

Higher-Order Defeat, Epistemic Luck, and Knowledge Norm of Belief

by

Talha Olgun

A Dissertation Submitted to the
Graduate School of Social Sciences and Humanities in Partial
Fulfillment of the Requirements for

the Degree of

Master of Arts

in

Philosophy



**KOÇ
ÜNİVERSİTESİ**

April 29, 2022

Higher-Order Defeat, Epistemic Luck, and Knowledge Norm of Belief

Koç University

Graduate School of Social Sciences and Humanities

This is to certify that I have examined this copy of a master's thesis by

Talha Olğun

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Assoc. Prof. Erhan Demircioğlu (Advisor)

Prof. İlhan İnan

Prof. Murat Bağ

Date: 29.04.2022



To everyone I love

ABSTRACT

Higher-Order Defeat, Epistemic Luck, and Knowledge Norm of Belief

Talha Olğun

Master of Arts in Philosophy

May, 2022

This thesis aims to explain why one would be wrong to stick to his belief upon learning that he has been cognitively unreliable during the reasoning whereby he formed the belief. I show that taking truth as a norm or a goal, which sets the standards of justification for belief, is untenable. For instance, on the one hand, those who take truth as the norm cannot give an explanation because they constrain epistemic reasons to evidential considerations. On the other hand, those who take truth as the goal are susceptible to trade-off objections against consequentialist and instrumentalist views in epistemology. However, I argue that a better explanation, which resides in theories that explain justification in terms of knowledge, does not face such problems. A wide range of accounts in these theories has it that epistemic justification requires being in a position to know the target proposition. By endorsing this requirement, I argue that learning about one's cognitive unreliability constitutes evidence that one is not in a position to know the target proposition because this evidence indicates that, under these conditions, one has no possible way to hold a non-accidental belief that would qualify as knowledge.

ÖZETÇE

Yüksek-Dereceli Yenilgi, Epistemik Şans ve İnancın Bilgi Normu

Talha Olğun

Felsefe, Yüksek Lisans

Mayıs 2022

Bu makale, inancını oluşturduğu akıl yürütme sırasında bilişsel olarak güvenilirmez olduğunu öğrenen bir kişinin, inancına bağlı kalmasının neden yanlış olacağını göstermeyi amaçlamaktadır. İnancı gerekçelendirme standartlarını belirlerken, doğruluğu bir norm veya hedef olarak almanın savunulamaz olduğunu gösteriyorum. Örneğin, bir yanda doğruluğu norm olarak kabul edenler, epistemik nedenleri kanıtsal düşüncelerle sınırladıkları için, bir açıklama yapamazlar. Buna karşılık, doğruluğu bir hedef olarak ele alanlar, epistemolojide sonuçsalıcı ve araçsalıcı görüşlere yöneltilen takas (trade-off) itirazlarına karşı savunmasızdır. Ancak, gerekçelendirmeyi bilgi üzerinden ele alan teorilere dayanan daha iyi bir açıklamanın, bu tür sorunlara sebep olmadığını savunuyorum. Bu teoriler içindeki geniş bir görüş yelpazesi, epistemik gerekçelendirmenin hedef önermeyi bilecek bir konumda olmayı gerektirdiğini ileri sürer. Bu gerekliliğe dayanarak, kişinin bilişsel güvenilirmezliğini öğrenmenin, hedef önermeyi bilecek durumda olmadığının bir kanıtı olduğunu savunuyorum çünkü bu kanıt, kişinin bu koşullar altında bilgi olarak nitelendirilebilecek tesadüfi olmayan bir inanca sahip olmanın olası bir yolunun olmadığını gösteriyor.

AKNOWLEDGEMENTS

I would first like to thank my supervisor, Professor Erhan Demircioğlu, whose expertise is invaluable in formulating the research questions and methodology. I would also like to thank my tutors, Prof. İlhan İnan, Prof. Murat Bağ, Prof. Kenneth Westphal, for their valuable guidance throughout my studies. Lastly, I would also thank my family, as I could not have completed this thesis without their moral and material support.



TABLE OF CONTENTS

Abbreviations	viii
Chapter 1: Introduction.....	1
1.1 Setting the Stage.....	3
Chapter 2: The Norm of Truth and the Evidence of Unreliability	7
2.1 Reasons to Believe	8
2.1.1 The evidence of unreliability is not a rebutting defeater	12
2.2 Reasons for Believing.....	12
2.2.1 The evidence of unreliability is not an undercutting defeater.....	14
2.3 Taking Stock	16
Chapter 3: Truth as a Goal and The Evidence of Unreliability	19
3.1 Evidence as a Means	20
3.2 Some Challenges Against the Truth-Goalers.....	23
3.3 Taking Stock Again: Some Explanatory Desiderata.....	26
Chapter 4: My Proposal: Knowledge Norm and The Evidence of Unreliability	29
4.1 Knowledge-Norm of Belief	30
4.2 Being in a Position to Know	32
4.2.1 Cooper is lucky to have a true belief in (2).....	33
4.2.2 Cooper is lucky to have a true belief in (3).....	34
4.3 The Evidence of Unreliability as a Defeater.....	35
4.3.1 Reply to Some Objections	36
Chapter 5: Conclusion	42
5.1 Meeting the Explanatory Desiderata	42
5.2 Summing Up	43
Bibliography.....	47

ABBREVIATIONS

HOD	Higher-Order Defeat
HOE	Higher-Order Evidence
FOE	First-Order Evidence



Chapter 1:

INTRODUCTION

A phenomenon called higher-order defeat received much attention over the last two decades. It dates back to the debates about peer disagreement [Christensen and Lackey, 2013; Feldman and Warfield, 2013]. Suppose that two peers, equally competent in gathering and evaluating evidence, come to disagree over a matter, e.g., who the murderer is. Would learning about their disagreement require them to revise their beliefs? Learning about a dissenting peer is called higher-order evidence, which seems to put a normative demand on what one should believe. And if it has a rational influence on what one should believe, acquiring higher-order evidence may lead, on some occasions, to higher-order defeat.

