1 Undoubtedly, Sosa's project is to construct a more defensible and desirable epistemological position by combining the merits of major theories. Sosa wants to maintain the reliabilist framework for his theory but also wants to pay attention to the other epistemic elements (such as

1) Even though it is not entirely clear how, Sosa seems to think that his perspectivism is a possible compromise between coherentism and foundationalism. If so, his account apparently accommodates the
internal perspective, which in turn presupposes accessibility, and coherence) that reliabilists or externalists have treated as non-essential for their epistemology. As the result of combining them all, Sosa’s virtue perspectivism is, as Alston dubs it, a kind of “we need it all” view.2

The hard part of this type of project, however, is how to combine those heterogeneous elements into one theory in such a way that each introduced element only contributes positively to the theory and does not create a tension that threatens the viability of the theory. More often than not, the ambition to serve both turns out to be serving neither especially when one constraint is characterized by the denial of the other constraint. With this regard, Sosa’s project may look suspicious from the beginning, and he owes an explanation for how his brand of combination does not fall into one of those unfortunate cases. Without such an explanation, his theory might look fundamentally misconceived. In this chapter, I criticize Sosa on these grounds. I argue that Sosa’s virtue perspectivism leads to inconsistency. If this is right, then Sosa’s claim of the superiority of his account over other versions of reliabilism is not justified.

Sosa categorizes his account of virtue epistemology as a type of generic reliabilism. However, his claim is that it is a better version of reliabilism.3 It is certainly true that one of his basic notions, i.e., the notion of epistemic virtue, strictly remains ideas from foundationalism, too. (See p. 97 in Sosa, E., Knowledge in Perspective, Cambridge University Press (1991))


within the boundary of reliabilism. However, an important portion of what makes his account a better version of reliabilism, or at least a distinguishable version from other forms of generic reliabilism (specifically in the sense that this important portion provides some ways to protect his "basically reliabilistic" account from the known criticisms of reliabilism) seems to be coming from non-reliabilistic sources such as requiring the agent's access to the reliability. If so, the claim that his account is a version of reliabilism might be debatable. In any case, regardless of the true identity of Sosa's account, if such an ideal combination of reliabilism and internalism could be done successfully, some difficulties of reliabilism are expected to be resolved. In fact, among many things that Sosa's epistemology aims to achieve, it is consciously designed to deal with the problems that the current state of reliabilism faces. One is the problem of justification in the New Evil Demon world. Here, the problem of reliabilism is that a belief of the demon victim, which cannot satisfy the reliability requirement, seems to be justified intuitively. Thus, according to this objection, reliability is not necessary for epistemic justification. The other is Bonjour's charge against reliabilism. He questions the epistemic status of true beliefs produced by reliable clairvoyance. Sosa calls this a meta-incoherence problem, where, contrary to the reliabilist's account of justification, the belief seems to be unjustified although it is reliable. Here, the sufficiency of reliability for justification and knowledge is challenged. Sosa develops his theory in order to meet these difficulties. For example, he distinguishes aptness and justification with the intention that this

---

1) Sosa, E., "Reliabilism and Intellectual Virtue" in his Knowledge in Perspective pp. 131 - 145. Sosa also presents his own solution of the generality problem. However, it seems to me that, basically, Sosa's solution of the generality problem is more related with his acknowledgment of yet another epistemological element.
distinction helps his account to deal with the New Evil Demon problem. He distinguishes animal knowledge and reflective knowledge and uses it to deal with the meta-incoherence problem. These two distinctions are unique and essential features of Sosa's virtue perspectivism. Therefore, my discussion mainly concerns these problems and distinctions. I argue that Sosa's use of these distinctions to handle the problems in his virtue perspectivism reveals some internal difficulties of his epistemology. Hence, despite his claims, Sosa's theory cannot be superior to competing epistemological theories.

1. The Basic Structure of Sosa's Theory

The complete picture of virtue perspectivism that emerges from several key articles in Sosa's anthology *Knowledge In Perspective (KIP, from now on)* is delicate and convoluted. Although providing a quick glance to the theory is not impossible, it can be challenging because his theory—or at least some parts of the theory—easily invites misunderstanding unless all the relevant details are filled in. However, a complete explanation of Sosa's whole theory with all the relevant details is not the initial purpose here. My main concern is whether Sosa's virtue perspectivism is an improved version of reliabilism, as he claims. For this reason, only some characteristics of his account, especially the features that I take it to be essential for overcoming the weaknesses of

---

namely a social aspect of knowledge (and justification), rather than utilizing the idea of virtue perspectivism itself. For this reason, I will not discuss the generality problem and Sosa's solution of it in this chapter.


6) I don't think that readers are solely responsible for such possible misinterpretations of Sosa's account. His thoughts on the essential components of his theory, such as intellectual virtue, justification, coherence, and internal perspective, are not really presented in an all-put-together-in-one account. Readers are frequently asked to go back and forth to his other relevant "chapters" in the book (in fact, they are
reliabilism, will be critically discussed, and other points, even though they are important for the completion of the theory, will be disregarded.

The basic ingredients that make up Sosa’s theory are the following: intellectual virtue or faculty, aptness, justification, animal knowledge, reflective knowledge, coherence and perspective. At bottom, knowledge, for Sosa, is a true belief resulting from intellectual virtue. What is intellectual virtue? According to Sosa, intellectual virtue is defined as “a competence to distinguish the true from the false in some field of propositions F when in certain circumstances C.”

Following Sosa’s own example, let’s consider ordinary human visual perception as a case of intellectual virtue. If S believes a proposition X (“This before me is white and round”), we can think of a field F to which proposition X could belong. Propositions about shapes and colors of the objects would be a possible example of F here. We can also think about a certain circumstances C that S might be in. Let’s say that S’s standing in front of middle-sized objects in daylight, at arm’s length is the one. According to Sosa, C could be either external as the one we are considering now, or internal (e.g., S’s being conscious and entertaining X). In addition to C, there also is E, an environment, which is external and generally broader than C (e.g., the surface of the earth). Now, if S has a certain inner nature (grounded in, plausibly but not necessarily, S’s having good eyes, healthy nerve system and brain) relative to E so that S would most likely to be right on any proposition X in field F when S is in C, then

---

107

---


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
S’s belief X is considered as being out of intellectual virtue. Here are Sosa’s own words for this;

A subject S's intellectual virtue V relative to an "environment" E may be defined as S's disposition to believe correctly propositions in a field F relative to which S stands in conditions C, in "environment" E.8

In order for a belief to be knowledge, it shouldn’t be true by coincidence. For Sosa, it means that a true belief has to turn out right by the involvement of such an intellectual virtue.9 With this regard, Sosa’s intellectual virtue is not different from the basic idea of reliabilism or the earlier version of externalism on knowledge (such as Armstrong’s). They all focus on objective truth-conducivity as the primary factor for knowledge. The epistemic agent’s subjective and psychological conditions (including having a justification in a traditional sense) are not the factor that determines whether the agent has knowledge. Rather, beliefs have to have the property of “being likely to be true” in order to be knowledge. Now, for Sosa, beliefs out of intellectual virtue are likely to be true.

For some reliabilists such as Goldman, Swain and Alston, a belief that is likely to be true is a justified belief. However, Sosa’s term for referring to a belief that has high objective probability to be true due to the intellectual virtue related to its production is called “aptness.”

The “aptness” of a belief B relative to an environment E requires that B derive from what relative to E is an intellectual virtue, i.e., a way of arriving at belief that

---

8) KIP p. 140 ("Reliabilism and Intellectual Virtue")
9) KIP, p. 277. ("Intellectual Virtue in Perspective")
yields an appropriate preponderance of truth over error (in the field of propositions in question, in the sort of context defined by C).10

Thus, all beliefs out of intellectual virtue are apt. Then, all beliefs that are qualified to be knowledge are also apt. Now, let's suppose that we have an apt true belief. Do we have knowledge? Here, Sosa’s answer is yes but with a condition that there are in fact two types of knowledge. Apparently, a merely apt true belief makes up only one type of knowledge, called animal knowledge. In order to be the other kind of knowledge called reflective knowledge, there should be more. First, let's see his distinction of animal knowledge and reflective knowledge.

One has *animal knowledge* about one's environment, one's past, and one's own experience if one's judgments and beliefs about these are direct responses to their impact—e.g., through perception or memory—with little or no benefit of reflection or understanding.

One has *reflective knowledge* if one's judgment or belief manifests not only such direct response to the fact known but also understanding of its place in a wider whole that includes one's belief and knowledge of it and how these come about.11

And concerning what more is required;

Animal knowledge requires only that the belief reflect the impact of its subject matter through the operation of a faculty or virtue. For reflective knowledge one not only must believe out of virtue. One must also be aware of doing so. Of course, one need not know with precision and detail the relevant C and F. Some grasp of them is required, however, even if it remains sketchy and generic.12

---

10) *KIP*, p. 144 ("Reliabilism and Intellectual Virtue") and p. 289. ("Intellectual Virtue in Perspective")
11) *KIP*, p. 240. ("Knowledge and Intellectual Virtue")
The picture that Sosa is drawing here so far is relatively clear. On one side, there is a mere apt true belief, which is acquired by an intellectual virtue alone, and which amounts to animal knowledge. By the definition of an intellectual virtue, an apt belief is a belief produced via a certain truth conducive inner nature of S in E when the proposition is one in F and when S is in C. On the other side, there is an apt true belief accompanied by a perspective ("understanding," "awareness," or "grasp" of how the given apt belief comes about "in a wider whole"), which is reflective knowledge. This kind of knowledge is acquired by a combination of an intellectual virtue and a perspective on it. Aptness coming from intellectual virtue is necessary for both animal and reflective knowledge, but additionally, "understanding," "awareness," or "grasp" of such a virtue as well as the circumstance (C) under which the virtue is working is necessary for reflective knowledge.

Some understand Sosa's account even more schematically so that they identify this additional perspective requirement as Sosa's notion of justification. Then, according to this understanding, reflective knowledge would require both aptness and justification while animal knowledge only needs aptness. In fact, the schema of "aptness - reliability (intellectual virtue) - external - animal knowledge" vs. "justification - perspective (awareness of the first-order intellectual virtue) - internal - reflective knowledge" is seemingly accepted by many Sosa's commentators and critics. They think that this schema is "rough" but generally unproblematic. However, I think this is misleading.

13) For example, see Bonjour, Laurence, "Sosa on Knowledge, Justification, and Aptness," Philosophical Studies 78 (1995) p. 208 and Foley, Richard, "The Epistemology of Sosa," Philosophical Issues 5: Truth and Rationality (1994) p. 2. Also compare them with Axtell, Guy, "Recent Work on Virtue Epistemology," American Philosophical Quarterly 34 (1997) p. 6. Here, Axtell distinguishes "justification" and "internal justification" and says that reliabilism is central in Sosa's "justification." This is my point. However, Axtell adds that Sosa's "internal justification is largely a matter of comprehensive coherence," which is closer to...
My understanding is that the perspective requirement alone is not equivalent to the justification requirement in Sosa. In other words, a belief is not justified in Sosa by merely having a coherent internal perspective, i.e., by having the awareness of the relevant intellectual virtue and the circumstance under which the target belief is produced. My reason is based on Sosa's text. We can notice that, when Sosa talks about the requirement for reflective knowledge in the above two quotes, he does not mention justification. Consider the following quote, too.

For reflective knowledge, you need ... an epistemic perspective that licenses your belief by its source in some virtue or faculty of your own. You trust your own correctness, holding your belief to be right through its origins in a reliable faculty or virtue.

Here again, a condition for reflective knowledge is discussed using a different terminology (a belief's being licensed by a perspective, and the agent's trusting one's own correctness) but there is no indication that this means a justification for the belief. A more explicit and crucial piece of textual evidence can be found in Sosa's claim that justification is one way for a belief to be apt. It is true that there is a place where he mentions justification specifically concerning reflective knowledge. For example:

---

Bonjou's and Foley's view where Sosa's justification is treated just like Axtell's notion of "internal justification." I do not find in Sosa whether there is such a distinction between "justification" and "internal justification" unless this is a distinction of "justification" and "internal perspective." I would agree with Axtell's understanding if what he is indicating with his "internal justification" is Sosa's notion of perspective requirement. Otherwise, saying that Sosa's "internal justification is largely a matter of comprehensive coherence" is misleading if not incorrect, and can be counted with another example as identifying Sosa's conditions for justification purely internal conditions.

14) Above footnote 10 and 11.
15) KIP, p. 277 ("Intellectual Virtue in Perspective")
16) See the quote below (footnote 19) from KIP, p. 245. ("Knowledge and Intellectual Virtue") Also see the following. "Apt then is perhaps what a belief must be to qualify as knowledge, in addition to being true (and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
By contrast, reflective knowledge always requires belief that not only is apt but also has a kind of justification, since it must be belief that fits coherently within the epistemic perspective of the believer.\(^{17}\)

However, this can be explained in a way that is compatible with my understanding. In my opinion, the correct thing to say about the relationship between perspective and justification is that an apt belief requires an internal perspective in order to be a justified belief (and thus, to be reflective knowledge). If so, it would be true that we reach reflective knowledge only through justified belief. However, at the same time, justification is not the same as having internal perspective. Based on the claim that justification is one way for a belief to be apt, the crucial difference is that, while internal perspective is something that can be independently achieved without aptness, justification cannot. For Sosa, a justified belief should also be an apt belief.

Now, let's consider how Sosa himself explains his notion of justification.

Virtue perspectivism distinguishes between aptness and justification of belief, where a belief is apt if it derives from a faculty or virtue, but is justified only if it fits coherently within the epistemic perspective of the believer—perhaps by being connected to adequate reasons in the mind of the believer in such a way that the believer follows adequate or even impeccable intellectual procedure.\(^{18}\)

The "justification" of a belief B requires that B have a basis in its inference or coherence relation to other beliefs in the believer's mind— as in the "justification" of a belief derived from deeper principles, and thus "justified," or the "justification" of a belief adopted through cognizance of its according with the subject's principles, including principles as to what beliefs are permissible in the circumstances as viewed by that subject.\(^{19}\)

---

\(^{17}\) KIP, p. 145 ("Reliabilism and Intellectual Virtue")

\(^{18}\) KIP, p. 145 ("Reliabilism and Intellectual Virtue")

\(^{19}\) KIP, p. 144 ("Reliabilism and Intellectual Virtue") and p. 289. ("Intellectual Virtue in Perspective")
Here, whether a belief "fits coherently within the epistemic perspective of the believer" is introduced as a necessary condition for justification. And, "being connected to adequate reasons in the mind of the believer" is suggested as one way that a belief fits coherently within the perspective. When we consider a perspective as the agent's awareness of how a belief is produced under what circumstance (the awareness of virtuousness), then we will have this: in Sosa, a belief that p (Bp) is justified when the agent is aware of how he acquired Bp, and Bp fits coherently, possibly together with some reasons he has for Bp, with this awareness.\footnote{I am not clear what kind of coherence notion Sosa is using here but I assume that a notion of "explanatory coherence" would be a safe understanding.} However, again, these are necessary but not sufficient for justification. The other necessary condition is aptness.

Good rhetoric suggests therefore that "justification" and its cognates be yielded to the argumentative account; in which case justification must likely fall from its status as principal concept of epistemology. ... Better to demote justification to the status of one way for a belief to be appropriate or apt for knowledge, while allowing other ways not dependent on already attained justification: perception, perhaps, or introspection, or memory.\footnote{I am not clear what kind of coherence notion Sosa is using here but I assume that a notion of "explanatory coherence" would be a safe understanding.}

Here, justification is considered as being on par with perception, introspection and memory because they are all appropriate or apt ways for reaching knowledge. In other words, just like a belief can be apt by being produced via intellectual virtue (reliable perception, for example), a belief can be apt by being justified.

According to the claim that justification is one way for a belief to be apt (thus, a belief cannot be justified without being apt first), aptness is a larger concept in its
extension. It includes justification. Then, despite the expectation based on the schematic understanding, this re-introduces the truth-conducivity issue into the Sosa's notion of justification. (In the schematic understanding, Sosa's justification does not have to be truth conducive.) Namely, aptness depends on intellectual virtue, and intellectual virtue is ultimately truth-conducivity, so Sosa's justification requires the reliability in the belief formation. Thus, a justified belief for Sosa is a belief that is an outcome of the involvement of intellectual virtue, and that also fits coherently within one's perspective. However, if this is the case, how can this notion of justification handle the known problems of a reliability theory of justification? As we can see in the New Evil Demon scenario (the demon victim's beliefs are not apt according to Sosa's account but they may well coherently fit within the victim's internal perspective), although not all beliefs that fit coherently within one's epistemic perspective are apt, we still want to call them "justified." This does not support Sosa's position because Sosa would have to deny the justificatory status of such a belief, i.e., a belief that lacks aptness even if it has coherence and internal perspective. How does Sosa react to this problem?

If the schematic understanding of Sosa's notion of justification (i.e., perspective and internal coherence is sufficient for Sosa's justification) were correct, we should expect that Sosa would grant the justificatory status of the demon victim's beliefs. However, the fact is, he denies the justificatory status of the demon victim's beliefs.22 Even though he allows that many internalism-favored components (coherence,

21) KIP. p. 245. ("Knowledge and Intellectual Virtue")
perspective, believer’s awareness) play important roles in his account, he cannot simply affirm the common sense intuition that the demon victim’s beliefs are justified (due to its coherence and internal perspective). Rather, he has to introduce yet another idea that justification is relative to the environment in order to deal with the New Evil Demon case. Sosa’s solution of the demon victim’s justification will be discussed later in this chapter. However, given what I have mentioned here, one can already expect that it is not as intuitive as the schematic understanding of his justification might suggest. Sosa says that although perspective is required for justification, and furthermore that the demon victim’s beliefs are OK in terms of such a perspective requirement, they are not justified relative to the demon world due to the lack of truth-conducivity. At a glance, this is no better result than the one from generic reliabilism, which is exactly the claim that the critics of reliabilism think is counter intuitive. One might wonder what Sosa’s motivation for introducing perspective and coherence in reliabilism is if the outcome fails to conform to our intuitions in the demon victim case. However, the real motivation and the straightforward effect of perspective and coherence business come in handy when Sosa deals with the other side of the problem in reliabilism, namely the meta-inconsistency problem. Sosa also thinks that a perspective is necessary in order to prevent an “accidentally apt” belief from being knowledge. Details on this will be discussed later in this chapter. In any case, despite its appearance, his internalistic feature such as coherence and internal perspective requirement is not the straightforward solution of the New Evil 

22) Precisely, what Sosa wants to do with the demon victims beliefs is to deny their justificatory status in the demon world (because they are not apt in the demon world) and then to restore their justificatory status relative to the normal world (because they are apt in the normal world).
Demon problem. And this I think is the final crucial evidence that coherence and perspective alone cannot be everything in Sosa's notion of justification. The aptness or truth-conducivity should also be a part of it.

Let me now summarize the first approximation of Sosa's position using the elements that I mentioned at the beginning of this section.

[1] A true belief (in an un-Gettierized situation) is knowledge iff it results from intellectual virtue.

[2] A belief is apt iff it results from intellectual virtue.

[3] A belief is justified iff it is apt and it fits coherently within one's perspective. (Or differently, a belief is justified iff it results from intellectual virtue and it fits coherently within the agent's awareness of the relevant virtue and the relevant circumstance.)

[4] If a true belief is justified, then it is reflective knowledge.

[5] If a true belief is apt but not justified, then it is animal knowledge.

Finally, there is one more thing to consider. What makes a belief have the property of aptness is intellectual virtue or faculty, i.e., the agent's "disposition" or "competence" to distinguish true from false. Then, what does make a belief have justification? Judging from Sosa's suggestion of the term "a faculty of faculties" as well as "coherence-seeking reason, Sosa seems to think that justification too results from a certain virtue or faculty.

23) KIP, p. 284 ("Intellectual Virtue in Perspective")
24) KIP, p. 291 ("Intellectual Virtue in Perspective")
What we care about in justification are the epistemic endowments and conduct of the subject, his intellectual virtues.25

A belief that $p$ constitutes reflective knowledge that $p$ only if one has a perspective on the source of that belief in a faculty or intellectual virtue of one's own.26

This overarching virtue is thus a faculty of faculties, a faculty that uses our brute animal endowments to raise us above that level and make of us the animal that is rational.27

However, even if it is Sosa's intention that there is a meta-faculty involved in the production of a justified belief, how exactly this meta-faculty works is largely unexplained. One suggestion is that, for Sosa, justification is conferred on an apt belief (due to its reliable first-order faculty) when it also passes a so-called screening test of a meta-faculty (or a "faculty of faculties"). A further speculation on this issue will come later in this chapter. At this point, I only indicate one thing that is certain for Sosa concerning this issue. Sosa insists that such a meta-faculty is involved in almost all of a rational being's belief formation. We normally think that a typical perceptual belief is a reaction (an animal reaction perhaps) to the stimuli without the hint of reasoning, while believing the conclusion of deductive or inductive inference is a result of deliberate reasoning. Such a dichotomy, however, is not found in Sosa. For Sosa, most of a rational being's perceptual beliefs are not merely apt, and do not merely constitute animal knowledge. On the contrary, most of our normal perceptual apt beliefs are justified beliefs as well. Therefore, they qualify to be reflective knowledge rather than animal knowledge.

25) KIP, p. 240 ("Knowledge and Intellectual Virtue")
26) KIP, p. 290 ("Intellectual Virtue in Perspective")
27) KIP, p. 284 ("Intellectual Virtue in Perspective")
This idea has to be brought out with a significant importance because Sosa eventually
denounces the animal knowledge as merely a metaphoric sense of knowledge. There,
Sosa is not saying that our perceptual knowledge, which has been considered by some
externalists as a typical case of “knowledge without (internal) justification,” is animal
knowledge, and thus only metaphorically knowledge. Rather, his idea is that our
perceptual knowledge is the result of genuine cases of “justified beliefs” due to the meta-
faculty involved in their production. Of course, the difficulty is how to explain the notion
of the meta-faculty and its involvement in most, if not all, belief formation.
Unfortunately, a detailed explanation on this issue does not seem to be ready in Sosa’s
texts.