These issues about higher-order evidence and defeat are a few of many that the debates about peer disagreement raises. Since then, higher-order evidence has been discussed in various contexts in which it comes in various forms [Whiting, 2020]. Suppose that a detective is working on a case to find out who the murderer is. (1) He may acquire higher-order evidence about what evidence he has available about the case. For instance, he may find out that a gun is detected near the crime scene, which suggests that a clue, some fingerprints, is to be found on it. (2) He may also acquire higher-order evidence about the strength and valence of his evidence, e.g., by learning that his assistant believes that the neighbor is the murderer. (3) He may acquire higher-order evidence about the normative import of his evidence bearing on the case. Some epistemologists claim that learning that his assistant believes that the neighbor is the murderer is not just further evidence, permitting the detective to agree with the assistant, but it gives him a compelling reason to believe that the neighbor did it. Lastly, (4) he may acquire higher-order evidence about his capacity to respond to (1)-(3). For instance, he may learn that the suspect is a friend, or he has been under the influence of some cognitive impairment that may affect his reasoning. In this thesis, I will focus on this last form of higher-order evidence.

This last form of higher-order evidence is relevant in cases where one acquires evidence that one has been cognitively unreliable. Here is a paradigm case¹:

¹ For similar cases, see Coates (2012), Horowitz (2014), Worsnip (2018).

Drugged Coffee

A homicide detective, Cooper, is working on a case to identify the murderer. He carefully gathers the relevant evidence that Diane's fingerprints are found on the murder weapon, and she has been seen around the crime scene at the time of the murder. After analyzing the available evidence, Cooper sees that the evidence points to Diane's guilt. He comes to believe that Diane is the murderer. Later, he learns from a reliable informant that someone had slipped a drug in his morning coffee before he started the investigation, which is highly effective in distorting his cognitive abilities, required for properly responding to the evidence bearing on who the murderer is.

This case elicits two opposing intuitions about what Cooper should do in his current condition. On the one hand, some epistemologists, who are called *Steadfasters* in the literature, hold that Cooper would be rational/justified to stick to his belief that Diane is the murderer [Kelly, 2005; Williamson, 2011, 2014; Titelbaum, 2015; Coates, 2012; Hazlett, 2012; Wedgwood, 2012; Lasonen-Aarnio, 2014; Weatherson, 2019].² According to them, learning about the drug does not give Cooper a reason to revise his belief. I do not aim to directly argue how this view is mistaken. However, I believe that the Steadfasters must explain away the intuition that it seems wrong to hold a belief under unreliable cognitive conditions.

On the other hand, some epistemologists, who are called *Conciliationists*, attempts to account for this intuition [Elga, 2007, 2010; Christensen, 2007, 2010; 2011, 2013; 2016; Feldman, 2005, 2006, 2007, 2009; Matheson, 2009; van Wietmarschen, 2013; Silva, 2017a; DiPaolo, 2018; Ye, 2020]. They accept that there is a sense in which it is wrong for Cooper to stick to his belief in the presence of the evidence that he has been unreliable in responding to the evidence relevant to the case. However, it is not exactly clear why it is wrong if Cooper decides to ignore the evidence of his unreliability and stick to his belief. There is not a consensus in the literature about how the evidence of Cooper's unreliability can put a normative demand on what to believe. I aim to give an explanation after objecting to some suggestions.

² Kelly (2010) and Lasonen-Aarnio (2020) later suggest how higher-order evidence may put a rational constraint on what one should believe on first-order matters.

1.1 *Setting the Stage*

Epistemologists draw various normative frameworks to back up their suggestions as to how the evidence of Cooper's unreliability may put a normative demand on what he should believe about the case. Many are concerned about the connection between truth and belief. However, I tend to see epistemic norms as related to knowledge and not merely the truth. So, in this paper, I will begin by objecting to these truth-centered views.

Before proceeding let me clarify Cooper's evidential situation to set the stage for our discussion. On the one hand, he has evidence pointing to Diane's guilt. For instance, Cooper learns that Diane's fingerprints are found on the murder weapon, and she was seen around the crime scene at the time of the murder. This set of evidence provides first-order evidential support for the proposition that Diane is the murderer. Let me call it 'first-order evidence' named after the way it relates to the target proposition. First-order evidence bears on the propositional content of the corresponding attitude by indicating whether the proposition is true. So, Cooper has evidence indicating that it is highly probable that Diane is the murderer.