2. The Distinction of Animal and Reflective Knowledge

Concerning Sosa’s distinction of animal and reflective knowledge, the following
seems to be a fair question to begin with: if there are two kinds of knowledge, which type
of knowledge is the knowledge that other epistemologists are also working on?
Especially, when reflective knowledge is the result of adding more conditions to animal
knowledge, what are the conditions for the knowledge that we think as a paradigm case of
knowledge? I think this question expresses our initial suspicion concerning Sosa’s
distinction. That is, isn’t this an easy way out from a major controversy among
contemporary epistemologists, namely the internalism-externalism debate, without really
solving it? Or, isn’t it just saying that externalism is correct concerning animal knowledge
but internalism is also correct when reflective knowledge is the issue?
Among several ways of distinguishing externalism from internalism, let me take the accessibility criterion because it seems to be basic. Given that a belief needs to have some extra properties (in addition to being true) in order to be knowledge (let’s call this set of extra necessary properties X), internalists think that at least some members of X should be accessible upon reflection by the epistemic agent. Externalism denies this and claims that knowledge is possible even though the agent has no cognitive access to any member of X. The tradition of internalism has been supported by the idea that justification is necessary for knowledge, and unless what provides justification (namely justifier) is accessible by the agent, a belief cannot be justified. Against this tradition, externalists either deny that justification is necessary for knowledge (thus, knowledge is possible without justification) or deny that justifiers need to be accessible (thus, a belief can be justified even if the agent has no cognitive access to its justifier).

Indeed, if we assume that knowledge does imply justification, internalists would attempt to point out that there is something counter-intuitive in the externalist position on knowledge. According to the idea of pure externalism, all that the epistemic agent needs to do for knowledge would be having a belief. Even when the agent believes something with absolutely no accessible reason, such a belief could achieve the status of knowledge depending on whether it meets the other external conditions that are true of the agent, true of the belief as a mental phenomenon and true of the environment. Yet, the agent doesn’t

---

28) Depending on what should be accessible, internalism/externalism distinction has different shape.
29) Or, the agent cannot be justified for having a belief.
30) Of course, in order to be knowledge, it has to satisfy some externalist conditions other than justification such as Armstrong’s nomological condition between the belief and what makes it true, Nozick’s subjunctive condition, Plantinga’s warrant condition.
have to be aware of any of those conditions. This would be the nature of uneasiness for going with externalists. However, externalists have something to say about this type of worry. They draw our attention to intellectually less sophisticated cases, such as the case of believing something directly out of our perceptual stimuli, out of our memory or out of our introspection. In those cases, there seems to be nothing that internalists typically think of as “justification on the basis of accessible reasons.” In fact, what can we say about the justification and the accessible reasons if I just believe that my feet hurt when I actually feel my feet hurt? Also, how about my perceptual belief that it is snowing when I see it is snowing? It would be true that there is an accessible reason state, such as sensational state of feeling pain, but the type and role of the “accessible reason” in these cases are very much different and limited when compared to what internalists would think an ideal case of believing something with an accessible reason. Especially, such a minimal sense of accessibility as in “an accessibility to the sensational state” is acceptable as the internalists’ case of believing something with accessible reason, the difference between this and the externalists’ case of believing with no reason becomes so slim that we are no

31) As a result, we have an externalist account of justification such as Goldman’s reliable process theory of justification.
32) Internalists will consider perceptual experience and the sensational state of pain as the accessible justifier, thus as the reason that satisfies internalism. Although this is a coherent view, it suggests a fairly minimal sense of accessibility. What we normally consider as a case of believing something without a reason does not deny the accessibility to the justifier (perceptual experiences) in this minimal sense. For example, even in the case like Norman’s clairvoyance belief, where Norman believes that the president is in New York due to his clairvoyance [see the section 3 in this chapter for the details of the example], which is normally agreed as a case of believing without a reason, we may assume that he has an accessible introspective state (maybe, a kind of inner image of the president in New York). However, it would be at least debatable whether this fact would be enough to change our intuition concerning the justificatory status of the belief. If the agent is not ready to “use” those experiences as the justifier of the belief (typically as in the case of small children), or cannot “grasp” the justificatory relationship between the belief and the experience (maybe, Norman would be such an agent), then the mere accessibility to the experience would not contribute much to the situation that internalists would consider an ideal case of knowledge and justification.

120

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
longer certain what is counter-intuitive in the latter idea. Nevertheless, those introspective or perceptual true beliefs constitute no less legitimate cases of knowledge. Indeed, a large portion of empirical human knowledge would be this kind. If this is plausible, asking all beliefs to be supported (justified) on the basis of the ideal sense of accessible reasons looks like rather a demanding requirement for knowledge, and might even be seen as “a tyranny of intellectualism.” It may very well be true that a belief justified as such should deserve higher epistemic status than beliefs directly formed from, say, perceptual stimuli. However, such a requirement wouldn’t be necessary for knowledge because the critical line between knowledge and non-knowledge would be drawn somewhere lower than where the reasons (thus, justification) are required. This is one main idea behind externalism.

Now, let’s put Sosa’s animal-reflective distinction of knowledge in this context. If animal knowledge is still a legitimate case of knowledge, it looks as if Sosa is repeating the above mentioned position of externalism. The internal perspective (endorsement by reasoning, understanding or awareness of the virtuousness of the target belief) will upgrade a lower degree knowledge (such as perceptual knowledge and knowledge retrieved from memory) to a more desirable higher degree knowledge. Nevertheless, animal knowledge is the bottom line of knowledge. Thus, a true perceptual belief is animal knowledge given that it is an apt belief (that is, the outcome of intellectual virtue), even though the agent has no idea of what is going on about this belief. If the agent gains an intellectual ability to put this perceptual belief in his own internal perspective so that the belief comes out as one that coherently fits in the perspective, then this belief would become reflective knowledge.
However, it is doubtful whether this is really the way to combine the merits of internalism and externalism. More importantly, I don’t believe that this is Sosa’s way. If this were the case, Sosa’s whole project of combining internal perspectivism with reliabilism wouldn’t be so interesting. His theory would merely amount to a position which states that, first, knowledge is acquired by satisfying purely externalistic conditions (animal knowledge by intellectual virtue) and second, only some portion of already acquired knowledge would need additional internalistic requirements in order to be a “better” kind of knowledge. If so, to put it bluntly, even if there is a certain condition that makes a case of knowledge a case of reflective knowledge, why do we have to be interested in those conditions? They are just conditions for what makes knowledge reflective. We are already given an account of what makes knowledge, namely the aptness and intellectual virtue.

The reason that I reject the above way of understanding Sosa’s animal and reflective knowledge distinction as incorrect is based on the following consideration.

How then can one rule out its turning out that just any true belief of one’s own is automatically justified? To my mind, the key is the requirement that the field F and the circumstance C must be accessible within one's epistemic perspective. (Note that this requires considering servomechanic and animal so-called "knowledge" a lesser grade of knowledge, or perhaps viewing the attribution of "knowledge" to such beings as metaphorical, unless we are willing to admit them as being endowed with their own epistemic perspectives.)

Here, Sosa describes animal knowledge as not really a “legitimate” sense of knowledge.

Then, what happens to the perceptual knowledge that we thought we have without the

---

33) KIP, p. 274-275. ("Intellectual Virtue in Perspective")

122

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
awareness of how it comes about? Is it too only a metaphorical sense of knowledge? Well, it would be an unwelcome consequence for Sosa if he would have to admit that our typical perceptual knowledge is knowledge only in a metaphorical sense. As a way to avoid this, Sosa claims that typical perceptual beliefs are in fact cases of reflective knowledge, not animal knowledge.

No human blessed with reason has merely animal knowledge. For even when perceptual belief derives as directly as it ever does from sensory stimuli, it is still relevant that one lacks contrary testimony. People automatically monitor their background information and sensory input for contrary evidence and normally opt for coherent hypotheses even when responding most directly to sensory stimuli. For even when response to stimuli is most direct, credible contrary testimony would change one’s response. The beliefs of a rational animal hence would seem never to issue from unaided introspection, memory, or perception. For reason is always at least a silent partner on the watch for other relevant data, a partner whose very silence is a contributing cause of the belief outcome.34 (Sosa’s emphases)

Thus, according to Sosa, a rational being’s perceptual beliefs are not only beliefs produced by the faculty of perception but also beliefs at least screened by the faculty of reason in their production.

Here, Sosa’s claim is that a rational agent’s reasoning ability participates in his acquisition of a perceptual belief. However, not much has been said about the exact nature of such participation by reason. If we understand the “silence” of reason as the “absence” of reason, his claim that reason is always a contributing cause of a perceptual belief outcome would become a strange claim such that reason helps to cause a certain belief without being involved in its causal process. Even if Sosa is thinking here of a
“conglomerate” faculty (neither a first-order faculty such as perception nor a meta-faculty such as reasoning, but a third one including both) as being responsible for the rational agent’s perceptual beliefs, it is not at all clear how “silencing” the reason in it when all things are fine (when there is no credible contrary testimony, for example) could be considered as one mode of “participation” by reason. A different way to understand this point is that, for any perceptual belief acquired, reason has already scanned the outcome from the perception, and has allowed it to stay because there was no credible contrary testimony. However, in addition to the difficulty of how to distinguish a silent involvement of reasoning from an absence of it (due to some kind of cognitive malfunctioning, for example), it is at best controversial whether the reason in our perceptual belief acquisition process indeed works like this. At least, it would be hard to settle this issue without heavily relying on the help from cognitive science and other empirical science. However, in any case, apparently this is how Sosa tries to get away from the implausible consequence of making all our perceptual knowledge to be just a metaphorical kind of knowledge.

If things are indeed like Sosa’s claim, this time, externalists might be challenged to defend the familiar ground that they have been using for their position. That is, if Sosa is right, virtually all of human knowledge is “screened” by reason in one way or the other even if the “origins” are different. Again, if we can set aside some important questions including whether this screening would be enough for “the contributing cause” of the belief, then Sosa’s view would eventually lead to the claim that there wouldn’t be any

34) Sosa, E., “Virtue Perspectivism: A Response to Foley and Fumerton,” Philosophical Issues, 5 Truth and
perceptual knowledge in the sense that externalists would like to use in order to make
their point (accessibility to the justifier is not necessary). Many alleged cases of
knowledge without accessible internal justification (for example, perceptual knowledge,
according to externalists) now become knowledge with coherence and internal
perspective in Sosa, thus knowledge with justification (reflective knowledge according to
Sosa).

Sosa’s questionable distinction of animal and reflective knowledge can be
summarized in the following.
[1] Animal knowledge is only a metaphorical sense of knowledge because it is merely
apt.
[2] Reflective knowledge is justified true belief, where justification is understood as a
belief’s being apt and also being coherently fit into the perspective.
[3] Typical perceptual beliefs (as well as others such as beliefs from memory and
introspection) are justified beliefs to a rational being, thus they should be considered as
cases of reflective knowledge.

Note that here Sosa’s verdict on the justificatory status of our perceptual beliefs is
the same as other reliabilists on justification such as Goldman. That is, they are justified.
However, the reason is quite different. For Goldman, the objective truth-conducivity (that
is, Sosa’s aptness) is why they are justified. Goldman believes that we cannot require
anything more than this because there indeed is nothing more in common in our ordinary
perceptual beliefs (if we include children’s, infants’ and dogs’ perceptual beliefs as we


125
should), yet it is intuitively certain that we (again, children, infants and dogs, too) know what our perception brings to us. Now, Sosa claims that there is more. At least in a case of a rational agent's typical perceptual beliefs—they are the beliefs that we would certainly consider as knowledge—the beliefs coherently fit the perspective. For this, there should be a perspectival grasp, i.e., understanding or awareness of the permissibility or reliability of the belief production under the given circumstance. (Thus, children's, infants' and dogs' perceptual beliefs are not included because their status as knowledge is not as certain as a rational agent's typical perceptual beliefs.) So, our perceptual beliefs are justified neither because our perception is objectively truth conducive (apt) alone nor because they coherently fit to the perspective alone. Instead, they are justified because perspectival grasp and coherence are added to the aptness of our perceptual beliefs. Thus, again, for Sosa, our perceptual beliefs are justified beliefs and cases of reflective knowledge. We now also understand why it is necessary for Sosa to argue that most of our typical perceptual beliefs and beliefs from introspection and from memory should not be considered as animal knowledge. It is because animal knowledge, i.e., apt belief without a perspectival grasp, is only a metaphorical sense of knowledge that applies literally to animals. (This would be what children, infants and dogs have, according to Sosa.) If Sosa is correct here, then he might be raising more than one critical point against Goldman and other reliabilists. First, Sosa might say that they fail to recognize the difference between a rational agent's perceptual belief and an animal's perceptual beliefs, namely the existence of perspectival grasp. Second, they fail to recognize the relatively dubious epistemic status of an animal's true perceptual beliefs. Whether they constitute knowledge or not is less certain than our perceptual beliefs. Finally, despite these
differences, they construct their reliabilist theory of justification (and knowledge, too) as if an animal’s perceptual beliefs were the typical case of the knowledge and a justified belief.

3. The Meta-Incoherence Problem: Norman, Samantha and Clairvoyance

3.1. The Problem and Sosa’s Solution

Sosa’s distinction of animal and reflective knowledge, his distinction of aptness and justification, and his understanding of perceptual belief and the role of reason in it are open to many challenges from critics. However, I do not want to press him in the direction that those notions are not theoretically rigorous enough. As I set the direction of my criticism earlier, here I will focus on whether Sosa’s particular blend of reliabilism indeed overcomes the difficulties of generic reliabilism. I believe that the theoretical difficulties of Sosa’s conception and distinctions can also be shown by checking his solutions to the problems of reliabilism. So, let’s consider how Sosa’s account deals with problems of reliabilism.

The first problem of reliabilism I discuss is the meta-incoherence problem. Bonjour has devised four different cases in order to point out that reliability is not a sufficient condition for justification. I will just briefly introduce two typical cases here, namely the case of Samantha and the case of Norman. Samantha, who has reliable clairvoyance, comes to have a belief that the president is now in New York using her clairvoyance. This belief is true and her clairvoyance is a reliable belief-forming method.
However, she is not aware of the fact that she has clairvoyance, nor that it is a reliable belief-forming method for her. Furthermore, she has massive evidence which indicates that the president is not in New York but in Washington D.C., such as a news report on presidential affairs, which is in fact fabricated. So, Samantha believes a true proposition due to a reliable clairvoyance without any evidence to believe so and with some evidence not to believe so.

The case of Norman is similar. He has a similar kind of true belief due to his perfectly reliable clairvoyance. However, in this case, unlike Samantha, Norman has no evidence for or against the truth of the proposition he believes. He also has no evidence for or against the possibility of such clairvoyance, and he has no evidence that suggests that he does or does not have such a power. In short, his belief that the president is in New York is neither undermined nor supported by any of the evidence that he possesses. Now, both in the case of Samantha and in the case of Norman, what Bonjour wants to conclude is that these beliefs are not justified, and this seems to have strong intuitive support. If so, it causes a problem in the reliabilist theory of justification because the unjustified beliefs of Samantha and Norma are the result of the reliable clairvoyance.

As a solution, reliabilists come up with the following condition. In order to be justified in believing that p, the belief should be a result of reliable process, and it also should be the case that there is no alternative reliable process available to the agent such

---

that if the agent were to use the alternative process, he would not have that belief.\textsuperscript{36} This will in fact be effective in handling the case of Samantha because we can agree on the existence of such an alternative process available to her there, namely believing testimony (news report) whose reliability is socially well established. However, it is debatable whether this additional condition can successfully deal with Norman’s belief.\textsuperscript{37} Goldman shows a sign of “bullet biting” and attempts to test externalist intuition here by drawing our attention to the fact that Norman’s clairvoyance belief is very similar to our other less sophisticated beliefs and their processes.\textsuperscript{38} For example, if Norman’s clairvoyance belief is based on some kind of “inner image” or an internal experience of the president in New York (although this image must have popped up in his mind), then it may remind us, again just for an example, of a belief about our past, which sometimes comes out of the blue. I may come to believe that the gift that I got in my 4\textsuperscript{th} birthday from my uncle was a toy truck wrapped with red and silver wrapping paper (with a bright image of that package) while I was taking shower last night. Suppose that my memory has actually stored this information accurately all along. My memory just brought it in my consciousness without some relevant context usually accompanied with such an occasion. Our intuition in this case seems to be at least little more generous than the clairvoyance case. By re-describing the process of clairvoyance in this way, Goldman hopes that at

\textsuperscript{36} For example, see Goldman, Alvin, \textit{Epistemology and Cognition} (Cambridge MA: Harvard University Press, 1980) p 63 and p 83 for his schematic construction of the undermining condition, and pp. 53 – 54 and pp. 109 - 113 for the application of this condition to various cases.

\textsuperscript{37} Reliabilists may think that there is an alternative process which satisfies above condition. For example, Marshall Swain suggests ‘the process of withholding beliefs in situations where there is no evidence to support them’ as a possible candidate.

least the initial hesitation expected in affirming that Norman’s belief could be justified (or could be knowledge) should be eroded somewhat. However, even so, it would not be enough to draw any wide and positive agreement that Norman’s belief is justified.

Now, Sosa’s account is surely in a better position to accommodate our ordinary intuition. According to Sosa’s virtue perspectivism, Samantha and Norman are both unjustified in believing that the president is in New York, just as our intuition affirms. The reason would be that they both fail to satisfy the perspective requirement. Recall that Sosa’s justification requires both aptness and perspective. Although clairvoyance can produce true beliefs and the beliefs so produced can be considered apt given that clairvoyance is a “disposition” to discern true from false (thus, clairvoyance would be an intellectually virtuous faculty), such a belief cannot be justified as long as there is a defect in perspective. This all sounds right. However, we can ask further, specifically concerning how exactly this belief fails to satisfy Sosa’s perspective requirement because, as we can expect, there could be more than one way that the perspective requirement is violated.

3.2. Does Samantha Have Animal Knowledge?

I will begin with asking about the knowledge status of the unjustified true belief that Samantha and Norman have. What I would like to draw attention to is that, according to Sosa’s account, both Samantha’s and Norman’s beliefs would have to be treated equally when their knowledge status is assessed. First, since both are unjustified beliefs (they both have problem in their perspectives), there’s no hope for them to be reflective

4. (Originally, Liaisons: Philosophy Meets the Cognitive and Social Sciences, by Alvin Goldman

130
knowledge. (Recall that reflective knowledge requires justification.) However, even if the beliefs are unjustified, the intellectual virtue of the clairvoyance belief (aptness) is nevertheless present in both cases. Given that truth-conducivity and perspective are two independent components of justification, both Samantha and Norman violate only one of them, namely the perspective requirement (although we can expect that the ways in which they violate this requirement might be different). Whatever happens in the perspective does not make their clairvoyance non-virtuous. Thus, Samantha's and Norman's beliefs are both apt. Sosa might want to say that they are accidentally virtuous. Be that as it may.

The point is that there seems to be no clear reason in Sosa's account that we should treat the way in which Samantha's belief is apt differently from the way in which Norman's belief is apt. Then, it looks that, if we apply Sosa's account to these cases, Samantha's and Norman's beliefs are both to be counted as animal knowledge on the basis of the truth-conducivity of clairvoyance. However, if this is the case, then it allows that, in general, even to the accidentally apt beliefs that worry Sosa, the status of animal knowledge can be conferred as long as they are (but happens to be) truth conducive (apt). In my opinion, this shouldn't be a welcome consequence for Sosa.\(^{(39)}\) One may claim that, at least concerning Norman, attributing the status of animal knowledge to his belief is acceptable. Well, I think it is debatable, at best. However, again my point is that, in Sosa, if Norman's belief is a case of animal knowledge, the same epistemic status should be attributed to Samantha's belief because they are equally apt. And this is where I see a

\(^{(39)}\) In fact, it is not clear from Sosa's texts, whether Sosa really wants to confer animal knowledge to Norman's, and especially to Samantha's belief. However, I do not see any clear explanation of how this
problem. It looks erroneous to me that we have to attribute animal knowledge to what Samantha has, even if it does not look to be so when we attribute it to what Norman has.

As Bonjour has originally intended in his examples, there is a difference between Samantha's epistemic circumstance and Norman's. Namely, Samantha has some evidence that indicates what she believes is not true, which is not just externally available (as the form of news report) but also internally accessible by her (as a form of "processed" internal state of Samantha) because, according to the example, she "has" that evidence. Despite the unclarity of what amounts to "having evidence," I understand this as something that the agent has an access to, i.e., that the agent can take notice under an appropriate circumstance. I think "having evidence" can be treated roughly the same as being in a certain reason state, which includes belief states and perceptual states, for example. Then, saying that Samantha "has" evidence for believing that the president is in Washington D.C., not in New York, should mean that this information is processed by Samantha, and in a form that Samantha can be aware of in an appropriate circumstance such as by reflecting on it. Norman does not have such evidence. That is, even if he is conscious about what he believes, he will not find any evidence that indicates the president is not in New York. The only fact that is accessible to Norman, if he is properly directed, is the fact that he does not have any evidence for his belief that the president is in New York. What I would like to know is whether this difference lies within the elements that determine the aptness of the belief for Sosa. In my opinion, that doesn't seem to be the case. The difference is in the way that Samantha and Norman are

should not be so, either. My claim is that a straightforward application of Sosa's account would allow
unjustified. It could be true that both of them are unjustified regardless of such a
difference, or maybe they are unjustified for different reasons because of such a
difference. However, their aptness would be in the same degree, and if so, attributing the
same kind of knowledge to both seems to be unavoidable. Sosa has to accept that each
has animal knowledge or that neither of them has knowledge of any kind.