On the other hand, Cooper has evidence that he is cognitively unreliable. He learns that he has been given a drug that is highly effective in impairing reasoning capacities. It provides first-order evidential support for the proposition that Cooper is cognitively unreliable. *The evidence of unreliability* indicates that it is highly probable that Cooper will make a mistake in evaluating the evidence bearing on who the murderer is. However, the evidence of unreliability does also seem to have a higher-order evidential bearing on what Cooper should believe about who the murderer is. Let me call it 'higher-order evidence' named after the putative way it may relate to the corresponding lower-order proposition.³ Also note that, when a piece of evidence has a higher-order evidential bearing on a corresponding lower-level doxastic attitude, it does not bear on the propositional content of the attitude. This means that the evidence of Cooper's unreliable cognitive condition does not speak to whether Diane is the murderer.

³ The term 'higher-order evidence' was introduced by Kelly (2005, 2010) and Feldman (2005). However, I do not plan to use this term to refer to what Cooper learns in hearing about the drug. I prefer to call it 'the evidence of unreliability' to differentiate this specific kind of higher-order evidence from others, such as, learning that oneself or the first-order evidence is biased in some sense. For a detailed classification of higher-order evidence, see Whiting (2020).

In brief, Cooper both has evidence indicating that Diane is probably the murderer and the evidence that he has been cognitively unreliable throughout the investigation, leading to his belief that Diane is the murderer. Our discussion, thus, mainly depends on how these two sets of evidence interact.

According to a group of views (§2) that takes truth as the constitutive norm of belief, only considerations speaking to who the murderer is can put a rational constraint on what Cooper should believe about the case. However, given that the evidence of Cooper's unreliability does not speak to who the murderer is, it seems normatively irrelevant to whether Cooper should believe Diane's guilt. So, they cannot explain how learning about his cognitive unreliability defeats Cooper's justification to believe that Diane is the murderer.

According to another group of views (§3) that takes truth as the ultimate epistemic goal of belief, as an epistemic agent, Cooper should aim to have true beliefs and avoid false ones. They hold that the evidence of unreliability indicates that Cooper is highly likely to hold a false belief about the case, i.e., he is at a high risk of failing to achieve the ultimate goal. They suggest that learning about his cognitive unreliability gives him a reason to refrain from believing Diane's guilt to avoid holding a false belief about who the murderer is. However, I will show that this suggestion is susceptible to objections leveled against consequentialist views in epistemology.

Having shown that the views which take truth as the norm that sets the standards of justification for belief are untenable, I will continue (§4) with my proposal. Following Williamson (2000), Sutton (2007), Smithies (2012), and many others, I hold that knowledge is the norm that sets the standards of justification for belief. Here is a preliminary outline of my argument: having a justification for holding a belief requires being in a position to know the target proposition; learning about his cognitive unreliability indicates that Cooper is not in a position to know who the murderer is; therefore, the evidence of unreliability defeats Cooper's justification to stick to his belief by overriding the prescriptive power of the evidence pointing to Diane's guilt.

Before proceeding, I should clarify a crucial issue about belief formation.⁴ Does Cooper have control over what he believes? Can he believe that Diane or someone else is the murderer at will? If not, can we hold him responsible for what he believes about

⁴ Thanks to Inan for raising this issue.

the case? If Cooper has no voluntary control over what he believes about who the murderer is, it seems absurd to claim that he is irrational or unjustified in believing so. I will not delve into the details of this age-old debate in epistemology. However, I believe that Cooper has a kind of control over whether to refrain from sticking to his belief after learning about his cognitive unreliability, even though he may have no control over what he believes. When he learns about his unreliability, he is given a choice to withhold belief. Although the evidence pointing to Diane's guilt makes it seem to him that Diane is the murderer, he still has a choice to suspend judgment over who the murderer is. Hence, merely ignoring the evidence of unreliability and sticking to his initial belief, regardless of whether it is formed at will, seems blatantly irresponsible.





Chapter 2:

THE NORM OF TRUTH AND THE EVIDENCE OF UNRELIABILITY

Virtually all philosophers accept that there is a connection between belief and truth. However, divisions occur about how to flesh out this connection. One group of views holds that there is a normative constitutive connection between belief and truth [Shah, 2003; Boghossian, 2003, 2005; Wedgwood, 2007]. Let me call them the *truth-normers*. According to the truth-normers, belief is subjected to the norm of truth due to its essential nature. By its nature and essence, belief has the truth as the norm that sets the standards of correctness for belief. Another group of views holds that the connection between belief and truth is rather goal-directed. Let me call them the *truth-goalers*. For the truth-goalers, epistemic rationality is a part of practical rationality, i.e., it is about achieving certain goals. Beliefs, they claim, are subject only to the norms that make the evaluation of beliefs depend entirely on how well they serve to realize the epistemic goal of having true beliefs and avoiding false ones [Foley, 1987; Steglich-Petersen, 2011; Cowie, 2014]. In this section, I will explore the prospects for the truth-normers to see whether they can account for the epistemic significance of the evidence of unreliability.⁵

As I said above, for the truth-normers, truth is the norm that sets the standard of correctness for belief. As Gibbard (2005) writes, “For belief, correctness is truth. Correct belief is true belief. My belief that snow is white is correct just in case the belief is true, just in case snow is white” (p. 338). To put it in a precise formulation:

Truth-norm of correctness: A belief that *p* is correct if and only if the belief that *p* is true.

This factual norm regarding the correctness of belief entails a prescriptive norm as to how one should go about in forming a belief. As Lynch (2009) puts out, “It is *prima facie* correct to believe that *p* if and only if the proposition that *p* is true” (p. 10). To put this norm in a precise formulation:

⁵ Suppose for this section that Diane is the murderer.

Norm of belief: One should believe that p if p is true.

Before proceeding, note that I will not discuss the problems that the truth-normers face in answering “What logical forms do these norms have?” or “How these norms can guide epistemic agents?” Instead, I will focus on how they answer the questions, such as “What can count as an epistemic reason to take a doxastic attitude?”, “Why should we follow evidence in forming beliefs?” My aim is to show that their views about epistemic justification and reasons cause them some trouble. I will show that forming a normative constitutive connection between belief and truth entails a strict form of evidentialism that excludes the evidence of unreliability as normatively irrelevant.

2.1 *Reasons to Believe*

To support my claim that the truth-normers limit the epistemic reasons to an exclusive set of considerations, where the evidence of unreliability has no place, I can survey several views that explain the evidential constraint on epistemic reasons by appealing to there being a constitutive normative relationship between belief and truth (Adler, 2002; Hieronymi, 2005; Shah, 2006; Raz, 2011; Wedgwood, 2013). However, that would be a daunting project. So, instead, I plan to present a widely acclaimed view (Shah, 2006) that has the main idea of establishing evidentialism as a conceptual truth about the nature of belief. I will also buttress my claim with an argument that I inspire from the object-given reason and state-given reason distinction in epistemology, especially from Hieronymi (2005).⁶

Shah’s argument relies on a psychologically unalterable fact about how we deliberate about whether to believe a particular proposition. He observes, “When we ask ourselves the deliberative question whether to believe that p , this question gives way to the question whether p is true, and so the only way for us to answer the former question is by answering the latter’ (p. 487), or alternatively put, ‘. . . the deliberative question whether to believe that p inevitably gives way to the factual question whether p , because

⁶ For the similar distinction with different terms, see Parfit (2001), D’Arms and Jacobson (2000), Olson (2004), Schroeder (2010, 2012), Sharadin (2015), and Way (2012).

the answer to the latter question will determine the answer to the former' (pp. 481–482). He calls this phenomenon *transparency*.

Transparency, he claims, results from a conceptual feature of belief. He accepts that the concept of belief is subjected to the truth-norm of correctness that governs belief within deliberative contexts. He writes,

When one deliberates whether to have an attitude conceived as a belief that *p*, one deliberates about an attitude to which one already applies the standard of being correct if and only if *p* is true, and so one is already committed to considering it with an eye exclusively to whether *p* is true. (pp. 488–489)

This constitutive connection between belief and truth, he holds, along with the conception of deliberations as a norm-governed activity explains the phenomenon of transparency.

One can see now how this account puts an evidential constraint on epistemic reasons. When one deliberates whether to believe that *p*, the only way to settle the matter is to answer whether *p* is true. What kind of considerations can serve one to answer the question whether *p* is true? They are naturally the kind of considerations that speaks to whether *p* is true. Such considerations are called evidential considerations—or evidence, so to speak—that bears on the truth of the target proposition. This entails that only evidence can be epistemic reasons to take a doxastic attitude towards a proposition within deliberative contexts.

Here is another argument for my claim that having the truth as the norm that sets the standard of correctness for belief puts an evidential constraint on epistemic reasons. Let me start by asking how can one justify oneself to believe that *p* according to the truth-normers? Note first that one could justify his belief that *p* in several respects. One can say that his belief that *p* promotes some value or a desirable end; for instance, by having the belief that *p*, one may promote some moral or practical value. However, for the truth-normers, a belief's epistemic status is not determined by whether it serves some external goal or value. They hold that a belief is subjected to a constitutive norm that distinguishes the correct beliefs from the incorrect ones based on the truth-value of their propositional contents. This shows that what can confer justification to the belief that *p* must be a constitutive reason indicating that *p* is true.

Moreover, putting this point about the ontological status of reasons aside, one can justify his belief that *p*—in a looser sense of justification, whatever that may be—by showing that he tries to comply with the truth-norm of correctness. That is, since this is a factual norm that makes an objective distinction between correct and incorrect beliefs, the truth-normers need a subjective norm that an epistemic agent takes as guiding him to comply with this norm of correctness. I stated above that the truth-normers could develop a prescriptive norm requiring one to believe that *p* if *p* is true. According to this norm of belief, then, one can justify his belief that *p* by pointing out that he abides by the norm of belief in forming the belief that *p*. Namely, he can justify himself by pointing out that he abides by the prescriptive norm of belief, i.e., he believes that *p* just in case *p* is true.

So, how could one point out that he abides by this norm of belief? It seems that the only way one can do that is to give reasons indicating that *p* is true. Note that the norm of belief prescribes one to believe that *p* only if/just in case/when *p* is true.⁷ It tells one to ‘do something’ under certain conditions. In order to follow, or be guided by, any such conditional prescription, one must see to it whether these conditions are met. Hence, given that one believes that *p*, one should show that he does so when the antecedent holds, i.e., it is the fact that *p* is true. And the only way to make a case that *p* is true is to give reasons indicating that *p* is true. So, again, the only way one to justify his belief that *p*, in this looser sense of justification, is to give reasons for which he believes that *p*, indicating that *p* is true.