Surely, animal knowledge is more or less a fuzzy notion. Even so, it still is a case
of knowledge, which should be more than a true belief. Then, it is questionable whether
the truth-conducivity is all that matters for knowledge in cases like Samantha, where
there is an obvious "defeater" outside the truth-conducivity. According to Sosa, this
"defeater" works against Samantha’s having reflective knowledge. That is fine. However,
isn’t this also working against Samantha to have any kind of knowledge, even a
"metaphoric" one? In my opinion, saying that Samantha “knows” about where the
president is in the circumstance described above (that is, in the face of such counter-
evidence that is available to her and accessible by reflection by her) even in a very
minimal, “metaphoric” and “servomechanic” sense, seems to be a serious compromise in
our notion of knowledge.40 Consider even a literal case of animal knowledge of a dog,
for example. Let’s stipulate that the dog has some minimal intellectual capacity just
each to form a first-order belief but lacks any sophisticated meta-level reasoning
ability. Now suppose that, this dog forms a belief that his master is in his bedroom due to

Samantha’s belief to be animal knowledge, and if so, that could be a problem.

40 I think that, despite the alleged difference, the case of Norman deserves the same verdict eventually.
However, assuming the difference here would be innocuous. My complaint here is mainly about the
epistemic status of Samantha’s belief, and I am not claiming that Norman’s case has to be knowledge while
Samantha’s is not. Rather, my point is that intuitively it is even harder to maintain that Samantha has animal

133

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
his reliable olfactory perception, and the belief is true. However, consider that this dog is also hearing (thus processing) the master's recorded voice coming from the living room, which he cannot recognize any other way but as his master's real voice. Despite this, the dog still believes that the master is in the bedroom. Now, the question is whether we want to think that this dog has animal knowledge concerning where his master is. We can surely agree that the dog has a true belief, but for knowledge, we need to see something more. What else does he have? He has a reliable olfactory perception that produces the belief, and presumably, that is also holding the belief throughout the moments that he is experiencing the master's voice. Would it be enough to bump this belief into the status of knowledge? Not enough to make it a full-fledged knowledge, Sosa would say, but at least some. So, that's why it would qualify as animal knowledge.

However, having the belief that the master is in the bedroom looks still "accidental" even for a creature like this. Why isn't the equally reliable belief-forming process (the dog's hearing) activated? If it had been, the dog would have ended up either with a belief that the master is in the living room instead of the belief that the master is in the bedroom, or possibly with both beliefs (since we assumed that he has no meta-level reasoning capacity). In this sense, the true belief state that the dog has is achieved accidentally. Similar things can be said about Samantha. The difference is that the source of accidentality in Samantha's case comes from the meta-level, but in both cases, the location of the source of this accidentality (or the defeater) is outside the causal chain to produce the target belief. However, we find it difficult to assign a status of animal knowledge (or any kind of knowledge) than Norman. At least in Norman, some might want to think that
knowledge to those cases solely on the basis of the appropriateness of the causal chain. If what the agent has is an accidental true belief, it would be highly controversial whether this belief is compatible with any sense of knowledge.

If the bottom line concerning the distinction between knowledge and mere true beliefs is our intuition that “accidental” true beliefs cannot be knowledge, admitting Samantha’s belief as any kind of knowledge would bother that intuition greatly due to the colorful “accidentality” in her belief. Surely, accidentality is an undefined notion, which could be understood in variety of ways. And, it wouldn’t be true that we have to get rid of all kinds of accidentality from a belief in order to have knowledge. Even so, there is a certain kind and degree of “accidentality” that even the weakest sense of knowledge cannot allow in it. Here, at least one kind of accidentality I spot in Samantha’s case is related with her awareness of the existence of the counter-evidence and undermining effect of that counter-evidence. It is true that a certain kind of accidentality is taken care of by the objective reliability of Samantha’s clairvoyance, the virtuous faculty that produces her true belief. However, for knowledge, a different kind of accidental nature found in Samantha’s belief should be eliminated too. Namely, the accidentality of maintaining a belief that “does not fit coherently in the perspective” (due to the agent’s awareness of the existence of the defeater and undermining effect of the defeater) given that the agent is able to recognize it as incoherent, cannot simply be overruled by the other kind of epistemic merit coming from objective truth-conducivity. However, apparently, Sosa has to assign the animal knowledge status to any such accidentally apt
belief because those defeaters do not affect the aptness of the belief. My contention here is that, even if truth-conducivity is a necessary element for any type of knowledge (both for animal and reflective knowledge) it may not be a sufficient element even for animal knowledge, not to mention for reflective knowledge. My worry is that Sosa’s category called animal knowledge solely based on intellectual virtue (truth-conducivity and aptness) has the danger of inviting not just many “less than obvious cases” of knowledge (or may be borderline cases of knowledge), such as Norman, but also many cases of “non-knowledge,” such as Samantha, into this category.

Now, let’s ask a question of how exactly beliefs fail to satisfy the perspective requirement. Again, Sosa does not explain much about this. At least concerning Samantha and Norman, what Sosa can say is that they do not have appropriate perspective. They do not have access to the relevant circumstances (C) under which the belief is produced and the field (F) that the proposition belongs to, thus they are not aware of the permissibility, reliability or relevant intellectual virtue concerning her belief. However, not all cases in which the agent lacks perspective would be the same.

One case where a belief cannot meet this requirement because it lacks any perspective at all is due to the agent’s inability to have such a perspective. We understand that this is probably what Sosa is thinking when he acknowledges animal knowledge. Given that small children are not able to have a perspective on their beliefs but still have reliably working (virtuous) perception that produces beliefs, their true but merely apt perceptual beliefs in this particular way are not justified beliefs. However, they still

41) This would be a different kind of accidentality that we can spot in the dog case.
constitute a case of animal knowledge. I am sympathetic to Sosa's idea to attribute a somewhat different sense of knowledge (but still not giving up the idea of knowledge) to this and similar cases with animals and infants, as long as the objective truth-conducivity condition is met. In these cases, we see that their beliefs are reliable indicators of truth under such circumstances (they are the product of intellectual virtue in Sosa's term). Nevertheless, they are 'merely' reliable indicators of truth, which we want to distinguish from a paradigm case of knowledge. However, we also notice that their not being more than a 'mere' indicator is due to the lack of ability as an epistemic agent to do anything more than be a reliable indicator. If the principle of "ought implies can" could be applied here, we cannot blame this type of agent on epistemic grounds. There is no way to find any defect in the mechanism that is used for the belief production, either. All of these make up quite a mixture of epistemic elements both positive and negative. Moreover, in my opinion, they are mixed together in a rather balanced way. Thus, the invention of a somewhat vague but intuitively appealing category such as animal knowledge could be defended only in these limited cases, if it can be defended at all.

However, Samantha's clairvoyance belief is not like a small child's true perceptual beliefs. Most relevantly, as a cognitive agent, she is "expected" to detect the inconsistency concerning what she believes, which is a cognitive function that we cannot expect from a small child. In other words, unlike small children, Samantha's cognitive system as a whole (or maybe the epistemic agent as a whole) is seriously defective although her belief-generating faculty (if it is the reliable clairvoyance) may be claimed to
be flawless. Children’s perceptual beliefs may well fit coherently within their narrow cognitive repertoires, while Samantha’s clairvoyance belief does not. Children may not be epistemically blameworthy by holding such perceptual beliefs without a reason, while Samantha is. Then, Samantha’s alleged problem in her perspective, which prevents her belief from being reflective knowledge, is not the same kind of problem found in small children’s perspective, which justifies small children’s beliefs to have animal knowledge status. If so, conferring the status of animal knowledge to both seems to be implausible. If Sosa’s position implies that, as long as a belief is apt, it qualifies at least as animal knowledge regardless of what goes wrong at the level of perspective, then it is problematic. The cases where Sosa’s animal knowledge may have some support from our intuition are where the agent violates the perspective requirement due to his general inability to do so as in the case of small children, or literally, animals. Samantha is not one of those cases.

3.3. "Incomplete" Perspective

So far, I have considered that the same type of beliefs in Sosa’s account, namely beliefs that are truth conducive but lack a perspective, cannot be treated all in the same way in terms of their epistemic status. Depending on how it fails to satisfy the perspective

---

42) I guess a similar thing can be said to the above example of a dog’s belief. That is, the dog as a whole is defective, although all he has is first order belief-forming mechanism. If so, the space that we can charitably allow for Sosa’s animal knowledge would be even narrower because, even among the reliably produced true beliefs of the agent who lacks the ability to satisfy Sosa’s justification condition (such as an infant and the animal), there would be a belief that cannot be considered as animal knowledge.

43) Sosa may claim that, due to a certain feature contained in Samantha’s failure to meet the perspective requirement, the belief does not even qualify as animal knowledge. I will soon consider how he can make sense of this option within Sosa’s account as well as the problems that beset it.
requirement, that is, depending on the nature of the problem in the perspective requirement, it may or may not qualify as animal knowledge. Now, a similar kind of problem can be spotted in terms of how exactly the perspective requirement should be met. I would call this a problem of “incomplete” perspective because, in my opinion, Sosa’s perspective requirement can be satisfied by a less than complete perspective. Here, an “incomplete perspective” means the case where the agent has a perspective on the permissibility, reliability, or the virtue concerning the target belief yet he is unaware of the “real” reason (or the “actual” C and F) for why the target belief is permissible, reliably produced, or virtuously produced, under the given circumstance. So, we may also add that, in this case, a belief would fit coherently within the perspective but only “up to a certain degree.” However, if we allow such an incomplete perspective, there would be a case that we wouldn’t want to consider as a case of a justified belief and reflective knowledge, yet it satisfies Sosa’s conditions for justification and reflective knowledge. This problem is not easily fixed by simply asking a perspective to be based on a complete awareness of the actual epistemic circumstance because the incompleteness is often our epistemic reality when the perspectival awareness is the issue.

Asking for the “complete” perspective is surely unfair and Sosa does not want this. Recall what he says.

For reflective knowledge one not only must believe out of virtue. One must also be aware of doing so. Of course, one need not know with precision and detail the relevant C and F. Some grasp of them is required, however, even if it remains sketchy and generic.44

Only under this lowered standard of accessibility, can Sosa make sense of his claim that most rational agents have at least a certain degree of accessibility to C and F when they have perceptual beliefs, and thus their perceptual beliefs are justified and become reflective knowledge. Otherwise, the requirement is set too high that it would be almost impossible to meet the requirement for typical epistemic agents even though they are rational.

However, if what is required in our perceptual beliefs are just some “sketchy” ideas on why the target belief is permissible, reliably produced or virtuously produced, then a different problem is expected. Let’s suppose that Samantha somehow acquires some perspectival awareness of the epistemic permissibility of her clairvoyance belief. Suppose that she now is aware of the following: “I have reliable clairvoyance, my belief that the president is in New York is produced by my clairvoyance, and therefore, it is likely to be true.” If we change the example in this way, now Samantha is considered as having a grasp of the perspective that licenses the aptness of the clairvoyance belief. Her belief coherently fits within this perspective. So, it may well receive the status of a justified belief and reflective knowledge according to Sosa’s account. However, we may still think about some further variations of the story concerning how Samantha comes to be aware of those facts that constitute her perspective. I am sure that these variations are relevant for our epistemic evaluation. Thus, for example, suppose that Samantha’s fantasy that she is an alien spy from Alpha Centauri makes her have such perspectival grasp concerning her clairvoyance. In this case, should we say that her belief is justified and reflective knowledge? I doubt it. This belief is far from being knowledge. However, this
revised Samantha’s belief would satisfy Sosa’s conditions for justification and reflective knowledge because of the partial and incomplete perspective that Samantha has. How can Sosa’s account exclude this kind of case from being justified?

As we can see, here Sosa’s perspective requirement seems to create a dilemma. Sosa cannot ask a complete perspective unless he wants to declare that a large number of our normal perceptual beliefs are not knowledge. However, an incomplete perspective is not at all helpful for Sosa’s purpose. It cannot prevent “accidentally apt” belief from being knowledge. In cases like the revised Samantha, it is unclear what would be the benefit of the incomplete perspective like hers, which is merely saying that her target belief is reliably produced. I do not deny that it is a positive epistemic element. However, it’s not sufficient to be the condition that can be added to apt beliefs and make them reflective knowledge. There seems to be no principled way to decide how much perspective would be sketchy and generic enough to be found in most typical non-sophisticated human perceptual beliefs, yet to be considered as the satisfaction of the perspective requirement for preventing accidental apt belief from being knowledge.

3.4. Coherence-Seeking Reason as a Meta-Faculty

In the case of original Samantha, the salient feature is that there is no perspective for her belief to fit in. In other words, Samantha has no awareness of the reliability concerning her belief production. However, I have argued that, if Sosa relies on this fact as the source of the epistemic mishap in this and other similar cases, then at least two problems would be foreseeable. One would be the problem of having incomplete awareness of the reliability. The other would be the problem of the animal knowledge
status that I think Sosa has to confer to cases like Samantha. Now, one might suggest that there is a way in which Sosa does not have to face these problems. According to this idea, Samantha’s problem can be described not as “a (correct) belief with no or wrong perspective” but rather as “an incorrect belief [a belief that does not coherently fit] in a given perspective.” That is, instead of seeing Samantha’s clairvoyant true belief as a “constant” and looking for the existence or non-existence of the perspective that endorses this belief, this interpretation considers Samantha’s current cognitive state as her perspective. Then, the existence of a belief that does not coherently fit to this perspective such as Samantha’s belief (because it can at best be described as a groundless belief from Samantha’s perspective⁴⁵) itself could be a violation of Sosa’s perspective requirement.

A ground for this interpretation is Sosa’s idea of meta-faculty. Following Sosa, we can suppose that we have a meta-faculty, “the overarching faculty of faculties”⁴⁶ or “coherence-seeking reason”⁴⁷ as Sosa names it, that scans the outcome of the first-order faculty. If this meta-faculty is working virtuously as it should be in Sosa, the agent ends up with having only beliefs that fit coherently in the given perspective. Thus, in Samantha’s case, if the meta-faculty, the “coherence-seeking reason,” had been virtuous, Samantha wouldn’t have ended up with the belief that the president is in New York. The

⁴⁵) Here, one could have even stronger case if the counter-evidence that Samantha “has” is somehow participating to her perspective. However, even if such a counter-evidence does not exist (as in the case of Norman), there would be no difficulty to spot the problem of Samantha’s belief as long as we can legitimately assume that Samantha is just like a normal epistemic agent who has some ideas of what to believe and what not as a part of her perspective.
⁴⁶) KIP. p. 284 ("Intellectual Virtue in Perspective")
⁴⁷) KIP. p. 291. ("Intellectual Virtue in Perspective")
existence of such a belief shows that the meta-faculty involved here is non-virtuous. For this reason, Samantha’s (and Norman’s too) belief does not have justification.

This is a possible interpretation of Sosa’s approach to justification reconstructed using the concept of meta-faculty. As I mentioned in the previous section, Sosa seems to think that there always is coherence-seeking reason as a meta-faculty working in our belief formation together with the first order faculty such as perception. If so, meeting the perspective requirement, and in turn, being justified, can be considered as a function of such a two-fold faculty (perception + reason). Indeed, this idea can explain many things in favor of Sosa. First, the problem of having an incoherent belief is now attributed to the defect of this meta-faculty because just like achieving a true belief is the goal of the first order belief-forming faculty, now we can suppose that achieving true and coherently fitting belief as the result would be the desirable goal when this “faculty of faculties” is involved. Second, if we follow this idea, it might be able to save Sosa from the awkward position of admitting animal knowledge status to Samantha’s belief. Relying on the success ratio, we can tell whether this coherence-seeking reason is intellectually virtuous in a given environment in a similar way that we can tell the faculty of perception is virtuous in a given environment. If it is not virtuous as in the case of Samantha (because it has generated a belief that is true but does not coherently fit to the perspective), then the outcome belief might be thought of as not virtuous as a whole despite the (partial) virtuousness of the clairvoyance. That is, it is not the reliable clairvoyance alone but the clairvoyance plus the unsuccessful coherence-seeking reason that has produced this belief. So, in this sense, the belief does not result from intellectual virtue, not to mention that it does not satisfy the perspective requirement. If we can use the term “aptness” here,
then, we have no reason to assign animal knowledge to Samantha's now *inapt* belief due to the non-virtue of the meta-faculty that produces this belief. She has neither reflective knowledge nor animal knowledge.

However, this seemingly convincing approach is not free from difficulty. First, I can surely expect some doubt about the criteria that make a meta-faculty virtuous. Is reliability still to be the criterion as in the case of the first-order faculty? If so, why can't we say that Samantha's meta-faculty is virtuous (reliable) just by letting her have the true belief that the president is in New York? However, saying that Samantha's belief is the result of a virtuous faculty (a combination of the first-order clairvoyance and the meta-faculty) is admitting that her belief is justified, which cannot be Sosa's verdict. So, that is why I have suggested above a different goal other than truth for the success of the meta-faculty. That is, given that Sosa uses "coherence-seeking reason" as a possible name of the meta-faculty, I have supposed that "generating (endorsing, screening) beliefs that coherently fit within the given perspective" is the desirable goal of Sosa's meta-faculty (just like generating true belief is the desirable goal of the first-order faculty). Also I have supposed that the high success ratio in producing such a belief is the criterion for a meta-faculty to be virtuous. That is, in addition to being a product of the first-order intellectual virtue (reliability), if a belief is also a product of a "disposition" or "competence" that discerns "coherently fitting beliefs" from "un-fitting beliefs" (namely, the meta-intellectual virtue), then we may say that the belief is a product of intellectual virtue, good

---

48) If this meta-faculty can have a belief as its input, it could be understood as an "input-dependent reliability." That is, the false input does not affect the evaluation of reliability of a faculty.
enough to have a justification. Would this much of speculation eventually save Sosa’s account? A possible problem is still lurking.

One of the basic assumptions in reliabilism and in Sosa’s virtue account approach is that the reliable faculty or the intellectual virtue is a general disposition, which does not guarantee 100% success. The same would be true of the meta-belief-forming faculty (coherence-seeking reason). This means that there is room for error in the performance of the given intellectual virtue. A virtuous meta-faculty may endorse an “unfit” belief as if it is the one that coherently fits in the perspective. Still, the faculty can be virtuous, just like a first-order faculty can be virtuous even if it may produce a false belief in a rare occasion. We cannot deny the existence of a whole disposition just because of one undesirable instance. For the same reason, a non-virtuous meta-faculty doesn’t mean that the faculty has absolutely no propensity to endorse fitting beliefs (that is, as a faculty that has a propensity to allow unfitting beliefs all the time). Now the question is, what if Samantha’s “unfitting” belief that the president is in New York is the result of one mistake of a virtuous faculty? For example, suppose that Samantha is usually an impeccable reasoner. She usually believes something only when she sees some coherent reasons in her perspective. Thus, if it were a usual case, Samantha wouldn’t have ended up having such a belief. However, in one particular incident only, for some unexplainable reason, she has made a mistake. Samantha’s reasoning simply fails to recognize the incoherence of having such a belief in this case only. By adding a story like this, it seems possible that Samantha’s meta-faculty is virtuous (allowing “fitting” beliefs in most of the similar circumstances) despite one mistake. However, I don’t think that such a story would change our intuition concerning Samantha’s belief. Samantha’s belief that the
president is in New York is still epistemically problematic and unjustified even if Samantha's coherence-seeking reasoning is "generally" virtuous. For me, it is hard to accept that Samantha's true but "unfitting" belief becomes a justified belief (and even a case of reflective knowledge) because this "unfitting" incident is the result of one occasional mistake of her otherwise perfectly virtuous coherence-seeking reason. However, as long as the faculty (either it is a first-order faculty or meta-faculty, or even it is a faculty made by a combination of both) is virtuous, it seems that Sosa no longer has a ground to maintain that Samantha's belief is problematic. In fact, it would make Samantha's belief justified in Sosa.

In general, I find it strange that an epistemic mistake in the meta-faculty level can be ignored in our justification and knowledge assessment. In case of the first-order faculty, a mistake from a usually reliable faculty can be considered as a "justified false" belief depending on what is happening in the meta-faculty level or the perspective level. However, damage caused by a breakdown of "coherence-seeking reason" itself cannot be restored by appealing to the general disposition of that faculty. Thus, here we find another tension in Sosa's virtue perspectivism. In order to satisfy the reliabilist constraint, intellectual virtue of the meta-faculty should be understood as a disposition of a faculty. However, in doing so, we probably have to sacrifice our "evidentialist" intuition concerning justification, which is no less strong than the reliabilist intuition. In evidentialism, justification is a function of whether a given belief fits to the evidence that the agent has at the point of evaluation. Thus, naturally, justification is something that we can have or lose at each given time-slice by satisfying, or failing to satisfy, certain conditions. If Sosa's intellectual virtue of a meta-faculty (coherence-seeking reason)
cannot serve for this intuition, his account is still missing an important aspect of epistemic justification even with the all the additional requirements such as perspectival grasp and coherence within the perspective. It is these additional requirements that are intended to serve for those who have non-reliabilist intuitions concerning justification.