I have presented two ways through which the truth-norm puts a strict standard on what can count as an epistemic reason to take a doxastic attitude towards a particular proposition. Firstly, it follows from Shah’s account that a piece of consideration can be an epistemic reason to take a particular doxastic attitude if and only if this consideration bears on the propositional content of the corresponding attitude. Namely, a consideration can be an epistemic reason to believe or disbelieve a proposition if this consideration bears on whether the proposition is true. If this consideration indicates that the proposition is true, it can serve as a reason to believe it. If this consideration

⁷ I am replacing the norm of belief with a weaker form because one cannot believe all the truths, see Boghossian (2003, p. 37). As I stated before, it is not my responsibility here in this paper to reformulate the truth-norm in ways that can defend it against possible objection from the literature on belief. I use this form for the sake of the point that I am trying to make.

indicates that the proposition is false, it can serve as a reason to disbelieve it or not to believe it.

For instance, when Cooper is working on the case, he obtains some relevant evidence. He learns that Diane's fingerprints are found on the murder weapon, or she was seen around the crime scene at the time of the murder. These are considerations speaking to who the murderer is by indicating that Diane is probably the murderer. In that sense, they can serve as reasons to believe that Diane is the murderer. Cooper may later learn from a trustworthy witness that Diane was somewhere else at the time of the murder. This testimony about Diane's whereabouts is a consideration indicating that she is not likely to be the murderer. And, thus, it can serve as a reason to believe that she is not the murderer or to disbelieve that she is.

So, according to the truth-normers, only an evidential consideration bearing on the propositional content of the corresponding doxastic attitude is a reason for which one takes that attitude. Cooper can come to believe that Diane is the murderer for the considerations about Diane's fingerprints and whereabouts because they are evidence pointing to her guilt. Can we fit the evidence of unreliability into this description? Of course, it has a role in this picture. After all, it is evidence that Cooper is cognitively unreliable. It can serve as a reason why Cooper may believe that he is unreliable. However, this does not say anything new about how Cooper's belief about his own unreliability can make a difference as to what it is rational for him to believe about the case.

According to this description, then, the evidence of Cooper's unreliability cannot be an epistemic reason bearing on what he should believe about the case. In learning about his cognitive unreliability, Cooper does not acquire further evidence speaking to who the murderer is. Precisely, the consideration regarding Cooper's cognitive condition does not bear on the propositional content of a doxastic attitude towards who the murderer is. If Cooper decides not to believe Diane's guilt for whatever reason provided by the evidence of his unreliability, he would not be avoiding a false belief. Instead, he would be refraining from exhibiting some other disvalue, for it still seems that there is something wrong with holding a belief while one is unreliable.

2.1.1 *The evidence of unreliability is not a rebutting defeater*

That the evidence of unreliability does not bear on whether the target proposition is true eliminates another potential explanation that can be classified under the evidentialist views. I am referring to whether the evidence of unreliability can serve as a rebutting defeater in the *evidential defeat account* [Pollock, 1986; Pollock and Cruz, 1999]. For instance, suppose that as wandering around the crime scene, Cooper sees that the wall before him is red and comes to form a belief that the wall before him is blue. When he asks his assistant to write down what he sees, his assistant tells him that the wall is not blue but red. His assistant's testimony is evidence that the wall is red. This evidence gives Cooper a reason to believe that the wall before him is red. If his new reason outweighs the reason that he acquires from what he sees, his perceptual evidence loses its prescriptive power. So, Cooper's justification to believe that the wall is red gets defeated. Based on his new set of reasons, it is not rational for Cooper to keep believing that the wall is red. His reasons to believe that the wall is red are defeated by new reasons to believe it is blue. Such reasons/considerations/evidence are called rebutting defeaters. A rebutting defeater serves as a reason against believing a proposition by being a reason to believe some other proposition.

Can this picture be used to explain how the evidence of unreliability and the evidence pointing to Diane's guilt interact? Now note that the evidence pointing to Diane's guilt indicates that she is probably the murderer. This gives Cooper a good reason to believe that Diane is most probably the murderer. Does the evidence of his unreliability give him a reason to believe that the murderer is someone else? It seems not. When Cooper learns about the drug, he does not obtain any further evidence bearing on who the murderer is. He acquires a consideration regarding his cognitive conditions, which does not have any first-order evidential bearing on who the murderer might be. So, the evidence of unreliability cannot serve as a rebutting defeater for his belief.

2.2 *Reasons for Believing*

Now let's proceed to the second way through which the truth-norm puts an evidential constraint on epistemic reasons and see what it may teach us about how to account for the evidence of unreliability. Note that, according to the second account I

have given above, one can justify his belief that *p* by pointing out that he abides by the norm of belief in forming the belief that *p*. And the only way one can do that is to give reasons indicating that *p* is true.

So, let's apply this account to Cooper's case. Suppose that Cooper decides to withhold belief about who the murderer is. How can he justify his attitude? He must show that he abides by the norm of belief in coming to withhold belief. He should give reasons against believing that Diane or anyone else is the murderer. However, the evidence that he is unreliable is not a consideration indicating that Diane is not the murderer. It does not bear on the first-order question about who the murderer is. This shows that it seems not be an epistemic but a practical or moral reason to decide to bring it about that he does not believe that Diane is the murderer for some external aim.

For instance, suppose that Cooper aims for higher positions in the homicide bureau. He believes that solving the case would get him an early promotion. So, he decides to believe that Diane is the murderer. In this situation, Cooper does not believe that Diane is the murderer because she actually is. His reason does not bear on whether Diane is actually the murderer. It is rather a reason to bring himself to take a particular doxastic attitude that can help him in achieving some external aim, e.g., getting a promotion. In this sense, his reason is a reason to *want* to be in a particular doxastic state to achieve some aim other than the truth of the matter at hand. The evidence of unreliability can play the same role. For Cooper, "getting an early promotion" can be a reason to want to believe that Diane is the murderer because believing so promotes a valuable end for him. In parallel, "being in an unreliable cognitive condition" can be a reason to want not to believe that Diane is the murderer because believing so would exhibit a disvalue. So, to refrain from exhibiting such disvalue, the evidence of unreliability can provide a reason for Cooper to decide to bring himself to withhold belief over who the murderer is.

It seems that, according to the truth-normers, Cooper cannot justify his decision to withhold belief since he cannot give reasons counting against believing that Diane is the murderer by indicating that she is not the murderer. This supports my aim to show that the truth-normers cannot account for the distinctive epistemic—not moral or practical—significance of the evidence of unreliability. However, it would be hasty to conclude that the truth-normers have failed merely because the evidence of unreliability does not bear on the truth of the target proposition. In fact, this feature opens up the prospect that

the evidence of unreliability gives a reason against believing a proposition without providing evidence for its negation. They still have another evidentialist explanation at their disposal that can utilize this peculiarity. I am referring to the undercutting defeat. For instance, suppose that as Cooper holds his belief that the wall before him is red, his assistant draws his attention to the red light shining on the wall. Learning about the red light shining on the wall seems to give Cooper a reason against believing that the wall is red without providing any evidence that it has any other color.

2.2.1 *The evidence of unreliability is not an undercutting defeater*

There are various accounts as to how this evidence—the red light shining on the wall—provides a defeating reason in the epistemic defeat literature. I do not want to commit to any specific account, so instead, I will give a general characterization. This evidence attacks the evidential relation between the perceptual evidence that the wall appears red and the proposition that the wall is red. Learning about the red light gives a good reason to think that the perceptual evidence is not a reliable indicator of what color the wall is. This is because it is more likely to mistake a non-red wall with a red wall when the lighting conditions can be highly misleading.

Given the above characterization, one come up with the following formulation to claim that the evidence of unreliability may serve as an undercutting defeater. Firstly, suppose that Cooper has first-order evidence indicating that it is 80% probable that Diane is the murderer. Here is the first doxastic attitude that Cooper takes then:

- (a) I have evidence indicating that Diane is most probably the murderer.

Secondly, it is virtually a universal acceptance in epistemology that one is justified or rational to hold a belief based on supporting evidence. If Cooper has evidence adequately supporting a proposition, he is rationally permitted to believe that proposition. Here is the second attitude that Cooper takes then:

- (b) Diane is the murderer.

Thirdly, one can draw our attention to a similarity between an undercutting defeater and the evidence of unreliability. For instance, when Cooper learns about the red light shining on the wall, the consideration about light provides a defeating reason since it is more likely to make mistakes under such deceptive circumstances; similarly, when Cooper learns about the drug, the evidence of unreliability provides a defeating reason because it is more likely to make mistakes under unreliable cognitive conditions. Here is the third attitude Cooper has then:

- (c) I have higher-order evidence indicating that the first-order evidence pointing to Diane's guilt is not a reliable guide to hold a true belief about who the murderer is.

Lastly, one may introduce a higher-order belief that Cooper has about how he come to believe that Diane is the murderer. Or we can see this higher-order belief as highlighting the basing relation between (b) Cooper's belief and (a) his set of evidence.⁸ Here is the last attitude he has then:

- (d) I believe that Diane is the murderer because/for the reason that I have evidence indicating that she is most probably the murderer.

Cooper may be rational to take (a), (b), and (d). I will not try to object to them. However, there is a problem with (c). It relies on the claim that, in both undercutting defeat and higher-order defeat, Cooper's evidential basis loses its prescriptive force by becoming unreliable. But this claim is false. Note that an undercutting defeater attacks the objective evidential support relations between the evidence and the target proposition. For instance, when Cooper learns about the light shining on the wall, the evidential relation between his perceptual evidence and his belief that the wall is red is objectively undermined. The perceptual input that the wall appears red can no longer be used as evidence for the proposition that the wall is red. Contrarily, the evidence that Cooper is cognitively unreliable does not influence the objective evidential support

⁸ I do not plan to go into detail about what kind of basing relation this higher-order belief picks out, for it does not make a difference to my point. See Sturgeon (2014) for an account that undercutting defeat requires the support of such higher-order attitudes.

relations between the evidence pointing to Diane's guilt and the proposition that Diane is the murderer [Christensen, 2010; Schechter, 2013]. When Cooper learns about his unreliable cognitive condition, it is still objectively highly probable that Diane is the murderer. Namely, the evidence of unreliability leaves the first-order evidential relations intact. In constituting a reason against believing that Diane is the murderer, it does not impose a rational constraint on Cooper's distribution of confidence across all propositions in his total set of evidence. He does not need to lower his confidence that the evidence pointing to Diane's guilt is misleading.