Let me summarize the discussion of this section. Sosa's verdict that Norman's and Samantha's clairvoyance beliefs are unjustified is the one we should expect from such an account. However, it is somewhat unclear how Samantha's belief violates Sosa's conditions for justification (aptness and perspective). If Samantha violates only the perspective requirement, then her apt belief should be considered as animal knowledge according to Sosa. However, I have argued that allowing animal knowledge status to Samantha's belief can be problematic. Second, it is also unclear how the perspective requirement should normally be met. Sosa allows a rough grasp of the permissibility of the belief in a given circumstance as being enough for the satisfaction of the perspective requirement. This is precisely because asking a complete awareness of the perspective for justification is practically impossible. However, an incomplete and partial awareness of the perspective cannot always provide the ground for excluding "accidentally apt" beliefs from knowledge because such an incomplete awareness may well be an awareness of an incorrect perspective. Even if what the agent is aware of is a legitimate portion of the correct perspective, we would have trouble excluding accidental aptness with such a "sketchy" awareness. This is because there could be no phenomenal difference between having a portion of the correct perspective and having a portion of an incorrect perspective, which happens to be the same as the former. Finally, utilizing the idea of a
meta-faculty, namely coherence-seeking reason, may have some merits but I have argued that this does not secure the success of his account.

4. Justification in the New Evil Demon World

4.1. The New Evil Demon Objection against Reliabilism

The other criticism against a reliabilist theory of justification is based on the familiar evil demon scenario. The basic idea presented by Cohen49 is this. Suppose an epistemic agent S is justified in having, for example, a perceptual belief that p in a normal world, on the basis of whatever reason R that S has. Then, his counterpart S* in the demon world (or S who now becomes a victim of demon’s deception unbeknownst him), who maintains exactly the same internal features as S including having R, should be equally justified in believing that p. The problem is that, due to a demon’s manipulation, all of S*'s perceptual beliefs would be false. Thus, they are considered as being produced by an unreliable mechanism. Since there is a justified belief deriving from unreliable mechanism, this is considered as a counterexample against the main claim of reliabilism that objective reliability is necessary for epistemic justification. In this section, I will consider how Sosa’s virtue perspectivism handles this case, and whether there can be any difficulty in Sosa’s position. I will begin with some discussion of the example itself. Although the New Evil Demon (NED form now on) example, as well as the charge against reliabilism based on the example, might be intuitively appealing even only with

the brief description above, there are details to be worked out in order to see whether it really constitutes a legitimate counter example against reliabilism.

Possibly, there is more than one way to set up the NED example. Typically, one considers the demon's victim to be in a similar epistemic state as the brain in a vat. Or any person whose mental world is completely disconnected from the outside world would work in the same way. Due to such a disconnection, the victim or the brain (let's call it Edmund) cannot really "perceive" the actual world. Edmund is an isolated being. Then comes the demon's deception concerning Edmund's "perceptual experience." What Edmund "perceives" does not originate from the things in the actual world but perceptual images (or percepts) that are provided by the demon through direct stimulation of the brain or through some other deceptive methods. Thus, despite the disconnection to the actual world, Edmund continues to "think" or "believe" in accordance with what he seems to perceive.

Before we go any further, the nature of the demon's manipulation needs to be clarified. For example, can the demon make Edmund believe that $1 + 1 = 3$? Or, more relevantly to the case of perceptual belief, can the demon make Edmund suddenly believe that it's snowing outside even without providing corresponding "snow-watching perceptual experience" in Edmund? Well, the demon may or may not have such a power, but for the purpose of the NED example, I think the critics of reliabilism should withhold giving such a power to the demon. If we may suppose that the demon can freely implant various beliefs directly into Edmund's cognitive system and Edmund has nothing to do about it except for just having those beliefs, the critics would face a difficulty in convincing us later that Edmund's beliefs are eventually justified. That is, if what the
demon can directly control is merely Edmund's beliefs rather than percepts (experiences), it follows that there is no way for Edmund not to believe what the demon makes him believe. Edmund would be just helpless in believing anything and everything as the demon feeds him. If he believes things regardless of what kind of evidence or counter evidence he has for them, or believes even the most obvious contradictions, then it would be hard to expect that our intuition considers these beliefs to be justified. The demon may be able to maintain Edmund's cognitive system impeccably with a perfect coherence and without any sign of contradiction. The demon may even be able to directly provide to Edmund necessary meta-beliefs, meta-meta-beliefs, and so on. However, if the demon retains the power to change Edmund's beliefs, and such a change would not be noticed by Edmund (except for the result of the change, that is, the belief itself), then I am sure that our ordinary intuition would have trouble to find that his beliefs are justified.

Also, it might give us pause when we think of whether such a belief-forming process in a demon world can be called perception. The essential characteristic of perception as a faculty is its capacity to interact with the external world. This is the very capacity that Edmund's relevant faculties do not have. Since Edmund is cut off from outside stimuli and cannot respond to them, I would like to think that this faculty cannot be properly called perception. However, even if it is not proper to call Edmund's belief-forming process perception, critics would not be bothered by that. They can make it at least an unreliable belief-forming process as long as reliability is understood in terms of

50 This may show that the notion of "objective" coherence, although an epistemically desirable element to have in one's belief system, is not exactly the crucial source of our "intuitive" idea of justification. For example, an objective coherence can be obtained purely by chance, even if the demon uses a "randomizer (a
the true belief production ratio of the whole process assessed in the world that the process is used.

Now, the critics have to establish two claims. One is that Edmund's beliefs are not likely to be true (that is, they are either produced by an unreliable process or fail to be a reliable indicator of truth). The other is that Edmund’s beliefs are epistemically justified. Let’s think about the first claim. In order to show that Edmund’s beliefs are produced by an unreliable process, critics rely on the fact that Edmund’s beliefs are mostly false. For instance, at t, if Edmund believes that the red sun is setting behind the purple mountain ridges (on the basis of the percepts which he thinks he is getting from his perceptual apparatus), that is a false belief because there is no state of affairs which makes the proposition (that the red sun is setting behind the mountain ridges) true at t. If he feels that something is poking him on his back, thus he believes so, that would be a false belief too, because nothing is in fact poking him and he does not have a back. Since it is true that Edmund is having perceptual and sensational experiences (such as pain from poking) even if the demon is the source of such experiences, if Edmund’s beliefs were about those experiences (for example, a belief that I am experiencing pain or a belief that I am experiencing reddish), these beliefs could be true. However, as we normally think, Edmund’s beliefs are mostly about the external world and about the existence of what he perceives. Thus, the critics may safely say that Edmund's "pseudo-perceptual" beliefs are

---

7) It might be possible that Edmund, the brain, happens to have a true belief on such matters. Suppose that he is placed in front of the setting sun at t. However, whether these states of affairs make Edmund’s belief true may well be controversial, especially, when we consider that his belief is in fact not about the very states of affairs.
mostly, if not all, false. And, from this fact, it follows that the process that is responsible for generating all those false beliefs is unreliable.

Note that, for this claim, a certain world is assumed. That is, there must be a world in which the falsity (or truth) of Edmund's beliefs is determined. In this case, it's the actual world where Edmund, the brain, is located. For simplicity's sake, suppose Edmund, the brain, is inside an opaque vat, which contains a bit of dark empty space filled with fluid. Then, that should be Edmund's actual world, and naturally, all of his perceptual beliefs about the sun and the mountains are false in this world. So, in this line of thought, "the demon world," where his beliefs have no tendency to be true, would be the inside of the vat. However, is it correct? Suppose that the vat is large enough to hold a whole person together with Edmund, so we put a normal person there. Then, the normal person would see and feel darkness, fluid and maybe the brain. Is this person in "the demon world"?

It seems to me that, for Sosa, this is an important question because many things are depending on the notion of the "world" (either a normal world or a demon world). However, note that, for the sake of the criticism, it is not necessary to stipulate "the demon world." In other words, it simply does not matter for them where the demon victim is (if that's what it is meant by "world") because the relevant feature for the NED example is that the victim is in a state of disconnection from the world. All that is needed for the example is a world where Edmund's beliefs are sufficiently false. Therefore, the demon world could literally mean the world where the demon (or a team of mad scientists) is working. If the demon is operating on the brain in this world, then critics wouldn't mind calling this world a demon world (if they have to) because this is the
world where Edmund has a lot of false beliefs. As long as they can also consistently claim that Edmund's beliefs are justified in that world (and there is no reason to think they can't with their conception of justification), it would not matter whether the world the brain is actually in is perfectly normal to some others. (For example, "the demon world" in this sense is a normal world to the mad scientists, or to the man in the vat who is watching a brain.)

Now, Edmund has to be epistemically justified in whatever world they pick as the world in which Edmund is unreliable. Again, suppose that the world is the inside of the vat. Is Edmund, being inside the vat, justified in believing a false proposition that the sun is setting on the basis of percepts of the sun and the mountains provided by the demon? This is a tricky question. Some reliabilists do believe that the intuition that Edmund's beliefs are not justified in that circumstance is at least competitive, if not predominant, against the critics' intuition that they are justified. Of course, critics have more to say about this case. Crucially, the actual world where Edmund is in is not the world he thinks he is in. And, by hypothesis, this is not something Edmund can discover. Thus, from Edmund's subjective point of view, how he perceives things is indistinguishable from the ways in which a normal person perceives things. That is, when Edmund has false perceptual beliefs, what is going on "inside" of him is exactly the same as what is happening "inside" of a normal person when he has a true belief of the same content. Therefore, whatever we think of as epistemically positive elements in the internal process of a normal person’s perceptual belief-forming process should all be found in Edmund, too. And, this would be enough for critics to say that Edmund's beliefs are as much
justified as a normal person's perceptual beliefs. Thus, here finally is the critic's case where a belief is justified although it is unreliably produced.

Since the critics of reliabilism do not provide the specific property by virtue of which any belief is epistemically justified, there surely is something unclear about their claim that Edmund's beliefs are epistemically justified in the demon world. However, as I described above, the critic's claim and the objection can still be maintained by relying on the fact that Edmund and the typical citizen of the normal world are exactly same "internally" (including the internal process of forming beliefs from experiences, their accessible reasons for a belief and their awareness of the coherence among beliefs). However, obviously this claim presupposes that all that is relevant for epistemic justification is internal, which begs the question against reliabilists. For reliabilists, Edmund might not be justified exactly because of the "external" element that constitutes the difference between Edmund and a normal person. Of course, critics would think that they have intuition on their side in any case. That is, Edmund is justified intuitively. However, if it comes down to the level of intuition, it is possible for reliabilists not to give up their main claim. Rather, they can avoid the charge by providing a reliabilist explanation for why we have such an intuition. This, I think, is what Sosa is doing about this case.

To sum up the discussion so far, the following scenario seems to be preferable especially when the victim in the NED example needs to be prima facie justified: the demon manipulates only the perceptual experiences of Edmund, but not the belief itself. The thought here is that we may be able to divide our normal perceptual belief-forming process into two distinguishable sub-processes. One sub-process, which begins from the
contact of external things (stimuli) using perceptual organs such as eyes and ears, is what leads to perceptual experiences. The other sub-process begins from these perceptual experiences and leads to beliefs. The upshot of the NED example then can be restated as follows: The demon intervenes in the first sub-process by his deception, disconnecting Edmund's perceptual organs with external stimuli and creating pseudo perceptual experiences instead, which have nothing to do with the external world. However, the second sub-process (from experiences to beliefs) is supposed to remain intact. It should be considered as Edmund's ability\^{52} despite the demon's manipulation. I have no intention to claim that this distinction reveals the real nature of how our perceptual beliefs are formed. However, apparently, this is what the critics presuppose in the NED example in order to have what they want to have, i.e., a belief that is produced by an unreliable process, which nevertheless has a ground for being justified. Furthermore, the way Sosa handles this case shows that he generally understands perceptual belief formation in this way.

There are more things to be cleared in order to conclude that this is a successful counterexample to show that reliability is not necessary for justification. First, it is debatable if the justification Edmund allegedly has is the so-called epistemic justification (justification which is asked for in order to make a true belief a case of knowledge). For example, it may be claimed by reliabilists that what Edmund has is a very special sense of justification that does not have much to do with knowledge.\^{53} If so, reliabilists wouldn't

---

\(^{52}\) On the contrary, Edmund has no ability to contact the external world using his eyes.

\(^{53}\) Surely, there would be more than one sense of justification, some of which are not considered as necessary for knowledge. A justification in a “deontic” sense could be argued as one of them.
get serious damage from the NED problem. Second, for the same reason, this example shouldn't be considered as a direct proof that there must be some “internalistic” elements in epistemic justification. At least, we should wait until we can find out exactly what kind of epistemically positive elements (there are surely more than one) we are talking about, other than reliability, when we say that Edmund's beliefs are justified just like a normal person. Finally, it is possible that the idea of reliability might be still playing the central role in our intuitive response that Edmund is justified. That is, despite the outcome (false beliefs in the world where Edmund is actually in), we might fail to fully appreciate Edmund’s perceptual process as an unreliable one. We might rather consider this demon world circumstance merely as an environmental glitch working against Edmund’s otherwise reliable processes (that is, reliable if there were no demon). Although many would focus on the property like “being epistemically responsible” as the source of an epistemically positive feature of Edmund’s beliefs, which can eventually allow his beliefs to be justified, describing him as “an otherwise reliable agent” seems to be equally legitimate. Then, it is possible that the justificatory status that we confer to Edmund’s beliefs could be the result of focusing on the “otherwise reliable” features of his belief-forming process more heavily than the actual “unreliability” of the process in a given environment. If this is the case, it might be premature to conclude that reliability is not necessary for justification because it looks as though we are still relying on the notion of reliability in order to determine the status of justification. Sosa’s solution to the NED objection is closely related to this point.
4.2. Sosa on the New Evil Demon Objection

Sosa is trying to combine internalism and externalism in his notion of justification. His idea is that justification is achieved when the belief satisfies both internal and external requirements. In my previous discussions, I have tried the following two slightly different versions of understanding concerning how Sosa’s notion of justification attempts to satisfy both internal and external requirements. First, we can interpret Sosa in such a way that a belief is justified when a perspectival requirement (internalism) is satisfied by a belief that is already apt due to the intellectual virtue in its production (externalism). Call this version (A). Or, it could be that a belief is justified when the belief is produced by an intellectually virtuous faculty as a whole, which includes the meta-faculty. The virtue or aptness of this faculty represents externalism but the nature of the meta-faculty requires internalism. Let’s call this (B). It is likely that (A) and (B) would mean one and the same thing for Sosa. However, the difference between (A) and (B) is this. In (A), the perspective that the agent has for the belief is not considered as a part of the belief-forming mechanism. It is rather considered as an extra requirement, which has nothing to do with actual truth likelihood of the belief. So, truth-conducivity is not an issue in this case. The reason why perspective is required is due to something other than truth-conducivity. (It is to prevent accidentally apt belief from being knowledge.) On the other hand, in (B), ‘believing with a perspective,’ ‘believing only the things that coherently fit in the perspective,’ or ‘being endorsed by the faculty of coherence-seeking reason,’ is considered as the part of mechanism that produces the belief. Let’s take our normal perceptual beliefs as an example to see how they are justified. If we explain it using the interpretation (A), it would go like this. First,
intellectual virtue (truth-conducivity) is secured by the reliability of the belief generating faculty, that is, perception. And then, the perspective requirement is met when the agent has a perspective about his perceptual belief and its formation (by being aware of the reliability of his perception). Thus, a perceptual belief that has a perspective is justified. Under (B), in addition to the reliability of perception, the meta-faculty as a whole, namely the coherence-seeking reason, is supposed to be reliable too. Or, if reliability is not the proper criterion for the meta-faculty, we may replace it with a success ratio for the meta-faculty to get its desirable goal. Of course, in this case, the fact that the meta-faculty is involved in the belief production guarantees the satisfaction of the perspective requirement. (Sosa's claim that most of typical human being's perceptual knowledge satisfies the perspective requirement best makes sense under this interpretation.) So, again, a perceptual belief, being helped by the meta-faculty in its production, is justified. It is unclear in Sosa's texts which alternative Sosa accepts. However, regardless of Sosa's exact view, Sosa's intention is to show that justification is the result of both intellectual virtue (the likelihood of a justified belief to be true, thus aptness) and perspective (awareness of the conditions under which the intellectual virtue is working, including the awareness of permissibility of the belief, reliability of the process, and coherence among them).

However, under any interpretation of Sosa, the evil demon victim's beliefs fail to satisfy the truth-conducivity requirement. That is, they are not reliably acquired in the world where the victim actually is, thus they are inapt in Sosa's terminology. Then, they are unjustified. If we follow the interpretation (A), the inaptness is due to the unreliable pseudo-perception of the victim. Or, in (B), beliefs could be inapt due to a non-virtuous
faculty although its meta-faculty portion is successful in securing coherence. In other words, in (B), the relevant belief-forming faculty including meta-faculty of coherently believing, or allowing coherent beliefs, is not truth conducive in the demon world. So, strictly speaking, the evil demon victim’s beliefs are not justified as long as the victim is in the demon world. However, if this were what Sosa is actually claiming about the demon victim’s justification, then there wouldn’t be any significant difference between Sosa’s position and what generic reliabilism has to say about this case. Then, the difficulty that the demon objection causes on reliabilism would be found in Sosa’s account too.

Sosa, or at least his wording, is very unclear on this issue. For example, in one place, he talks as if the demon victim’s beliefs are justified.

Both the beliefs of S [in actual world] and those of T [in demon world] are justified (and reflectively apt), since they both fit equally coherently and comprehensively into their respective experience-plus-belief frameworks. ... [However] T depends on some falsehood for justification of many particular beliefs therefore [sic] blocked from constituting knowledge. (Compare here the Gettier problem.)

However, this statement is puzzling. Sosa’s concept of aptness is about the truth-conducivity in a given world. So, in the demon world, however well the beliefs coherently fit in the “experience-plus-belief frameworks,” they cannot be apt as long as they are not coming out of the tendency to be true in the demon world. And if beliefs are not apt, they cannot be justified. Recall Sosa’s claim that justification is one way for

54) Sosa seems to think that, in a normal world, the coherence-seeking reason is truth conducive.
55) KIP, p. 281 (“Intellectual Virtue in Perspective”)
beliefs to be apt. In the above quote, Sosa sounds as if he forgot all about this.\textsuperscript{56} The following quote is equally puzzling for the same reason.

Justification is therefore by our view \textit{internal}—and not even an evil demon could deprive us of it merely by tampering with our external context. For justification is defined as the correct "application" (by our logic, naturally) of our deepest intellectual procedures (which makes justification relativist and indexical). [Sosa's emphasis.]\textsuperscript{57}

Again, my problem is that, if justification is one way for a belief to be apt, then it is unfair to describe it as "internal." At best, it is partially internal because it has to be apt, and the aptness is an environment-dependent and external notion. And if Sosa's claim that justification implies aptness is correct, then, contrary to Sosa, justification would be exactly a kind of thing that the evil demon \textit{can} take away from us by tampering with the external context (that is, by making our beliefs inapt).

Then, should we discredit his claim that justification implies aptness on the basis of inconsistency we finds here? Well, that is hard to determine as well because what Sosa takes as his position eventually would be the following.

\textit{On the present proposal, aptness is relative to environment.} Relative to our actual environment A, our automatic experience-belief mechanisms count as virtues that yield much truth and justification and aptness. Of course, relative to the demonic environment D, such mechanisms are not virtuous and yield neither truth nor aptness. It follows that relative to D the demon's victims are not apt, and yet relative to A their beliefs are apt.\textsuperscript{58} [Sosa's emphasis.]

\textsuperscript{56} The other claim that we can find in this quote (the falsehood involved in justification prevents the justified belief from being knowledge) is no less problematic, or at least unsupported in Sosa's theoretical framework, although I will not pursue this any further here.

\textsuperscript{57} \textit{KIP}, p. 291 ("Intellectual Virtue in Perspective")

\textsuperscript{58} \textit{KIP}, p. 289 ("Intellectual Virtue in Perspective")
And

Relative to the demon's D, therefore, the victim's belief may be inapt and even unjustified—if "justification" is essentially honorific—or, if "justified" simply because coherent, then, relative to D, that justification may yet have little or no cognitive worth. Even so, relative to our environment A, the beliefs of the demon's victim may still be both apt and valuably justified through their inner coherence.59

Here, Sosa attempts to show that his previous idea that the demon victim’s “application of his deepest intellectual procedures” seems to be correct (thus the victim could be justified) is compatible with the opposite fact that the demon victim’s beliefs are inapt in the demon world (thus unjustified). His solution is based on the idea that the demon victim’s beliefs would be apt due to his correct application of his intellectual procedures if he were in a normal world.

Given that aptness is relative to the world in which a belief is formed and held, the notion of justification, which implies aptness, is a relative notion, too. According to Sosa, the (half of the) reason why we want to say a normal epistemic agent’s normal perceptual belief is justified is that, in a normal world, such a belief is truth conducive.60 So, what we really mean is that, given that the perspective condition is satisfied, a normal epistemic agent’s normal perceptual belief is justified relative to the normal world. On the contrary, Edmund’s pseudo-perceptual beliefs given his belief-forming mechanism (including his coherence-seeking reason as a part of it) are not truth conducive in the demon world, thus the beliefs are not justified relative to the demon world. However,

59) KIP, p. 290 (“Intellectual Virtue in Perspective”)
60) The other half of the reason is that such a belief fits to his perspective.
most importantly, when we consider all the intellectual characteristics of Edmund including his non-defective belief-forming mechanism (especially, the part from his experiences to beliefs), Sosa’s idea is that Edmund is justified relative to the normal world even if he happens to be in the demon world. Sosa believes that this verdict fits to our “surface intuitions.” According to Sosa, our intuition about demon victims is that “they lack knowledge but that internally they are blameless and, indeed, virtuous.”

Despite some misleading statements, I take this as Sosa’s final position on this issue.