Another reason why the evidence of unreliability differs from an undercutting defeater is that the former seems to defeat with a retrospective aspect (Lasonen-Aarnio, 2014). For instance, before Cooper learns about the red light shining on the wall, his belief that the wall is red is rational. He was justified to believe by relying on his perceptual evidence that the wall is red. However, he was not justified in forming a belief under unreliable cognitive conditions. Even before Cooper learns about the drug, his belief that Diane is the murderer is not justified. This is because as he was forming the belief, he was lacking the required cognitive reliability to properly respond to the evidence bearing on who the murderer is.⁹

2.3 *Taking Stock*

In this section, I presented a view holding that epistemic normativity is grounded in the constitutive norm of the concept of belief. They maintain that the truth is the constitutive norm setting the standard of correctness for belief. I proceeded to show that this factual norm of correctness entails a prescriptive norm of belief that compels one to consider only evidence in doxastic deliberation. This entailment puts an evidential constraint on epistemic reasons by excluding considerations not bearing on the truth of the target proposition as normatively irrelevant. However, it turned out that such an evidential constraint on epistemic reasons prevented the truth-normers from explaining how Cooper would be justified to withhold belief upon learning that he is cognitively unreliable. I have shown that undercutting defeat is no help either as the evidence of unreliability has some features that prevent it from serving neither as a rebutting nor as

⁹ For similar views holding that higher-order defeat is different from undercutting defeat, see Christensen (2010), Schechter (2013), DiPaolo (2018), McGrath (2021). For an opposing view, see Steup (2022).

an undercutting defeater. These features also give a clue about how it puts a normative demand on what to believe about first-order matters.

The most important thing that we learned from this section is that the evidence about how reliable Cooper is makes a difference as to what he should believe *in a higher-order way*. It defeats Cooper's justification to believe that Diane is the murderer by bearing only on *his rational standing* to the question of who the murderer is. It bears on the normative status of Cooper's belief in virtue of being a reason that his belief is formed in a rationally unacceptable manner. Regardless of who the murderer might be, the evidence of unreliability seems to serve as an independent reason for withholding belief on that matter [Lord and Sylvan, 2021]. In the next section, I proceed with a view that appreciates the importance of this lesson learned. They see that the evidence of unreliability has a prescriptive force not because it influences how likely it is that Diane is the murderer but because it bears on how likely it is that Cooper is going to hold a belief with the desired normatively positive status—true belief. However, I will show that although this view is more successful than the truth-normers, they still incur undesirable costs.



Chapter 3:

TRUTH AS A GOAL AND THE EVIDENCE OF UNRELIABILITY

I have shown that the evidence of unreliability has no effect whatsoever on the objective probability of Diane's being the murderer or who the murderer is altogether. This has opened up the prospect that it exercises its defeating force by bearing on Cooper's rational standing to the question who the murderer is. So, let me ask, accordingly, what does the evidence of Cooper's unreliability say about whether Cooper can come to hold a belief that has a positive normative status in deliberating about who the murderer is? To answer this question, consider these three possible ways that the events may turn out in the *Drugged Coffee*¹⁰:

- (1) Cooper gets affected by the drug and holds a false belief about who the murderer is.
- (2) Cooper does not get affected by the drug and holds a true belief.
- (3) Cooper gets affected by the drug but manages to hold a true belief still.

These possibilities allow several interpretations. For instance, given that the drug is potent, one may infer that Cooper is more likely to hold a false belief rather than a true belief. In other words, in this unreliable condition, Cooper is more likely to be affected by the drug and come to hold a false belief about the case. A view that focuses on this interpretation may hold that following the evidence of cognitive unreliability leads to more true beliefs, for it helps one be responsive to the risk of holding a false belief. According to such a view, although the evidence of unreliability does not change how likely it is that Diane is the murderer, learning that he has been cognitively unreliable when he started the investigation gives Cooper evidence about *how likely he is* to properly evaluate the evidence bearing on the case and thereby find out who the murderer is. So, Cooper would be justified to withhold belief about who the murderer is upon learning that he is cognitively unreliable because in withholding belief he avoids

¹⁰ There is also another possibility that Cooper holds a false belief because of another cause that is not due to the effect of the drug. As I find this possibility small and insignificant enough to be ignored, I choose to exclude it from this discussion.

the risk of holding a false belief, which serves to the ultimate epistemic goal he has, i.e., having true beliefs and avoiding false ones.

In this section, I will show that the truth-goalers are inclined to take this interpretation. However, I do not plan to adopt this approach because it is susceptible to objections leveled against consequentialist views in epistemology. Instead, I will argue that the evidence of unreliability bears on the rational standing of Cooper's belief by indicating how likely Cooper is to *know*—rather than how likely he is to have a true belief about—who the murderer is.

3.1 *Evidence as a Means*

For the truth-goalers, epistemic rationality is a part of practical rationality, i.e., it is about achieving certain goals. They hold that we, as believers, have an ultimate epistemic goal of having true beliefs and avoiding false ones. It is hard to deny that this is an epistemically valuable end. Having more true beliefs and fewer false ones about matters of interest to us is an intuitively valuable end. Hence, an epistemic agent is subjected to a set of instrumental norms that compel him to take the necessary or best means to this goal. One ought to take actions and attitudes that can help one to have true beliefs and avoid false ones. Accordingly, an attitude or an action acquires its respective value according to how well it is conducive to realizing this goal.

This normative view is called epistemic consequentialism. What is epistemic consequentialism? I can give a general definition that is similar to its counterparts in other normative domains, e.g., ethics. According to the consequentialist approach in epistemology, we can identify epistemic goods/values and explain normative facts in terms of their conduciveness to these goods/values. There are various forms of epistemic consequentialism as they may disagree about what epistemic goods/values are and what relation this conduciveness consists in. Below, I will object to a certain form of epistemic consequentialism, which is called accuracy-first program, by elaborating on Schoenfield's (2015) view on the normative significance of higher-order evidence. But, before that, let me further elaborate on how this instrumentalist/consequentialist normative picture responds to Cooper's situation.