4.3. A Problem of Sosa’s Position on the New Evil Demon Objection

The charge against reliabilism is that, according to our intuition, Edmund is justified in believing that p. due to some unspecified internal features of him, regardless of where he is, and regardless of whether or not he is a reliable truth gatherer in a world where he happens to be placed. However, Sosa’s idea is that, if conditions for the internal elements (perspective and coherence) are all met, justification then depends on reliability, and reliability is assessed relative to the environment or the world. Sosa seems to have a way to admit our intuition (that Edmund is justified) on one hand and to maintain his claim that reliability is necessary for justification (if a belief is not reliable in a given world, it’s not justified in that world). Sosa would say that, since justification is such a relative notion that is dependent on the world, we cannot really say that Edmund is justified regardless of where he is just because he satisfies the perspective and coherence requirement. Rather, we should say that Edmund is justified relative to a certain world.

61) KIP. p. 289 (“Intellectual Virtue in Perspective”)
And, in this case, Edmund is justified relative to a normal world because that's the world where Edmund's beliefs with perspectives are also apt. That's Sosa's explanation for why we intuitively think that Edmund is justified. What distinguishes Sosa from other reliabilists at this point is that, for Sosa, our intuition that Edmund is justified is not a confusion or mistake.\textsuperscript{62} That is, Sosa's claim is not that Edmund's beliefs \textit{would} be justified if he \textit{were} in a normal world even though he is \textit{actually not} justified in the demon world. It is rather that Edmund’s beliefs \textit{are} justified in the demon world, and that's a justification relative to the normal world (let's call this N-justification). In fact, if Sosa is correct, all the typical perceptual beliefs of both Edmund and any normal epistemic agent are N-justified regardless of where they actually are (again, given that they maintain their internal coherence and perspectives). No wonder that even an evil demon can't take away N-justification from Edmund’s beliefs. Of course, an evil demon can and does take away the truth-conducivity of N-justification by changing Edmund's world from a normal world to a demon world. However, in this case, what is affected by this environmental change is only the justification relative to the demon world, namely, D-justification. That is, the cognitive features of Edmund which satisfy Sosa’s perspective requirement are not truth conducive in demon world. Since N-justified beliefs are not apt in the demon world, it is not really the kind of justification that Sosa wants for reflective knowledge. Thus, an N-justified true belief in the demon world wouldn't be a case of knowledge. What we need in the demon world would be a D-justified true belief. In this way, Sosa also can say that his account accommodates the other strong intuition of

\textsuperscript{62}) For example, Frederick Schmitt does not agree that Edmund is justified, and he thinks that our intuition

163
ours that Edmund does not have knowledge even if his belief happens to be true and he might even be justified intuitively. And this would be considered as another good thing of his position.

However, what exactly would be the demon “world” and the normal “world”? As I have considered at the beginning of this section, the demon world probably isn’t the world in a literal sense of physical space. That is, “the world Edmund actually is in while he is being deceived” cannot be the right description of the demon world that Sosa has in mind. Rather, the fact that Edmund is disconnected from the physical world should be the crucial part of the description of the demon world. Then, the demon world has to refer to, not the physical world itself, but a certain circumstance that can be applied to any physical world, namely the demon circumstance. As long as the agent is in the demon circumstance, he is considered to be in the demon world regardless of where he actually is. And in the same way, the normal world should mean the circumstance where the agent is properly connected to the physical world.

However, this would bring some twist in our understanding of Sosa so far concerning the relativity of justification. First, let’s consider a straightforward case where justification may well be relative to the world as Sosa thinks. Suppose that, in a world E, elephants are invisible but emit a certain frequency of ultra sound. People in this world are able to detect this ultra sound so that they reliably believe that the invisible elephants are near. However, suppose that one citizen in E, Jane, is zapped by a beam and transferred to another world, namely the Twin E (TE), unbeknownst to her while she is misled for that matter.
sleeping. In TE, everything is exactly the same as E except that no such invisible elephants exist. Although the particular frequency of ultra sound does exist in TE, and although people in TE can detect it just like people in E, it has nothing to do with the elephant at all there. Now, if Jane picks up the familiar pitches of ultra sound in the TE and believes that an invisible elephant is near, then this belief is false. However, we can see how it could be justified in Sosa’s sense. That is, this belief is not apt in the TE, thus not justified in TE. However, if she were in E, with the same perspective and cognitive features that she is showing in TE, she would most certainly have true beliefs. Thus, in E, her beliefs are apt as well as justified. So, it is possible to say that even if she is in TE, she is justified relative to E. (She has E-justification in TE.)

Presumably, what Sosa has in mind would be a situation like this. It surely bears some similarity to the NED example. However, the fundamental difference between this example and the NED example is that, in this example, the agent is still connected to the physical world both in E and in TE. So, it is relatively easy to figure out what would happen to this agent throughout different worlds. We just need to imagine the agent in TE who acts as if she is in E. However, this way of stipulating worlds becomes tricky in the NED example. Given that whether the agent is connected or disconnected to the world is the crucial difference between the normal world and the demon world, changing one’s physical environments wouldn’t be all we have to do. In other words, observing what would happen to Edmund if he were in a normal world cannot be as simple as the case of Jane in E and TE. As Greco attempts to criticize, it is surely possible that Edmund would not be a reliable truth gatherer in a normal world even if he shows all the features to be so when he is in the demon circumstance. Greco’s point is that, even if we stipulate the case
in such a way that Edmund becomes unreliable in a normal world, we would nevertheless maintain our intuition that he is justified in the demon world. For example, suppose that Edmund, the brain, has been a bad reasoner until he is victimized by the demon. Something in the process of the demon's deception makes him reason well. Then, we can say that even in a normal circumstance such that Edmund is somebody's brain properly connected to a normal world through real perception, Edmund wouldn't be reliable due to his bad reasoning. Despite this, if we still acknowledge Edmund's justification in the demon world, then it would be possible to argue that the justification Edmund has in the demon world has nothing to do with what would be the case if he were in a normal world.63

The question here might be related to what constitutes the epistemic agent “Edmund” throughout different worlds, or what kind of change is allowed and what should be preserved. As a reply to Greco, Sosa may say that we should construct a normal world that contains Edmund in accordance with the cognitive characteristics that Edmund is showing in the demon world. Given this, I don't know whether this problem would be insurmountable for Sosa. However, the other thing we need to note is a kind of asymmetry we can find in this straightforward case of changing environments. Suppose this time that a citizen in TE, Jack, has a radical fantasy about invisible elephants. (Recall that there's no invisible elephant in TE.) He believes that this kind of animal exists. And he also believes that he can detect it by perceiving a certain frequency of ultra sound. Due to this fantasy, he believes that an invisible elephant is near whenever he perceives the

particular pitches of ultra sound in TE. All of those beliefs are false because there is no invisible elephant in TE. Moreover, I don’t think those beliefs are justified. For reliabilists, this fact does not require much explanation. The method that by which he acquires the belief is unreliable. However, Sosa has to say more. If what he says about the NED case and Edmund are correct, for the parity of the reasoning, he should say the following. Although Jack is not justified in TE, with all the cognitive characteristics he is showing in TE, such as perspective and coherence, he would be justified in E.

However, it is my point that this does not seem to be right. Suppose that Jack is transferred to E unbeknownst him. He maintains his incorrect perspective on invisible elephants, and continues to believe accordingly. However, now, he is getting truth because, in E, there indeed are invisible elephants emitting ultra sound. His faculty of detecting invisible elephants becomes reliable due to the world change, and he is now considered as having a correct perspective. But are those beliefs really justified? Do they qualify as reflective knowledge? If Jack’s belief that there is an invisible elephant near would not be justified in E, then we cannot accept Sosa’s possible position that this belief is justified relatively to E in TE.

In my opinion, Jack’s belief as it is described above is not justified in E although it is reliable there. Having a perspective on this belief would not help because the perspective he has in E is not the result of his being properly connected to the world in E. Jack is not really “aware of” the reliability of his belief-forming mechanism in E because his perspective is merely what is carried over from TE. Obviously, in TE, the perspective
Jack used to have on the basis of radical fantasy does not contribute to improving his epistemic situation. So, it does not help Jack’s epistemic situation in E either. Of course, if Jack somehow acquires a “new” and an appropriate perspective in E after he moves into E, then this new perspective would be exactly the same as the one he used to have in TE on the basis of fantasy. Then, and only then, would he be justified. The alleged asymmetry (or symmetry, depending on where you focus) that I find here is the following. If Jane in the previous example is justified in E with some internal cognitive features X, which are appropriate ones to have in E for justification, she would not lose this justification (justification relative to E) as long as she maintains X regardless of where she is relocated. So, she is E-justified in everywhere else including TE. However, when Jack is unjustified in TE with some internal cognitive features X, he remains to be unjustified in E even if the reliability is restored and even if the X becomes appropriate internal features to have in E for justification. So, concerning Jack’s belief in TE, we cannot say that it is E-justified. Thus, here is a problem of Sosa’s “justification relative to world” business: with Sosa’s explanation for why Jane has E-justification, we cannot explain why Jack does not have E-justification.

Can Sosa once again point out that the way in which I stipulate the world E for Jack is not correct as he might do concerning Greco’s criticism? In the case of Greco, the internal cognitive features that constitute Edmund in the demon world are not fully transferred to a normal world. So, Sosa may have some complaint. However, in the case of Jack, all the internal features that Jack is showing in TE are transferred to E. Sosa should not have a problem at least there. Rather, he should complain about some parts that shouldn’t be transferred from TE to E. Namely, the fact that Jack’s perspective is
acquired by his fantasy in TE are transferred to E. And this might be the source that makes us say that Jack is not justified even in E. So, in this case, "Jack has justification relative to E" means more than "Jack would be justified in E." Rather, it means more like, "Jack would be justified if he were Jane, or any normal citizen in E, who has the same perspective as Jack but acquires it in E through the real connection with things including invisible elephants." At this point, it looks as though only the perspective causally acquired from (or causally based on) how the world actually is plays a meaningful role as a perspective. That is, even if a perspective that I have is well "representing" how things are around me (just like Jack's perspective in E), I am not considered as having the necessary perspective for justification unless I acquired it from how things in fact are.

This reply might be tenable. However, note that if Sosa has to distinguish proper perspectives from improper ones in this way, what kind of perspective we should have for knowledge and justification is determined purely by external factors. Recall that Sosa's perspective requirement is asked for because we need to sort out the cases where a belief is accidentally apt. Now, it turns out that only a proper perspective can do this. That is, the agent has to have access to, for example, the very reason for why his perception is reliable. And whether a given perspective is the very proper perspective or not is determined by external considerations. Then, here may come again the possibility that the agent comes up with the proper perspective by chance. In this case, meeting Sosa's perspectival requirement for justification does not really matter for the positive epistemic status of the given belief including justification and reflective knowledge. The proper connection between the world and the agent deserves attention as being the key factor for our epistemic appraisal. This proper connection is not well emphasized in Sosa's account,
and this is a suspicion that remains behind Sosa’s “relativized world” approach to the
demon victim’s justification.

5. Conclusive Summary

Let me summarize the discussion of this chapter. I began this chapter with the
issue of how Sosa’s account should be interpreted. I discussed Sosa’s distinction of
animal and reflective knowledge in details as well as the distinction of aptness and
justification. I found that these distinctions are questionable and difficult to maintain
without creating a tension within the theory. That is, if Sosa’s intention is to combine the
merits of internalism and externalism, it is doubtful whether his account achieves this
goal. In order to see whether Sosa’s account is at least a better version of reliabilism than
other versions of generic reliabilism, which is one of Sosa’s claim, I discussed the result
of applying Sosa’s account to two major objections to reliabilism, namely the
clairvoyance objection by Bonjour and the New Evil Demon objection by Cohen.

In clairvoyance cases, I focused on Samantha, who has a true belief out of reliable
clairvoyance but also has counter evidence to what she believes. I first pointed out that,
according to one interpretation of Sosa’s account, Samantha’s unjustified belief should be
considered as animal knowledge, which is hard to accept. Then, I raised a general
question of how Sosa’s perspective requirement for justification should be met. I found a
theoretical difficulty of Sosa’s account here. On the one hand, Sosa has to allow
incomplete perspectival grasps to be the fulfillment of the requirement because a
complete perspectival grasp is practically impossible to obtain. However, on the other
hand, an incomplete perspectival grasp conflicts with Sosa’s motivation to opt for this
requirement, namely preventing true beliefs out of an accidentally reliable mechanism (that is, accidentally apt beliefs) from being knowledge. I also considered an alternative interpretation of Sosa’s account using the idea of a meta-faculty, namely coherence-seeking reason. I argued that this interpretation also generates a counterintuitive result. As long as the general tendency of the meta-faculty is what matters for justification, Samantha can be considered as having justification and reflective knowledge, which is unacceptable.

I also criticized Sosa’s relativized world approach to the New Evil Demon problem using the counterexample of Jane and Jack in two different worlds. In this counterexample, Jack is not justified relative to TE, the world where all of his invisible elephant beliefs are false. According to Sosa’s account, however, Jack has a justification in TE, namely, justification relative to E. However, the problem is that Jack, who allegedly has justification relative to E, wouldn’t be justified in E even with the exact same characteristics he has in TE. Here, I found a possibility that both reliability (aptness or virtue) and internal perspective requirement are met but justification and knowledge are not acquired. My diagnosis is that if a perspective is shaped without proper connections with the world, its presence does not prevent accidentally apt beliefs from being knowledge. I will accommodate the lesson from this case in my own account, which will be presented in the next chapter.
CHAPTER 4

EPSTEMIC GOODNESS EXPLAINED

1. Two Phases of a Belief and Epistemic Goodness in Each Phase

1.1. Two Phases of a Belief

In chapter one, I defined epistemic goodness as a property that is necessary for knowledge, i.e., as what makes a true belief knowledge when a proper Gettier solution is added. In chapter two and three, I have argued that neither Sosa’s intellectual virtue nor Plantinga’s warrant is quite up to its billing in terms of analyzing epistemic goodness. In this final chapter, I will attempt to give my own proposal of how this property could be better explained.

What distinguishes my proposal from others is that I consider a belief as having more than one aspect to be evaluated epistemically. That is, although a propositional belief that p [Bp] may refer to a single mental phenomenon metaphysically, it can be seen in two different points of view in terms of epistemic relevancy, and thus should be evaluated in accordance with them. First, Bp can be seen as a result of certain belief-
forming and belief-holding processes, mechanisms or faculties. Second, at the same time, Bp also can be seen as what leads the agent to a certain consequence, namely an event of genuine agreement on p or an event of thinking that p is the case. Regardless of the real nature of what a belief is (whether it is a tendency or an occurent mental event), these are two important characteristics of any given belief. The former is a description of Bp when it is seen as a causal consequence of some of its antecedent conditions, including its belief-forming mechanism. The latter is a description of Bp when it is seen as a causal precursor to a different type of mental activity such as S’s assent to p. S’s assent to a certain proposition is a mental event accessible by S, which can be distinguished from a belief itself. We may call this “belief manifestation.” Belief manifestation itself is not a necessary part of having a belief. Belief manifestation or a genuine assent to p occurs only when some appropriate circumstances obtain. Even so, the idea of how a belief would be manifested in those appropriate circumstances is an essential part of understanding the characteristics of Bp. If S has Bp, S is either in the state of manifesting it (agreeing on p at the given moment) or in the state of having a tendency to manifest it in a particular way, namely in an agreeing-on-p-way. We may think of a belief-forming mechanism as if it is a tendency. That is, a belief-forming mechanism could be understood as a tendency to produce a particular belief depending on the input (and that’s the causal role of a belief-forming mechanism in the actual production of a belief). An individual belief could be understood similarly. That is, except for the moment that a tendency is manifesting, which is an occurent sense of a belief, Bp itself can be understood as a tendency to produce an assent to p under some appropriate
circumstances. In this picture, Bp itself is an antecedent condition, which causally contributes to produce an assent to p (just like a belief-forming mechanism is an antecedent condition that causally contributes to produce Bp), and the assent to p under a certain circumstance is considered the consequence of Bp.

When I describe the nature of the contribution of Bp to the production of an assent to p as causal, the term “causal” used here is a less than straightforward sense of the term. A belief is a causal contributor to the event of assenting just like the tendency of fragility is “causally” contributing to the event of a cup’s being broken, or the dryness is “causally” contributing to a wild fire. If there hadn’t been Bp, there wouldn’t have been a manifestation of Bp, i.e., a genuine assent to p, even if a direct cause had been provided (such as facing a direct question of whether p is the case, or any other stimulation on S). The same thing can be said when we consider belief as a consequence of certain belief-forming mechanisms. The strict sense of the cause of Bp would be, for example, perceptual stimuli, but the mechanism of perception is causally contributing to the outcome.

Now, I would like to call the aspect of a belief as a consequence the C-phase, and the aspect of a belief as an antecedent the A-phase.

[C-phase of a belief]: A belief is a consequence of what is responsible for the existence of it such as belief-forming and holding mechanism.

[A-phase of a belief]: A belief is an antecedent for the manifestation of an agent’s genuine assent to the proposition that is believed.
Epistemological evaluative concepts such as justification, warrant and epistemic goodness primarily apply to beliefs. Now, given the distinction of C-phase and A-phase, any belief can be picked out by focusing on its C-phase and by focusing on its A-phase at the same time, just like a coin can be described in terms of the characteristics in its tail and the characteristics in its head. In fact, any given belief should be described using the characteristics of both, if it is to have a thorough epistemic description. Thus, my idea is to approach the epistemic evaluation of a belief on these two separate grounds.

1.2. Epistemic Goodness in C-Phase

If we consider a belief as a terminal product of a certain mechanism, it is natural to think that the quality of the belief would depend on the quality of the mechanism. Similarly, the immaculate state of the assembly line can be a reason why a given product should be considered highly in terms of its quality. Therefore, I think that this should be the way we measure the epistemic goodness concerning C-phase. That is, in order to assess the epistemic qualification of a belief when it is considered as a consequence of belief-forming mechanism, we can go to the antecedent conditions that are responsible for the existence of the belief, and evaluate them. We may say that this is a backward looking consideration concerning the legitimacy of having a certain belief.

Due to recent developments in epistemology such as the causal theories of knowledge, the process reliabilism and the faculty reliabilism, the view to find an epistemically significant aspect of a belief in its causal relationship with its antecedent
conditions (such as its being a product of a belief-forming mechanism or faculty) is a part of a strong tradition. Following Goldman and Plantinga, we can expect that this mechanism would determine the epistemic quality of the belief. In fact, the history of a belief is very influential for our intuitive epistemic appraisal of any given belief. We sort beliefs as perceptual beliefs, inferential beliefs, beliefs formed by mathematical reasoning, beliefs formed by wishful thinking, and so on. Namely, describing a belief as “being caused by perception” is a legitimate way of identifying and categorizing it. What then would be a criterion for evaluating the goodness of those belief-forming mechanisms? For this, we should look for epistemically desirable features of Bp. One of those desirable features is the truth of the proposition that is believed. Moreover, what we need for knowledge is not merely a true belief but a “non-accidental” true belief. Then, what I mean by the evaluation of a belief concerning C-phase is:

[Evaluation of an epistemic goodness of Bp concerning C-phase]: Evaluation of the antecedent conditions (belief-forming mechanism) of Bp in terms of a desirable feature of its consequence (Bp) such as Bp’s being a non-accidental true belief.

In fact, the most fundamental requirement for the property that makes a true belief knowledge is that believing truth should not be “by chance.” A true belief resulting from guesswork or wish-fulfillment are the typical examples of true beliefs by chance. We can

1) The phrase “by chance” represents a group of words such as luck, fluke, coincidence and accident.
think that an epistemic agent ends up with having a true belief by chance by using a causal mechanism that has tendency to produce false beliefs. That is, if a certain belief-forming mechanism is not reliable in producing true beliefs, such as guesswork or wishful thinking, then even if the outcome belief is true, it would be only a matter of a luck. This would be one way to explain accidentality involved in a belief out of guesswork, and it focuses on the C-phase of a belief.

Let’s then consider the ability of producing true beliefs as a criterion for measuring the desirable quality of a belief-forming faculty. If so, as it is acknowledged by the process reliabilists and faculty reliabilists, a reliable belief-forming mechanism provides epistemic goodness to the outcome belief. That is, the epistemic goodness is conferred to Bp by its belief-forming mechanism if a true Bp is an outcome of a mechanism that has a tendency to produce Bp more often when p is true than when p is false. This is the basic idea of how to measure the epistemic goodness of a belief in the C-phase. I would like to call this a “properly acquired belief for knowledge.” Of course, this is an incomplete idea. We would need further considerations especially concerning the individuation of the mechanism and the relevant circumstance and the environment in which a belief is properly acquired. We may think of a special circumstance that a proper acquisition of a belief is not contributing enough to secure the ideal of having “non-accidental” true belief. I will consider some “defeater” condition later. Thus, at this point, let’s define a prima facie epistemic goodness in C-phase as the following.

177

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
[A *prima facie* epistemic goodness in C-phase = A properly acquired belief]: Bp is a properly acquired belief iff the conditional probability of p being true in the actual world given that Bp is produced by a belief-forming mechanism M is high.

This idea is structurally similar to the way in which a reliable indicator theory understands epistemic justification. However, note the difference that, in my proposal, the high conditional probability is based on the reliability of the belief-forming mechanism M.

1.3. Epistemic Goodness in A-Phase

1.3.1. What to Evaluate in A-Phase

In the A-phase, a belief itself is considered as an antecedent condition, as if a tendency, and its consequence is considered as a genuine assent to the proposition that is believed. Such an assent is a mental event that typically includes linguistic behaviors.

The consequent of Bp can be characterized as manifesting an assent to p (not to q or to r). Describing Bp as "what causes S’s assent to p (rather than q or r)" seems to be equally legitimate as the description of "what is caused by perception," for example. That is, my belief that there is a computer screen in front of me is what is produced by my perception

---

2) In order to fully develop this proposal, a technical distinction between belief-dependent and belief-independent mechanisms concerning M would need to be introduced. Goldman makes a similar suggestion in his *Epistemology and Cognition* (p. 83).

3) Whether assent is equivalent to some kind of linguistic behavior, such as responding affirmatively to the content of a proposition p, is unclear. Although I want to leave a room for such a thing as non-linguistic affirmation as a kind of assent, it doesn’t seem to be crucial at this point whether my notion of assent amounts to purely linguistic behavior. What matters more for my project is that assent is a conscious event, where S is aware of her affirming as well as is grasping the propositional content of the proposition.
presumably, and what makes me genuinely assent to the proposition that there is a computer screen in front of me. So, in the A-phase, a belief itself, that is, what causes S’s assent to the proposition itself, is evaluated. Here, I think that what should be the center of the focus is the propositional content that is believed.

A belief can be acquired and discarded without being noticed by the agent. Moreover, acquiring and holding a belief is not a matter of the agent’s volition or control. A strong version of doxastic voluntarism, which is a doctrine that, at the bottom line, we can freely choose what to believe or disbelieve, is difficult to be maintain. Nevertheless, manifestation of a belief is a conscious event. The agent is required to be aware of the fact that she is taking an affirmative attitude toward a proposition that she grasps. And a certain degree of ability to assess the things going on in one’s consciousness seems to be a reasonable request upon a rational agent. We assume this as an important feature of an epistemic agent. In this sense, the events that are going on in the agent’s consciousness may define what kind of epistemic agent she is, and this adds another dimension that is epistemically relevant for evaluating a belief. That is, to qualify as knowledge a belief should not turn out to be problematic in this dimension.

Suppose an agent manifests Bp and B(not-p) at the same time. This amounts to the event of assenting to p and assenting to not p at the same time. Furthermore, suppose that this agent also manifests a belief that it cannot be both p and not-p at the same time. It might be that some or even all of those beliefs have legitimate causal histories. Even

---

so, we detect an epistemic problem here in a different dimension. We may think that this
type of epistemic agent is problematic or we may think that the relationship between
those propositions that are believed is problematic. Whatever that might be, it is certain
that it makes us, the evaluators, reluctant to assign a high epistemic status to any of those
beliefs. That is, we think that neither Bp nor B(not-p) qualifies as knowledge in this
circumstance. I do not want to deny the rationality of an agent just because she is holding
Bp and B(not-p) at the same time in a dispositional sense. That is, causally holding Bp,
B(not-p) and B(either p or not-p but not both) at the same time might be compatible with
the fact that the epistemic agent is rational enough to have knowledge. What seems to be
incompatible with being rational, however, is the occurent event of manifesting all three
beliefs at the same time. If an agent is manifesting all three beliefs at the same time, it is
doubtful whether we can confer knowledge to any one of those manifested three beliefs.
Maybe it could even be a reason to discredit the agent as a whole, and therefore a reason
to deny the knowledge status of many other beliefs of the agent. Now, if this is true about
belief manifestation, then having Bp and B(not-p) even in a dispositional sense cannot
avoid the negative epistemic evaluation. It is because Bp and B(not-p) are antecedent
conditions for the manifestation of Bp and B(not-p), i.e., the event of genuine assent to p
and not-p. In short, we negatively evaluate these beliefs because they lead to an event that
is not supposed to happen to a qualified rational agent.

5) For this, we would have to assume that the agent manifests those beliefs (i.e., brings them into
consciousness) one at a time. For example, Bp is manifested with Bq and Br but never with B(not-p), and
B(not-p) is manifested with Bs and Bt but not with Bp, and so on.
Even in many less extreme cases, having Bp without any other manifestable state to support Bp could hurt the epistemic status of Bp because Bp leads to an event where the agent genuinely assents to p with no other manifestable state to support p. Recall the case of Mr. Truetemp mentioned in chapter one. When an appropriate circumstance obtains, it is possible that Mr. Truetemp is manifesting a belief that the current temperature is 62 in his consciousness together with the recognition that there is no reason to be so. Furthermore, if there is an epistemic agent who shows this type of manifestation rather normally (i.e., manifestation of contradictory propositions or manifestation of propositions that have no support from other manifestable states), then we might even want to discredit the epistemic status of all or many other beliefs of such an agent.

What I can see here is the following. The decision of whether an epistemic status of a given belief should be discredited or not can be made depending on other beliefs that the agent has. Whatever the exact nature of the epistemic value that a belief creates within the relationship with other beliefs, it is certain that our epistemic evaluation takes notice of this value. That is, the epistemic quality of Bp depends on a certain epistemic value that the proposition p has when it is put within the background of other propositions that the agent also believes. Eventually, I think that the candidate of such a value is the “subjective correctness” of the belief. However, before I go into the details of

---

6) Perceptual beliefs of a normal epistemic agent are not good examples of what I have in mind as beliefs without any manifestable states that support those beliefs. I will soon introduce the concept of “cognitive system” and “potential belief,” whose members include propositions that the agent has a disposition to believe, as well as propositions that the agent already dispositionally believes. A proposition that the agent has a disposition to believe would qualify as a propositional content of a manifestable state that supports Bp. If so, the perceptual beliefs of a normal epistemic agent would be considered as having those manifestable states that support those beliefs.
the notion of correctness, let's consider the general idea that a believed proposition has a
certain epistemic value relative to other beliefs that the agent has.

Earlier, I used an analogy of an assembly line and its product as a justification for
the required quality of the product. To continue this analogy, let's suppose that the result
of the above mentioned immaculate assembly line is an LP record rather than a CD, or a
wooden cart wheel rather than a rubber tire. Regardless of the justification we may have
concerning how this LP record or a wooden cart wheel is well-made without a defect,
another kind of justification from a different direction seems to be required. Since the LP
record is seen as one of many commodities in contemporary society and culture, all of
which have their own roles and bear various relationships with others, we can ask for a
justification of the LP record making with this regard. For example, if an LP record
player no longer exists, the practical value of the LP record that emerges from the
relationship with other commodities should be reconsidered, and maybe "unjustified"
relative to the elements in current society and culture, even if it is a very well made one.
The LP record just doesn't fit in there.

Back to beliefs, when Bp is acquired, it means that, from that moment on (and
until Bp is discarded) the agent has a tendency to agree on p. In this way, Bp introduces p
in the agent's cognitive repertoire. Let me introduce the term "cognitive system" in the
following way.

[A cognitive system of S at t]: A cognitive system of S at t is the sum of the
propositions that S actually believes at t both occurrently and dispositionally (thus,
propositions that S assents to) and propositions that are “potentially” S’s beliefs at t (thus, propositions that S would assent to).

[A potential Bp]: A proposition p is S’s “potential” Bp at t iff (i) S does not have Bp at t but (ii) S has a disposition to believe that p because S possesses evidence for p, and the existence of such a disposition is not counteracted by other beliefs that S has.

For example, consider a normal epistemic agent who believes that he is seeing red. Suppose he does have a red experience. Even so, it is possible that S does not believe the proposition describing S’s perceptual experience such as “I seem to see red.” However, this proposition, “I seem to see red” can be a potential belief of S, thus a member of S’s cognitive system. It is because, although S in fact does not have this belief, S has a disposition to believe the proposition that describe his red experience. However, suppose that S is an epistemic agent in Plantinga’s Red Bus example. Based on false but legitimately authoritative testimony, a person in this example is convinced by the idea that there is no such thing like color red in this world. He believes that everything that appears to be red is in fact not red. He trained himself so that when something appears to be red, he could have an almost automatic belief that this is not red. One day, on the street of London, he was almost hit by a bus. When the big red bus was fast approaching toward him, he momentarily acquired a belief that the bus is red despite his training because this experience was so overwhelming. Soon after the incident, however, he corrected himself that the bus was not red. If the case is like this, there are certain beliefs.
that we cannot attribute him as his "potential" beliefs despite his disposition to believe. For example, "I see red" cannot be the member of this person's cognitive system even though he has a disposition to believe this proposition. The problem is that he would be consciously denying the ways in which this disposition works because of other beliefs that he has.

For the same reason, considering the conclusion of the epistemic closure principle [if S has Bp and S also has a belief that if p then q, then S believes that q] as a "potential belief" of S and as member of S's cognitive system seems to be acceptable. The epistemic closure principle may be false in terms of the relation between actual beliefs. However, in many cases, attributing such Bq to S on the basis of S's disposition to believe might be reasonable. For example, if S believes that x is red, and also believes that a red thing is a colored thing, then the belief that x is a colored thing would be S's "potential belief" as long as S has a disposition to believe this.

The reason why the propositions that are merely "potentially" S's beliefs should be in S's cognitive system is this. If S's cognitive system is only to represent the propositions that S actually believes, it would be too narrow for the relevant background where the manifestation of Bp takes place. For our purpose of epistemic evaluation, Bp should be evaluated in comparison to all the evidence that S possesses, not just to all beliefs that S has. Roughly, the idea of a cognitive system is to make it represent the total evidence that S has for current and future beliefs of his own. Some evidence that S has may not be the propositions that S actually believes although it could easily be one of them, as in the above mentioned case of Bq, the conclusion of an obvious entailment.
Also, some evidence that S has would not be in the form of a proposition (such as perceptual experience). Nevertheless, this non-propositional evidence may as well be the evidence that S has. In spirit, the idea of potential belief is to capture “what S would believe on the basis of easily available evidence that S already has” and to treat the contents of those potential beliefs as a part of the background into which a new belief is put and evaluated.

Using the idea of a cognitive system, now the consequence of Bp (assent to p) can be considered more generally. Namely, as a consequence of Bp, p is introduced in the cognitive system as a new member. And in the parallel way that we understand the belief-forming mechanism (epistemic goodness in C-phase), using the desirable feature of p (especially, it’s a desirable feature rising from its relationship with other members in the cognitive system), we can evaluate the Bp itself, because it is an antecedent condition from which such a desirable feature derives.

[Evaluation of an epistemic goodness of Bp concerning A-phase]: Evaluation of the current state of having Bp in terms of the relationship that p makes with other propositions in the agent’s cognitive system.

If we understand the evaluation of a belief concerning A-phase as focusing on the current state of having the belief, then this stays within a tradition of epistemology too, just like the evaluation of an epistemic goodness of Bp concerning C-phase is in the tradition of the process and faculty reliabilism. In order to see if a belief is knowledge, we
have been normally looking for some epistemically positive features in the current state of the belief, such as whether it is well supported by evidence and other beliefs that the agent has. Coherentism would be a notable example where the features in the current state of a belief essentially determine the epistemic quality of a given belief. My thought is that this is a case where only the A-phase of a belief is emphasized. The state of having an optimal explanation of p using other elements in one’s cognitive system, the state of p enjoying mutual evidential support among the elements in the cognitive system and also the state of having other propositions as the reason for believing p are all possible consequences of Bp when a belief itself is considered as an antecedent of them. The position held by Feldman and Conee, which is called “Evidentialism,” claims that what is essential in epistemic justification is that a belief should fit to the evidence that the agent has. This would also be viewed as focusing on the A-phase of a belief. I also believe that a certain aspect of reliable indicator theory (at least the idea of a belief’s being an indicator of something else) could be understood in a similar way. Finally, what we have seen in Sosa in the previous chapter (whether a belief fits coherently in the perspective that one has) would be the same kind of view. It may be said that these approaches amount to an evaluation of the “static” nature of a belief (rather than the “dynamic” nature as in belief formation), and constitute what might be called a ‘Current Time-Slice’ theory of epistemic goodness. Thus, in spirit, the distinction of C-phase and

A-phase is to capture both the dynamic and static natures of any given belief, and reflect them in our epistemic evaluation.

1.3.2. A “Correct” Belief

Now, let’s consider the criterion to measure the epistemic goodness of a belief in the A-phase. Again, I think that the prevention of accidentality would be the primary goal here, just like the C-phase. However, the kind of accidentality that I think is involved in A-phase is somewhat different from the one involved in the C-phase, i.e., the case where a belief produced by an unreliable mechanism (that only has a low probability to produce true beliefs) turns out to be true. For an analogy, we would agree that it is lucky if I beat one out of 100 odds and win the lottery. It is mainly because one out of 100 is such a low chance to land on. Now, suppose that buying a lottery ticket is the least expected thing that I would do because I believe that all lotteries are rigged. Thus, I believe that I have absolutely no chance to win any lottery, and given that the only purpose for me to participate in a lottery is to win, I believe that I have no reason to buy any lottery ticket. If a person like me buys a lottery ticket out of drunkenness, then there seems to be another kind of accidentality involved, regardless of whether it eventually hits a jackpot. In the first place, I am not supposed to buy a ticket. It is accidental because buying a ticket is not the thing I am expected to do. Some beliefs are accidental in this sense. That is, regardless of whether it turns out to be a true belief or not, having the belief itself is not something the agent is expected to have given the epistemic situation that she is in. For example, as the result of guesswork, a normal epistemic agent usually ends up
believing a proposition that has no evidential support from her cognitive repertoires, regardless of whether the proposition believed is true or false. That is, the epistemic situation that the agent is in does not contain anything that she might consider as the reason for believing the given proposition. This is a case in which a true belief is paired with no accessible reasons or accessible evidential support. Such a belief can be seen as an "incorrect" belief relative to the epistemic situation that she is in. That is, given all the elements in her cognitive repertoires, the agent is not supposed to hold the belief. In this sense, having this belief, and eventually have a true belief, is accidental.

I will consider that the source of the accidentality in such case is that the given belief is "incorrect from the agent's point of view," where the point of view is constituted by the beliefs and evidence that the agent has. Thus, a belief qualifying as knowledge should be at least "correct from the agent's point of view." This, I think, would be another fundamental requirement for knowledge. In many test cases, presence of "incorrectness" in the above sense hinders us from attributing knowledge status to a true belief even though the belief may have other epistemically desirable features such as the epistemic goodness in the C-phase. Consider again the Mr. Truetemp case. Intuitively, his true belief on the current temperature is "incorrect" in the above sense. The same goes for Bonjour's clairvoyance cases. These are the beliefs that the agent is not expected to have relative to the other elements in her cognitive system although there are legitimate causes for the agent to acquire and to hold those beliefs.

However, given that there could be several different ways to capture the idea of "being correct from the agent's point of view," the next thing is to explain how this
notion should be understood. First, one might think that the criterion of correctness is subjective approval through some kind of meta-awareness. For example, one might think that a belief is correct from the agent's point of view when the agent has a meta-belief (awareness or recognition) that it is correct (or reliably acquired). However, demanding the presence of a meta-belief in every correct belief is not plausible. If the propositional content of a meta-belief is required to be correct, then there obviously is a regress problem. If we allow a regress, then, due to the agent's cognitive limits, some beliefs cannot be accompanied by such a meta-belief. However, not all those beliefs are negatively evaluated in terms of their epistemic status. The reason I think we must require correctness is that we think this kind of property is necessary for knowledge. Then, for the notion of correctness that I want to have here, I must reject any combination of purely subjective criteria based on the agent's cognitive access (such as what the agent thinks "correct" from her point of view, which she thinks she has) because these are not really relevant for having knowledge. It is possible that a normal epistemic agent has no idea of her own point of view yet the agent may have knowledge. Thus, the notion of correctness is objective and does not require any meta-awareness.

Even if the notion of correctness is not subjective, it would surely be relative to the agent's other cognitive repertoires because, so far as we can see, this value emerges from a given belief's relationship with others. Now, my suggestion is that correctness should also be understood in terms of reliability. However, it wouldn't be the actual truth conducivity relative to the actual world but rather a "truth" conducivity relative to the agent's subjective world. That is, given that we may be able to construct the subjective
world of the agent, which I would like to identify as a world represented by all the items in the agent's cognitive system, a belief that is a reliable indicator of "truth in that world" is what I would like to call a correct belief. This idea can be expressed in the following way.

[Bp is correct relative to S's cognitive system] = S believes that p and p is correct relative to S's cognitive system.

[A proposition p is correct relative to S's cognitive system] = the conditional probability of p being true given the relevant portion of propositions in S's cognitive system is high.  

This is to capture the idea that a correct proposition p is what is likely to be considered as true in the world where S thinks S is in, where the sum of the relevant portion of propositions in S's cognitive system represents such a subjective world. For example, consider a normal epistemic agent S's perceptual belief that there is a coffee mug on the table. (Call this proposition p.) Suppose that S has a proposition that describes her perceptual mug watching experience in her cognitive system (call this q) either as a propositional content of an actual belief or of a potential belief, and furthermore, there is no proposition that S believes (again, either actual or potential) which may conflict with p or q. Then, the conditional probability of p being true given the relevant portion of

9) Currently, I do not have a detailed account of relevance. However, I have in mind the standard probabilistic notion of relevance, in accordance with which p is relevant q on evidence e if the probability of q on (p & e) is different from the probability of q on e alone.

190
propositions in S's cognitive system, which includes q, would be high. Thus, p would be correct relative to S's cognitive system, and thus Bp would be S's correct belief. However, suppose that S also has a belief that a holographic mug image is frequently tested on this table, which is r, and a belief that a holographic mug image is indistinguishable from the real mug just by looking, which is s. Then, the conditional probability of p being true given q, r and s is not so high. Thus, the proposition p in this circumstance is not correct, and therefore, in case S believes that p, Bp is not correct either.

1.3.3. Intellectual Virtue

Even if a given belief is correct as in the sense that I proposed above, that wouldn't be all that is needed for epistemic goodness in the A-phase of a belief. Incorrect beliefs are "accidental," and thus fail to achieve higher epistemic status in our epistemic evaluation. However, not all correct beliefs are epistemically valuable enough to confer knowledge status. Correctness is an evaluation of a belief in terms of its "current time-slice." Then, it is possible that this requirement overlooks the so-called epistemically questionable characteristics of an agent whose beliefs appear to be correct in its relationship with other beliefs at a given time-slice. For example, consider an epistemic agent S who has propositions q, r and s in her cognitive system. Suppose that a proposition p being true given q, r and s is high, and therefore, the negation of p [not-p] being true given the same conditions is low. Then, p would be a correct proposition for S to believe, and (not-p) would be an incorrect one. However, suppose further that S has no
tendency to prefer to believe $p$, a correct one, rather than ($\neg p$). That is, $S$ is generally indifferent between $Bp$ and $B(\neg p)$, thus sometimes believes $p$ and other times believes ($\neg p$) depending on the belief-forming mechanism that happens to be activated. ($Bp$ and $B(\neg p)$ could be produced by two different mechanisms in terms of their reliable belief-forming tendency, but there is no problem even if we assume that both beliefs are produced by one and the same mechanism.) In this scenario, $Bp$ is correct but it is not certain that $Bp$ is epistemically good. The problem is that, in this case, having a correct $Bp$ itself is accidental given $S$'s characteristics as an epistemic agent as a whole. In short, $S$ is the kind of agent such that the correctness of $Bp$ does not play any role in $S$'s state of having $Bp$. Because of that, even though the state of holding $Bp$ is epistemically desirable (thus correct), the correctness itself could be accidental just in the sense that the agent could have ended up with incorrect beliefs with the equal chance. There must be some requirement to insure that the desirable state of having a correct belief is not accidentally created.

I think that the term “intellectual virtue” or “cognitive virtue,” which has recently been introduced in theories of epistemology, would be a convenient notion for my needs.\(^\text{10}\) Let me consider “intellectual virtue” in the following way.

\[ \text{[Intellectual virtue of } S \text{ for believing } p \text{ under a circumstance } c] = (i) \text{ There is a group of propositions } F \text{ to which } p \text{ belongs. (ii) There is a group of circumstances} \]

C to which c, the particular circumstance where S is in, belongs. (iii) Concerning propositions in F under a circumstance in C, S has the tendency, grounded in S’s inner nature, to only believe correct propositions.

My understanding of intellectual virtue here is different from what is normally suggested by virtue epistemologists like Sosa and Greco. The crucial difference is that, in my understanding, securing truth is not the primary goal of intellectual virtue, while for Greco and Sosa intellectual virtue basically overlaps truth conducivity. It is directly connected with reliability of the faculty working in a given environment. As Greco suggests, it seems right to understand a virtue as an ability and as “a stable disposition to achieve certain results under certain conditions,” as Greco’s analogy of Don Mattingly’s ability to hit the baseball is supposed to show.\(^1\) However, concerning the result that intellectual virtue should bring about, Greco thinks that true belief is the one.

A mechanism M for generating and/or maintaining beliefs is a cognitive virtue if and only if M is an ability to believe true propositions and avoid believing false propositions within a field of propositions F, when one is in a set of circumstance C.\(^2\)

A similar definition is found in Sosa. For him, intellectual virtue is “a competence to distinguish the true from the false in some field of propositions F when in certain circumstances C.”\(^3\)

---


\(^{2}\) Ibid., pp. 414 - 415

If cognitive virtue is understood in this way, virtue epistemology is no doubt a kind of reliabilism. But, the question is whether it is a good thing for virtue epistemology. In both Sosa and Greco, cognitive virtue is measured in terms of its outcome in a certain environment in which the virtue is actually used. In other words, they would call something a cognitive virtue in a given environment only if it is efficient in achieving the favorable outcome, i.e., getting truth and avoiding error, in that environment. Thus, according to their view, perception could be a cognitive virtue in a normal world (because we can get true beliefs most of time as a result of using it), whereas a similar or even an exactly same faculty is not virtuous in a demon world (because it is unable to bring about true beliefs in that world). This allows that a virtue could depend on the world or environment. A cognitive faculty (or a cognitive agent as a whole) which was virtuous in an environment E1 could no longer be virtuous in a different environment E2 even though the faculty is working exactly in the same way in both environments. This could be problematic when we think about an agent who was in the normal world but now transferred to a demon world unbeknownst her. This agent was virtuous but through the night she has become non-virtuous even though there hasn’t been any internal change in the way she maintains her cognitive life.

Maybe this consequence itself isn’t so seriously counter-intuitive for some people. For those who think that a virtue is synonymous with the efficiency of getting a certain goal, this consequence might be unproblematic. However, an intellectual virtue in this sense is not different from a reliable process or faculty approach. Then, I think, the motivation for why we need to introduce the concept of virtue in epistemology becomes
unclear. For one thing, the evil demon victim's cognitive faculty does not have any ability of fetching truth in the given world. If we have to call this faculty a non-virtue because of its inability to get truth, then virtue epistemology seems to inherit the same problem which has threatened generic reliabilism. For another, Norman's clairvoyance has an ability to arrive at truth in the given world. Again, if we have to call this faculty a virtue because of its reliability, the problem of reliabilism is repeated in virtue epistemology. These results are unfortunate because I think that the concept of virtue, if it is properly accommodated in epistemology, has a potential to avoid these two chronic problems of reliabilism. It would be at least theoretically unprofitable to identify cognitive virtue simply as reliable process or faculty.

In my opinion, the goal of intellectual virtue should be something other than truth. If we consider reliability as the primary objective constraint of intellectual virtue, then the notion of intellectual virtue becomes environment-dependent because it presupposes the notion of truth. However, there is something strange about the idea that whether a belief results from intellectual virtue or not depends on the environment. In most cases of ethical virtue, virtues are considered as a so-called environment-neutral inner character of the agent. My action is not judged as being brave or noble on the basis of its actual result in the actual world. If I step in front of a lady who is about to be battered by a guy in the back streets of LA without knowing that they are acting for a film, I do not achieve the objective goal of being brave (or any type of virtue you want to ascribe to such a behavior) in this case. However, my act would still be considered as bravery, thus virtuous, not on the basis of the actual environment (the filming scene) but on the basis of
the environment that I perceived as actual (a damsel in distress). If a certain analogy can be established between moral and intellectual virtue, this fact does not favor the understanding of intellectual virtue in Greco’s or Sosa’s way. In my understanding, intellectual virtue is a tendency primarily aiming for internal correctness, not truth. Just like a reliable belief-forming mechanism is an ability that the agent has even if it is not under control of the agent, the ability that maintains only the beliefs that are correct from one’s point of view also belongs to the agent even if it is not under a voluntary control of the agent.

Now, for the epistemic goodness in A-phase, I think we need both the correctness of a belief and backup of this correctness by intellectual virtue. Let me use a term “properly maintained belief”\(^{14}\) as the term to describe the epistemic goodness in A-phase.

\[ \text{[A prima facie epistemic goodness in A-phase} = \text{A properly maintained belief]} \]
\[
\text{A belief is properly maintained in S at } t \text{ iff (i) the belief is correct relative to S's cognitive system and (ii) S’s has the tendency, grounded in S’s inner nature, to only believe propositions that are correct relative to S’s cognitive system.} \]

Finally, I think that knowledge requires epistemic goodness both in C-phase and in A-phase. However, for knowledge, what is needed is not just \textit{prima facie} epistemic goodness.

\(^{14}\) My use of the term “maintenance” would be rather technical because it doesn’t exactly represent what it normally means. It does not have much to do with a “causal maintenance” or “holding a belief.” When a formed belief is sustained (stayed) in one’s cognitive system due to some causal power, I would use the term “being held” rather than “being maintained” and treat it as a characteristic of the C-phase.

\(^{15}\) I recognize that sometimes, giving up existing beliefs might be considered as an intellectually virtuous thing to do. However, in order to explain this, a so-called theory of rational belief revision would be needed. I cannot sketch such a theory here.
goodness but rather undefeated epistemic goodness. So, what follows will be mainly an explanation on those conditions that defeat prima facie epistemic goodness in each phase.

1.4. Application to Cases

So far, I have discussed two distinguishable aspects of a belief that should be evaluated and about evaluative criteria in each phase. It can be summarized in the following ways. First, believing a true and “correct” proposition is required for knowledge. Furthermore, a certain degree of “non-accidentality” should be added. A true belief is what a belief-forming mechanism concerns, and the required “non-accidentality” can be established by the reliability of a belief-forming mechanism. This is the evaluation of a belief concerning C-phase and this constitutes epistemic goodness of that phase, which can also be called a proper belief acquisition. Next, the requirement of “correctness” is justified by the fact that an “incorrect” belief brings about a kind of “accidentality” that also should be eliminated in order to have knowledge. Then, the possibility that this correct belief is maintained “accidentally” is dealt with by requiring the intellectual virtue of the agent. These considerations are eventually summed up as constituting epistemic goodness of A-phase, or the proper maintenance of a belief. Now, given that these considerations are about prima facie epistemic goodness, I will go on to explain some defeater conditions for them, i.e., the circumstances where such epistemic goodness is overridden, thus cannot amount to knowledge. However, before that, it would be helpful to see how the basic idea of epistemic goodness can be applied in some standard cases.
I believe that any proposed account of knowledge should be able to explain how our
typical cases of true perceptual beliefs qualify as knowledge. According to my proposal, a
normal epistemic agent’s perceptual belief in a normal (non-Gettierized) situation, for
example, my belief that there is a computer screen in front of me, has *prima facie* epistemic
goodness in both C-phase and A-phase of a belief. That is, it is a case where a belief is
acquired properly and maintained properly. If my perception is a reliable belief-forming
mechanism under a normal condition (in a normal world, in a normal environment), it
satisfies the reliability requirement for proper acquisition. Furthermore, this belief is correct.
Either I actually believe the proposition that describes my computer screen watching
experience or it would be reasonable to consider such a proposition as the content of my
potential belief. Thus, the proposition that describes my computer screen watching
experience is in my cognitive system. And my cognitive system does not contain any
particular proposition that indicates what I see is not like what really is. Given that, the
probability of the proposition that there is a computer screen in front of me to be true is high.
So, my belief is a correct one. Finally, there seems to be no reason to doubt that this belief is
out of intellectual virtue. As a normal epistemic agent, I do show a disposition to maintain
those correct beliefs. Then, my belief satisfies the intellectual virtue requirement as well.
Then, my perceptual belief that there is a computer screen in front of me has all the
necessary epistemic goodness (both in C-phase and A-phase) to qualify as knowledge if
there is no defeater.

Now, we can think of a case where the epistemic goodness in A-phase is satisfied
but the belief fails to be knowledge because of the problem in C-phase. Consider an agent
who (for whatever reason) actually believes that crystal ball gazing is a reliable way to
tell tomorrow's temperature. She looks at the crystal ball and comes to believe that
tomorrow's high would be 62, which happens to be true. If she has the proposition about
the reliability of crystal ball gazing in her cognitive system, 16 and she lacks beliefs that
undermine this proposition (for example, a belief that it didn't get the correct temperature
yesterday would be one such proposition, but suppose that she never checked the actual
temperature to confirm her beliefs), we can imagine that her belief that tomorrow's high
is 62 may turn out to be correct relative to her cognitive system. If we can assume that
she is also intellectually virtuous, this belief may be considered as epistemically good in
A-phase. However, even if this is the case, the lack of objective reliability, i.e., the failure
to satisfy the epistemic goodness condition in C-phase, prevents this belief from being a
case of knowledge, and this is a correct result.

How about a case where a belief is epistemically good in C-phase but problematic
in A-phase? The case of Mr. Truetemp would fit this category. Here, the reliability of the
belief-forming mechanism provides the required epistemic goodness in C-phase, thus
proper acquisition. However, Mr. Truetemp's cognitive system does not contain any
proposition to provide a condition where the proposition he believes (a proposition about
the current temperature) is likely to be true. As it is described in the example, Mr.
Truetemp is just like an ordinary epistemic agent with regard to his cognitive system.
Then, his belief is incorrect. We may also think that he is not intellectually virtuous given

16) The correctness of the proposition that the crystal ball gazing is a reliable method should be doubted.
Later, I will explain one of the defeater conditions using this "global incorrect" problem in the elements in
one's cognitive system.
that he has a tendency to allow such incorrect beliefs. Thus, Mr. Truetemp’s belief is not properly maintained and it lacks the epistemic goodness required in A-phase. He does not know the current temperature, and this corresponds to our intuitive verdict.

Also, in the case of Plantinga’s Red Bus, which I mentioned earlier in this chapter, the agent’s belief that the bus is red is incorrect because the agent’s cognitive system contains the propositions about the denial of the existence of color red. The proposition that the bus is red cannot have a high probability to be true in the world represented by the relevant portion of the agent’s cognitive system. Thus, this belief is incorrect, and therefore it cannot be a case of knowledge. My verdict on this case is the opposite of Plantinga’s. He thinks that the agent knows that the bus is red because this belief is a result of a “properly functioning faculty,” namely, his perception. However, it is strange that Plantinga ignores the fact that this agent as a whole is clearly “malfuctioning”—if I may use Plantinga’s term—because he allows an incorrect belief. Having incorrect belief would be what the epistemic agent is not supposed to do for knowledge. So, I think the difference here between Plantinga and me concerning this example rather favors the plausibility of my proposal.

Finally, Bonjour’s clairvoyance cases including Norman should receive the same treatment. They are incorrect beliefs. In the case of Samantha who has evidence to believe that the president is not in New York but in Washington D.C., her cognitive system contains this proposition. Then, the probability that the proposition that the president is in New York is true given only the elements in her cognitive system is very low. That is, the probability that the president is in New York when he is in Washington
D.C. (or when it is reported to be in Washington D.C., depending on which proposition is in Samantha’s cognitive system) is either 0 or very low. Even in the case of Norman, the probability cannot be high because, by hypothesis, there cannot be any proposition about the president’s whereabouts that Norman may actually and potentially believe. And, just like the case of Truetemp, we have reason to think that both Samantha and Norman lack intellectual virtue if we assume that they would allow such incorrect beliefs whenever their clairvoyance kicks in.

2. Defeater Conditions for Epistemic Goodness

2.1. Defeater in C-Phase

Suppose that science can tell us what constitutes the belief-forming faculty or inner mechanism that is responsible for the existence of Bp, and thus help to individuate the relevant faculty. In order to assess the reliability of this individuated faculty, what we need to know next is the relevant domain where this faculty works (and would work). A belief-forming mechanism, as long as it is understood as an inner mechanism or faculty of the agent, always works under a certain environment, and it is surely imaginable that the same faculty may have different reliability depending on the environment. Not until we can fix the relevant environment for a given belief production can we settle the

17) The problem of the relevant faculty individuation can be considered as an adaptation of the generality problem to faculty reliabilism. (Concerning the generality problem, see chapter 1, p. 28.) The generality problem was originally raised against Goldman’s version of process reliabilism. It is based on the fact that there could be infinite number of ways to describe the process type to which a given process token belongs. Some epistemologists such as Alston and Plantinga claim that this problem might be avoided if there are psychologically real internal processes (Alston) or inner faculties (Plantinga). However, even if that is the case, how to individuate the relevant faculty involved in a given belief formation would be no less difficult.
reliability standard that I use to measure the epistemic goodness in C-phase. How do we approach this issue?

Here, I would suggest a distinction between “a world” and “an environment” (or “a circumstance”) that constitutes a world. The thought is that a belief-forming mechanism is always used in a specific environment, but this environment is one possible environment among many, which altogether constitute the relevant world in which the reliability of a belief-forming mechanism is measured. If science can come up with an idea of what amounts to the belief-forming inner mechanism, then the environment under which the mechanism is working would have to contain all and only elements that the mechanism takes into account for producing the belief. Suppose that “the particular shape, color and size of a furry object that just darted across my visual field” is the element that my belief-forming mechanism takes into account for producing the belief. Many different environments contain this particular shape, color and size of a furry object as its element. For example, there would be an environment that contains this particular shape, color and size of a furry object and being on the grass hill on Wednesday morning. This environment would be different from an environment that is made up with the same particular shape, color and size of a furry object and being in the woods on Thursday evening. There also would be an environment that is made up of this particular shape, color and size of a furry object and having no rabbit around, so on and on. However, by virtue of sharing the same elements that the mechanism takes into account for producing the belief (that is, the particular shape, color and size of a furry object), we can think of a broader category that is made up by these many different environments. This broader
category, the sum of these environments, is what I would like to think of as the world in which the reliability of the mechanism is measured. Then the reliability of a given mechanism is understood as the reliability of it in the world where the belief formation is taking place. That is, it is reliable in the world constituted by the sum of actual and possible environments, each of which contains a common element for the activation of the given mechanism. In the above example, again, the common element in all environments that make up the relevant world is the particular shape, color and size of a furry object.

Sometimes, a particular environment in the world may contain defeating elements. That is, we may have a case where a belief-forming mechanism, which is reliable in most of the environments—thus it is generally reliable in the given world—is no longer reliable in a particular environment. I consider a defeating element to be an element that exists in the environment but the belief-forming mechanism does not take it into account for the activation of the mechanism. For example, consider the barn façade case. Suppose there is a belief-forming mechanism that takes into account a particular shape of a barn in order to activate itself. Let's call that particular shape of a barn N. We may suppose that this mechanism is reliable in the world where it is used, which is the sum of all environments that contain N. However, in addition to N, the particular environment where this mechanism is now being used also contains barn facades, the perceptual equivalent of what the mechanism takes into account (N), which, therefore, is enough to activate the same mechanism. Let's call these barn facades, the perceptual

---

18) A particular environment itself, or being in a particular environment, can be considered a defeater by virtue of the defeating elements in it.
equivalent of a barn, M. Now, to be sure, this particular environment that contains both N and M is repeatable, and an extension of it would make up a world, too. However, since the world where the reliability of the mechanism is measured is a sum of environments that contain N, a world that contains N and M would be a portion of the world that only has to contain N. We can reasonably measure whether the given mechanism is reliable in that portion of the world, that is, relative to the "sub-world" that is made by a sum of environments containing N and M. We may say that the mechanism is unreliable in this "sub-world" because in this "sub-world," the mechanism may take a perceptual equivalent (the thing that is perceptually indistinguishable from N) in order to activate the barn belief-forming mechanism, and if it does, it would generate false beliefs. So, the reliability of the mechanism becomes low in this particular environment because in the extension of this environment (i.e., in the "sub-world" that contains N and M), this mechanism is unreliable.

However, this would not be the problem of the mechanism itself because it is still a reliable one in the (larger) world, which is made up by the sum of environments that only need to contain N. That is, the mechanism would generate true beliefs in many other environments that only contain N and no M. The characteristic of the mechanism should be determined in this larger world, not in the "sub-world" that always contains a perceptual equivalent M. Such a characteristic is what is needed for knowledge. Therefore, the epistemically positive features coming from being produced by a reliable belief-forming mechanism in this (larger) world should be preserved. That is, we cannot deny that the belief is produced by a reliable mechanism. Because of that, rather than
denying or alternating the reliability of the mechanism in accordance with the
environment, postulating the notion of defeater makes a lot of sense here. Thus, in such a
case, we should say that, even though the portion of epistemic goodness in the C-phase
(namely, the proper acquisition by virtue of being produced by a reliable mechanism) is
still there, its contribution to making a true belief knowledge is now limited due to the
special environment in which this mechanism becomes unreliable. We can do this
without denying that this mechanism is a reliable one in general. In fact, there is little
doubt that, for example, a true belief gotten from perception in a barn façade environment
is properly acquired, thus has epistemic goodness in the C-phase. However, this proper
acquisition cannot be the kind of epistemic goodness that is required for knowledge in the
C-phase because the specific environment is the defeater. This is what I will call defeated
epistemic goodness. To summarize, in the barn façade case, barn façades are defeating
elements existing in the specific environment, and this specific environment is the
defeater. And what is defeated is the effect of \textit{prima facie} epistemic goodness that makes
a true belief knowledge. As the result, what the accidentally true barn belief has is
defeated epistemic goodness.

If the way I approach the barn façade case with the distinction of the world and
the environment is acceptable, we may also have an effective way to block at least a
certain type of Gettier problem. In the original Gettier case where Smith believes that the
person who will get the job has ten coins,\textsuperscript{19} the initial false belief that Jones will get the
job has been considered as justified despite its being false. The reason would be that

\textsuperscript{19) See chapter 1, p. 4.}
Smith’s boss who had told him about Jones’ getting the job is usually a reliable source of such information. That is, the belief-forming mechanism of Smith (namely, testimony) is reliable in the given (larger) world. However, in that particular environment in which the boss had decided to make a practical joke on him, such a mechanism fails to be reliable. Thus, instead of saying that Smith’s belief that Jones will get the job is a justified false belief, we could now say that it is a belief whose epistemic goodness (due to reliable production) is defeated (due to a special environment). Then, we may go on to say that a belief that has only defeated epistemic goodness cannot transfer the kind of epistemic goodness required for knowledge to a new belief produced by another mechanism. This mechanism would be a second-order mechanism that uses beliefs as input. In this specific case, this mechanism (presumably, the deductive reasoning of Smith) uses a belief that has defeated epistemic goodness (the belief that Jones will get the job) as input. Thus, the mechanism that allows him to extrapolate to a new belief that the man who will get the job has ten coins in his pocket, although reliable, cannot produce an epistemically good belief that qualifies as knowledge because its input is a belief whose epistemic goodness is defeated. It is not to deny that *prima facie* epistemic goodness can be transferred in such a way. Similar to the cases where justification can be transferred through a justified false belief, *prima facie* epistemic goodness can be transferred to a newly formed belief even if a belief that has defeated epistemic goodness is used as an input. However, I think that the new belief cannot acquire undefeated epistemic goodness, the kind of epistemic goodness that

---

20 A mechanism that uses other doxastic states as its input can only be “input dependently reliable,” which means that a reliable performance of the mechanism is expected only when input is true.
goodness required for knowledge, if its *prima facie* epistemic goodness is rooted in defeated epistemic goodness.

Indeed, if we generally apply the distinction of the world and the environment that I suggest here, the reason why we acquire a false belief would be either because the belief-forming mechanism is unreliable in the world (thus, in this case, the produced belief does not have *prima facie* epistemic goodness in the C-phase) or because the reliable belief-forming mechanism becomes unreliable in a given specific environment. Thus, in the latter case, the epistemic goodness is defeated due to the specific environment. Then, many Gettier examples that are constructed on the basis of a reliable second order belief-forming mechanism that uses justified *false* belief as input can be blocked. According to my proposal, this justified *false* belief now amounts to belief whose epistemic goodness is defeated. Thus, for a solution of a Gettier problem, what we need to add is that defeated epistemic goodness is not appropriate as an input for any second-order belief-forming mechanism to produce a belief that has undefeated epistemic goodness even if the second-order mechanism is input dependently reliable. In my opinion, this explains the epistemic defect of the original type of Gettier example such as Smith’s job and coins. When the environment and the world are distinguished as I propose, it also provides a slightly different way to look at the case of global deception, which has been a troublesome case for reliabilists. This will be discussed in the next section.

207

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
[Indefeasibility condition for epistemic goodness in C-phase]: Bp has indefeasible epistemic goodness in C-phase only if *prima facie* epistemic goodness is not defeated by the environment where the belief is produced.

[Environment that defeats *prima facie* epistemic goodness in C-phase]: An environment defeats *prima facie* epistemic goodness in C-phase iff a reliable belief-forming mechanism becomes unreliable in the given environment due to the elements in the environment that this mechanism does not take into account.

2.2. A Global Deception: The New Evil Demon Case

According to the core idea of process reliabilism such as Goldman’s theory, a belief is justified if and only if it is produced by a reliable belief-forming process. As we can see in the case of the New Evil Demon (NED) scenario, the alleged problem of reliabilism is that, if a completely unreliable mechanism, such as “perception in the demon world,” can produce “justified” (or epistemically good) beliefs, we do not need the reliability requirement for justification.\(^{21}\) However, first, we should note that, although this case may work against the alleged necessity of reliability or truth-conducivity for justification (i.e., the reliabilist position on justification), it is not strong enough to question the legitimacy of a reliability requirement for knowledge (i.e., the reliabilist position on knowledge). And, furthermore, if we can explain the reason why we are compelled to think that such a belief is “justified” or has at least some kind of positive epistemic features using the idea of reliability, then the force of this criticism

\(^{21}\) For the details of this example see chapter 3, p. 148.
will be sharply reduced. In my proposal, the NED problem will be treated as a case where the victim’s belief has only a part of epistemic goodness (epistemic goodness in A-phase through a proper belief maintenance), which is another way to explain the source of the epistemically positive factor in this example. However, before that, I first want to consider the idea that even reliabilists do not have to be in the difficult position that their critics want them to be in concerning this example. This is related to the “defeater” in the C-phase relying on the previous distinction of the world and the environment. The reason we judge the perception of the demon victim as unreliable is because it is unreliable in the “demon world.” However, what is the demon world? Would it be a world, or should it rather be a special environment similar to what we have seen in the case of a barn façade environment? Depending on which, the force of the demon world example against reliabilism can be appreciated differently. My contention is that, since it could be either way given that we utilize the distinction of the world and the environment, reliabilists can choose a less damaging interpretation of this example.

We may think of the case in the following way first. To every environment E, there would be a corresponding environment E* that is characterized by the demon’s total deception of all the elements in that environment. E and E* are completely different environments. For example, in E, there is an actual barn with a particular shape and size, whereas in E* there is the image of the exact same barn. If E-1 contains a real barn and propped up real barn facades, E-1* contains mere images of the exact same barn and the exact same propped up barn facades. As we have considered, E is considered as a part of many environments that constitute the relevant world, W, where the reliability of
perception is evaluated. E-1 belongs to a “sub-world” of W. Perception is generally reliable in W, and thus provides the *prima facie* epistemic goodness in the C-phase except that it is used in E-1, a special environment where perception becomes unreliable. If it is used there, the *prima facie* epistemic goodness is defeated. Now, the question is whether E* should be considered as a part of W, or as making up an independent world W*, namely the demon world. The victim is in E* and the belief-forming mechanism he has there is unreliable to be sure. But, is this unreliability due to the victim’s being in E* in W or due to being in W*?

I think that the normal assumption is that the victim is in W* by virtue of being in E*. That is, the reliability of the victim’s belief-forming mechanism should be evaluated in W* and that’s a sure way to claim it is reliable. However, this option makes the following fact completely irrelevant for this case: The victim’s belief-forming mechanism would be reliable in W. If the critics of reliabilism cannot utilize the idea that the victim’s belief-forming mechanism would be reliable in W, it does not help them. The problem is that the other essential claim by the critics, namely the claim that the demon victim is “justified” in W*, seems to lose intuitive support. If we consider the demon victim’s false beliefs in W* without envisioning W where the victim’s same beliefs could be all true, the only positive epistemic worth of those false beliefs is coming from relationships among the beliefs, such as coherence. Then, the critics have to say that those false beliefs are justified because of that. However, if so, they have to presuppose that epistemic justification is a matter of coherence. And this is attacking one notion of epistemic
justification with another notion of epistemic justification, which I believe is not the intended strategy taken by the critics who present this as a counterexample.

Or maybe it is their strategy. It might be the case that the victim is epistemically justified in $W^*$ with mere internal coherence. However, my point here is that we do not have to interpret the NED example in this way because there is an alternative way to interpret it. I think it is possible to treat $W^*$, the world made up by $E^*$, as a "sub-world" of $W$. This sub-world is constructed by environments that must contain both (a) my perceptual experience $X$ and (b) the demon's deception of providing $X$ directly in me, for example. The purpose of doing this is to make the status of $E^*$ to $W$ similar to the status of $E$ to $W$. The intended result is to provide a ground for thinking that $E^*$ or the sub-world $W^*$ would be merely a portion of $W$, which does not affect the reliability assessment of the victim's mechanism. The victim's mechanism is reliable in $W$, thus there is a *prima facie* epistemic goodness in a belief produced by this mechanism regardless of where it is produced. Of course, when a belief is in fact produced in $E^*$, it defeats the portion of epistemic goodness in C-phase that comes from proper acquisition because the mechanism is unreliable in that environment. It is just like the barn façade case where the barn façade environment defeats the portion of epistemic goodness coming from the reliable perception. In other words, the demon world problem could be treated as a case where a reliable mechanism produces a belief, which has *prima facie* epistemic goodness, but such epistemic goodness is defeated because it is produced in the environment that defeats it, i.e., in the environment where the same mechanism is no longer reliable. As a consequence, the belief whose goodness is defeated in this demon
environment is either false (as is normally the case when a demon deceives a victim) or cannot qualify as knowledge even when it is true. In this way, we can explain what is epistemically positive about belief acquisition in the demon victim, which is often ignored by both reliabilists and their critics. That is, what is epistemically positive is still the reliability of the perception of the victim assessed in the relevant world W, namely the *prima facie* epistemic goodness in C-phase.

Of course, this solution relies on the fact that there is in fact a non-demon world where the victim uses his perception. If so, we can reasonably expect that such a non-demon world is a major part of his world. So, at least intuitively, this solution works nicely when the global deception is a temporary condition on the agent, e.g., when the agent is a so-called “abducted” demon victim, who used to be a normal epistemic agent in a normal world but now belongs to a demon world, which is in fact a standard way to set up this example. However, alternatively, if we have to assume that the victim is permanently trapped in the demon world, it might be insisted that judging the victim’s perception as being unreliable is intuitively more plausible. It might be that, if the whole world is infested with barn façades such that I cannot distinguish from real ones, saying that my barn perception is unreliable is more acceptable than saying that it is reliable but it is in an environment that makes it unreliable. However, again, in this case, it seems hard to maintain that the victim’s beliefs are epistemically justified.

What this shows is that the NED example is not compelling enough to force us to give up the reliability standard for epistemic evaluation entirely since the intuition that the unreliably produced belief still has epistemically positive status ("epistemic
justification”) can be explained using the same idea of reliability. This is what I believe Sosa attempts to achieve in his approach to this problem with the idea of a relativized world. However, according to Sosa, even though a belief is “justified in a normal world” and “unjustified in the demon world” at the same time, it is eventually “unjustified” as long as it is in the demon world. The fact that the same belief would be justified if it were in a normal world is not a significant factor concerning the epistemic status of the belief, although one might claim that this is the source of our intuition that there is something epistemically positive about this belief. In my account, so-called “justification in a normal world,” or epistemic goodness based on truth-conducivity in the relevant world, is rather a basic epistemic status of the belief instead of possessing an alternative epistemic status as in Sosa. However, the actual environment (the demon environment) is the reason why this prima facie epistemic goodness is defeated.

Now, even if we must judge the victim’s perception as unreliable because we believe that the relevant world should be the demon world rather than a normal world (as in the case of a permanent demon victim), we can still explain what is epistemically positive in such a belief. In fact, several attempts so far made by virtue reliabilists, including the one from Sosa, seem to presuppose that the beliefs in the demon world are unreliably produced. Yet, if we truly utilize the idea of intellectual virtue as I have suggested (and also including the notion of correctness of a belief), it has a clear advantage in explaining the issue of what is epistemically positive in demon victim’s belief even we accept the unreliability of the belief-forming faculty. Now, let's briefly

22) Sosa’s approach to this problem was discussed in the previous chapter. See pp. 213

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
consider how Plantinga's account might approach this issue to see how this type of "virtue" theory can do more than generic reliabilism does with this problem.

Plantinga does not mention the NED problem explicitly, at least, not as the one he needs to solve for the success of his theory. It is because, for him, what is more important in epistemology is "warrant," the property that makes true belief knowledge. And if we apply his conditions for warrant to the NED case, the demon victim's belief would not be a warranted belief. One of conditions for warrant demands that the environment be similar to the one where the designed faculty is supposed to work. A demon environment wouldn't be the one that, for example, the designer of our perception would have in her mind when she designed it. Since it is our intuition that the victim's beliefs cannot be knowledge even when they are true (and even justified), Plantinga's verdict that his beliefs do not have warrant creates no problem. Furthermore, Plantinga does not need to worry about the intuitive justificatory status of the beliefs. Presumably, Plantinga would agree with the critics of reliabilism, and would say that the victim's belief may well be justified. For him, this fact would only support his other claim on epistemic justification, namely, that epistemic justification is essentially deontic, and justification in this sense is neither necessary nor sufficient for warrant.

What is interesting here is that, in Plantinga's theory, it is possible, and indeed very plausible, to think that the victim's faculty (say, perception) is a properly functioning faculty even if it fails to fetch truth. That is, in Plantinga, the NED example would be understood as a case of a properly functioning faculty in an unintended (hostile) environment. The unreliability, the tendency of not getting true beliefs, is not the problem
of the faculty that he is using but it is due to the world where the victim is located. This is similar to my attempt to explain the circumstance relying on the idea of defeaters, the environmental glitch. However, in my attempt, instead of proper functioning, a notion which eventually needs philosophical qualification and clarification especially concerning the idea of the design plan, I use reliability as the standard. Also, instead of thinking of a specific “favorable” environment as a portion of essential elements for “warrant,” I consider those specific environments that makes a given faculty unreliable as a defeater.

However, what we are talking about here is a different approach to the NED case using the idea of intellectual virtue. Even here, Plantinga’s idea gives us a hint. A noteworthy perspective is that the demon victim’s faculty of perception (or maybe the victim as a whole) is properly functioning in the given unfortunate world or environment. At least in spirit, the idea of intellectual virtue in my proposal is intended to capture a certain aspect of Plantinga’s idea of a properly functioning faculty, especially in the sense that the primary goal of a faculty’s proper function is not truth but something else such as doing what it is supposed to do or following design. Given that the appropriateness of the faculty here does not have to mean truth-conducivity, the intellectual virtue required in A-phase of a belief or in belief-maintenance in my proposal is similar to the idea of proper function. Thus, the positive epistemic feature that we find in the demon victim’s belief, which is usually identified as at least of a kind of “justification” by both

23) As a solution to the NED problem, Goldman has suggested a distinction of strong and weak justification, and attributed weak justification to the demon victim’s beliefs. Goldman, Alvin, "Strong and Weak Justification", in Philosophical Perspective Vol. 2: Epistemology (edited by J. Tomberlin), Ridgeview, 1988.
reliabilists and their critics, is due to the victim’s being proper as an epistemic agent regarding belief-maintenance. Of course, there are also significant differences. First, I do not think that a faculty’s proper function should be measured in comparison to its design plan as Plantinga claims. Furthermore, in Plantinga, it is a belief-forming faculty that should be appropriate (properly functioning), and as a consequence, a belief that is likely to be true is produced. I don’t think that is right. In my proposal, I distinguish appropriateness of a belief (being true as the desirable quality of a belief in C-phase, and “being correct” in A-phase) and appropriateness of the belief-forming faculty (reliability), yet neither of them is identified as intellectual virtue. Intellectual virtue could be understood as the appropriateness of a belief-maintaining faculty because it is the agent’s tendency to maintain these correct beliefs. As a result, my proposal can explain what is going on in the NED example without relying on heavy items like a design plan. As we have seen, there surely is something epistemically good in those beliefs even without the actual reliability. It is the generic sense of proper functioning of the epistemic agent. In my proposal, it is intellectual virtue. I believe that this explains the source of our intuition that those beliefs are “justified.” However, this part of epistemic goodness in the A-phase is not accompanied by the epistemic goodness in the C-phase. If we have to say that the belief-forming mechanism is unreliable since we assume that the demon world is the relevant world where the reliability of the mechanism should be measured, then this part of epistemic goodness is the problem of the victim’s belief, and it is the reason why it cannot be knowledge.
2.3. Defeater in A-Phase

Earlier, I have mentioned that the case of Mr. Truetemp or Bonjour’s clairvoyance cases can be dealt with in my proposal as cases of having “incorrect” beliefs. Thus, they lack epistemic goodness in the A-phase. However, in these cases, due to the reliability of the machine that tells temperature, and due to the clairvoyance, the beliefs should be considered as properly acquired, thus as having the epistemic goodness in the C-phase. However, it might be possible that, by changing the story, we may make the belief correct. For example, let’s consider the Mr. Truetemp case, and let’s change the example in the following way. Instead of having no reason at all for believing that the current temperature is 62, suppose that Mr. Truetemp believes a plausible but false story of why he is so sensitive to and reliable in telling the truth about the current temperature. Suppose that he thinks he is a victim (or beneficiary) of special genetic mutation due to his father’s job in a nuclear plant. He thinks that, due to this mutation, he somehow has a special ability of guessing temperature correctly. All of these propositions are the result of his wild imagination but suppose that they are in his cognitive system as a form of his actual beliefs. Because of that, whenever a number comes in his mind concerning the temperature, he just believes that as the current temperature. These are true beliefs about current temperatures reliably produced by his belief-forming mechanism. Furthermore, these are correct beliefs because they are reliable indicators of truth relative to the world represented by the relevant portion of his cognitive system. That is, suppose that the world is like what the revised Mr. Truetemp thinks it is (including his possessing a special ability due to genetic mutation). If so, the probability that the number
spontaneously coming into his mind is indeed the accurate current temperature would be
high. Thus, it constitutes the correctness of a belief in A-phase. Finally, we admit that, if
Mr. Truetemp is described in this way, there is no particular reason to doubt that he has
intellectual virtue. So, the temperature belief of the revised Mr. Truetemp seems to satisfy
all of the conditions for epistemic goodness I so far have proposed. Now, the worry is
that, even so, such a belief is not a case of knowledge.

I have considered one way to deal with a similar type of example when I discuss
how Gettier problems (especially the ones that rely on false beliefs) can be handled using
the concept of a defeater in the C-phase. There, the idea was that, in case of false belief,
the portion of epistemic goodness coming from proper acquisition is always questionable
because either its belief-forming mechanism is unreliable or its epistemic goodness is
defeated. Thus, we might want to say that those false beliefs of revised Mr. Truetemp do
not have epistemic goodness, and therefore they cannot transfer it to the final belief.
However, this wouldn't be exactly what is going on in the revised Mr. Truetemp case.
Unlike the Gettier case where the false belief is being used as an input to the so-called
second order belief-forming mechanism that generates the final belief (such as inductive
or deductive reasoning), Mr. Truetemp's temperature beliefs do not depend on his false
beliefs in terms of their acquisition. There is a separate belief-forming mechanism,
namely the Tempcomp machine in his head, which is reliable yet does not use any false
belief as the input. And, even if we might be justified in degrading all false beliefs in
terms of belief acquisition, we cannot expect any immediate and intuitive problem in the
idea that a cognitive system may contain false beliefs as well as the idea that having a belief complying with them still should be considered as epistemically desirable.

However, even though a suggested Gettier solution may not work here, it reminds us that the portion of epistemic goodness in the A-phase could be defeated just as the portion of epistemic goodness in the C-phase can be defeated in a special environment. What then would be such a special circumstance in the A-phase? When we admit that certain subjective elements, such as correctness relative to one’s point of view, should be necessary for our knowledge, the reason is that it would eliminate a certain type of accidentality, namely the accidentality coming from doing an unexpected thing. Believing only what is allowed in one’s cognitive system is necessary for knowledge in this regard. However, not all cases of believing what is allowed in one’s cognitive system are epistemically worthy enough to produce knowledge eventually. It depends on the “quality” of the cognitive system that a given belief complies with. And, in this case, the quality should be determined by how the cognitive system contributes to having true beliefs. For example, consider a case of a mentally ill person (who incoherently thinks he is Napoleon). If the elements in his cognitive system are completely inconsistent with each other, first, at least concerning some beliefs, it would be hard to make sense of judging whether a given belief is a reliable indicator of truth in his world (or in the relevant segment of the world). 24 Second, even if we somehow bypass this problem, it

24) Although there is a problem in this person’s cognitive system, the perceptual beliefs of this person, for example, might be correct because the relatively independent portion of his subjective world to judge the correctness of his perceptual belief does not have to include the propositions concerning who he thinks he is. In the same way, we can find enough epistemic goodness in Mr. Magoo’s belief on things that are very close to him provided that we might be able to separate his world into more than one relatively independent and self-sufficient segments.
would still be questionable how epistemically valuable it is to be a reliable indicator in such a “badly shaped” world. So, the issue of the quality of the subjective world is relevant and needs to be considered.

I think this type of example bears a similarity to the example of the Coherent but Inflexible Climber. In this example, a rock climber hit by a burst of high-energy radiation becomes non-responsive to experience. As a consequence, the elements in the agent’s cognitive system that constitute his subjective world are causally disconnected to the features in the world although the whole system maintains internal coherence. For this reason, although his belief that a hawk is circling the sky above the valley is correct from his “point of view,” enjoys high probability to be true given the other elements in the cognitive system, there is no real sense of “point of view” or “subjective world” for this person. What is the common feature between this case of the Rock Climber and in the cases of hallucination or extreme fantasy such as the case of the revised Mr. Truetemp or the case of a madman? In my opinion, it is the lack of causal contact with the actual world in the process of shaping the relevant segment of one’s cognitive system (regardless of whether it ends up with a correct representation of the world or not). Then, even if such a fantasy may avoid a blatant contradiction (and thus, provide at least a coherent segment of the world in which the correctness of the belief can be assessed, as in the case of the rock climber), it would be problematic to admit such a point of view as a contribution to the eventual epistemic goodness in the A-phase. Thus, I think that a defeating circumstance that negates the epistemic goodness coming from apparently


220
proper belief maintenance would be that elements in one’s cognitive system do not have causal contact with features in the actual world.

[Indefeasibility condition for epistemic goodness in A-phase]: Bp has indefeasible epistemic goodness in A-phase only if *prima facie* epistemic goodness is not defeated by the bad quality of the cognitive system.

[Cognitive system that defeats *prima facie* epistemic goodness in A-phase]: A cognitive system defeats *prima facie* epistemic goodness in A-phase iff the relevant portion of the agent’s cognitive system contains propositions that do not have causal contact with features in the actual world.

This would allow that a portion of a cognitive system that contains significant falsity would always be able to defeat the epistemic goodness in the A-phase. That is, even if Bp is correct relative to the portion of the cognitive system that contains a proposition that the agent is Napoleon, such correctness, and thus *prima facie* epistemic goodness, has not much to do with knowledge. Furthermore, my defeater condition allows that, even if the cognitive system may not contain falsity, it can defeat *prima facie* epistemic goodness if there is no causal contact between the true propositions in the cognitive system and the actual world. As an example, consider the case of the Coherent but Inflexible Climber who returns to the original spot of the accident. Here, his beliefs about the hawk, the blue sky, the shape and color of the cliffs, and so on, could all be true. However, none of them is caused by the features in the actual world at that moment.
It would be only a miraculous coincident that those propositions are all true. For another example, consider an isolated being who believes himself to be a brain washed CIA secret agent out of his own self-generated fantasy. If he is indeed a brain washed CIA secret agent, this proposition in his cognitive system is true. However, since this proposition does not have any causal link to the features in the actual world, again it would only be a coincidence. My defeater condition deals with these cases too. If we apply this idea to the demon victim, we may say that even the portion of *prima facie* epistemic goodness in the A-phase, which I admit that there is, would be eventually defeated when the quality of the cognitive system is considered.

Finally, this idea would lead to an interesting result concerning the knowledge status of Foley’s Lightening Victim example. Here, Foley wants us imagine that an epistemic agent’s beliefs are as accurate and as comprehensive as is humanly possible. Thus, she has true beliefs about the laws in universe, and using them, she can explain almost everything in the universe. In short, she has a perfectly veridical and explanatorily perfect system. Foley’s claim is that, if this is the case, the agent has knowledge about all those things that she believes even if the beliefs are all caused by a strange way such as a hit by a bolt of lightening. Although it might be unclear, let’s take it that what Foley means in this example is that all the beliefs of this agent are created by a bolt of lightening at the moment that she is hit. Thus, all the elements in the entire cognitive system of the agent are created at once. Furthermore, this cognitive system represents the

---


---

222
features in the world accurately almost up to the degree of being omniscient. Then, naturally, any proposition in this system is correct from the agent’s point of view.

This example can be easily controversial. In one string of thought, it is possible to understand this example as a similar kind of isolation and fantasy case, only on a large scale. In this interpretation, the victim’s cognitive system happens to represent the world accurately. It would be just like the inflexible rock climber coming back to the original spot where he had been hit by the lightening but still being disconnected. Then, no matter how reliable the lightening victim’s belief is in her cognitive system, we may want to disregard all these factors as if it is a groundless fantasy. If so, it would be considered another case of epistemic goodness in the A-phase being defeated. However, the equally plausible interpretation is that the lightening makes an already existing cognitive capacity of the victim super-accurate. Then, the elements in her cognitive system, all the true propositions about the universe, can be considered as being newly entered in it by those renovated faculties. In this case, there are causal connections between those elements and the features of the world. And in this case, saying that the victim’s beliefs are correct from her point of view is a lot more meaningful than the previous understanding. We may think that, in this case, the epistemic goodness in the A-phase is unproblematic.

However, even with the latter understanding of the maintenance phase, whether those beliefs are epistemically good enough to be knowledge still depends on what we consider as the belief-forming mechanism in this case. If the refined faculties after the lightening hit are the ones that produce all the beliefs, then there is no reason to deny their reliability, thus those beliefs are knowledge. Even though the lightening itself is the
cause of all of his beliefs, if it is a reliable belief-forming mechanism (i.e., it will provide any human being it hits with massive amount of true beliefs as well as a perfectly veridical and explanatorily perfect system), then I think the victim knows all the things she believes. However, if the lightening is the cause of all of his beliefs, yet it is just a one time accident (thus, if it is not a reliable way to acquire true beliefs), then I don’t think any of her belief would be knowledge.

3. Conclusive Summary

In this final chapter, I have proposed an account of epistemic goodness, a necessary property for knowledge. I have noted that a belief can be seen in two different directions and thus should be evaluated accordingly. A belief’s C-phase is the result of seeing a given belief as a consequence of a belief-forming mechanism. This belief-forming mechanism should have a tendency to produce true beliefs, thus should be reliable, in order to provide the epistemic goodness in the C-phase. Meanwhile, a belief’s A-phase is the result of seeing a given belief as an antecedent for the agent’s genuine assent to the proposition that she believes. In this sense, a belief itself is considered as a tendency to generate genuine assent to the proposition believed. So, using the idea of a cognitive system and the notion of correctness, I suggested a way to evaluate a belief as such an antecedent condition. Then I added the requirement of intellectual virtue also as relevant to the A-phase. Therefore, a correct belief out of intellectual virtue, a tendency to maintain only correct beliefs, provides the epistemic goodness in the A-phase. Only a belief that has the epistemic goodness in both phases can be considered as a candidate for
knowledge. After that, I considered the defeater conditions in each phase. A special environment has been considered as a defeater in the C-phase, while the quality of the cognitive system has been a defeater in the A-phase. Eventually, a belief that has undefeated epistemic goodness in both the A-phase and the C-phase is knowledge. Throughout the chapter, well known examples in contemporary epistemology (such as the example of Mr. Truetime, Bonjour’s clairvoyance cases, the New Evil Demon example, Plantinga’s examples such as the Red Bus case and the Coherent but Inflexible Climber, and finally Foley’s Lightening Victim case) have been discussed and received new perspectives. My proposal produces acceptable results in all of those cases.
LIST OF REFERENCES


__________________________, "Reliability and Justification," *The Monist* 68 (1985): 159 - 174


______________________________, (eds.) *Essays on Knowledge and Justification* (Ithaca, N.Y.: Cornell University Press, 1978)


__________, *Warrant and proper Function* (New York: Oxford University Press, 1993)