So, according to this normative picture, when Cooper is to decide what to do, he must assess which action is better option to achieve his goal. That is,

- (I) If Cooper aims to achieve 'X',
- (II) And if he knows that Q-ing is a reliable way to achieve 'X',
- (III) He should (is required by instrumental rationality) Q.

Assuming that Q-ing is practically the only reliable way for Cooper on this occasion to achieve 'X', ignoring Q-ing would be an irrational thing to do. I believe that a similar reasoning can be applied to the *Drugged Coffee*,

- (Ia) Cooper aims to have true beliefs and avoid false beliefs.
- (IIa) Withholding belief upon learning that he is cognitively unreliable helps Cooper to avoid holding presumably a false belief, which, in turn, serves the ultimate epistemic goal.
- (IIIa) Accommodating the evidence of unreliability into his doxastic deliberation is a reliable way to ensure that he holds true beliefs and avoid false ones.
- (IVa) Cooper should consider the evidence of his unreliability, indicating that he is at a high risk of holding a false belief, and withhold belief to avoid a false belief.

According to this line of reasoning, it seems that the evidence of unreliability puts a normative demand on what to believe as an indication of how likely Cooper is to achieve his ultimate epistemic goal. To elaborate on this point, I can use Schoenfield's (2015) accuracy-first account of higher-order evidence as an example that can be classified under the truth-goal views, i.e., with the relevant instrumentalist and consequentialist inclinations. Here is a quotation that can reveal her goal-directed approach:

One might take a similar stance when one considers the expected value of various epistemic plans - plans about how to form beliefs. You might be more

interested in the expected accuracy of the beliefs you would form as a result of having made a certain epistemic plan than the expected accuracy of the beliefs that you would form by conforming to a certain epistemic plan. If so, you might think of the rational plan as the epistemic plan that is best to make, rather than the epistemic plan that is best. (p. 651)

Schoenfield holds that when one deliberates about whether to believe a proposition, he should seek some epistemic guidance and plans to have the best outcome. One should be concerned with the outcomes because, for Schoenfield, epistemic rationality is about achieving certain goals, i.e., accuracy in credal states. An epistemic plan, then, is evaluated according to how well it serves one to have higher overall accuracy. For instance, suppose that we are programming a robot that is unsusceptible to any kind of cognitive malfunctioning. In programming this robot, we would not be concerned about whether the robot can properly follow the plans to form credence about the world. After all, it is perfectly capable of following the plans without any mistakes. So, we would focus on how good the plan is, i.e., how well it serves the robot to have higher overall accuracy. However, this approach does not apply to imperfect agents like us. As we are vulnerable to a wide range of cognitive impairments, we should be concerned instead with the plan's viability. Namely, we should make plans to compensate for our potential drawbacks.

According to Schoenfield's view, following the first-order evidence is ideally the best plan to maximize expected accuracy. It is the best plan to make for an ideal agent who is unsusceptible to any malfunctioning and mistakes. Turning back to the robot example, if we want the robot to have higher overall accuracy in taking doxastic attitudes about the world, the robot will be more successful if it follows the evidence bearing on the propositional content of the corresponding attitudes. The robot should ignore any evidence that it is unreliable because any negative evidence regarding its cognitive functioning would be misleading. If the robot were to consider such evidence of unreliability, it would have lower accuracy. However, there is one problem with this plan. Although it is ideally the best plan to follow, it is not the best plan to make for fallible agents. We cannot expect a cognitively imperfect epistemic agent to follow that plan flawlessly because he is more prone to mistakes than usual due to various occasional impairments in his reasoning.

Having said these, consider Cooper's situation again. If Cooper had a perfect cognitive structure—like that of a robot—he would be rational to consider only the evidence about how likely it is that Diane is the murderer. But he does not have such cognitive capacities and is vulnerable to distorting effects that a drug can cause. So, Cooper should consider his imperfect nature when weighing his total evidence. He should not ignore the evidence of unreliability, which speaks to how likely he is to respond to the first-order evidence without making any mistake. Namely, the evidence that he is cognitively unreliable speaks to how likely Cooper is to follow the best plan for ideal epistemic agents. Learning that he is not expected to follow the best plan to follow without any shortcomings, he should abide by the norm governing the evidence of his unreliability, which is the best plan for him to make.

3.2 *Some Challenges Against the Truth-Goalers*

As I stated before, although the truth-goalers has a way to account for the evidence of unreliability, they are susceptible to trade-off objections raised against the consequentialist views in epistemology.¹¹ What is a trade-off objection? Here is a simple trade-off from ethics. One can save his family by sacrificing five persons that are unknown to him. If he takes the trade-off, he sacrifices the lives of five people to rescue his family [Thomson, 1985]. However, such a verdict comes out as intuitively problematic. Similarly, these kinds of objections have their counteracts in epistemology as well. In trade-offs in epistemology, one has the choice to sacrifice an epistemic good/value in exchange for another epistemic good/value. For instance, suppose that a divine being gives you a choice that if you believe that *p*, which is a false proposition, it will make your epistemic situations very good, i.e., you will have many justified true beliefs amounting to knowledge [Fumerton, 1995]. Should you take the trade-off? I will show below that, from a consequentialist viewpoint, you should take the trade-off, although it sounds blatantly counterintuitive.

Before proceeding to my trade-off objection, suppose that Cooper is more likely to make mistakes in reasoning when he is hungry. So, he makes a plan to have his breakfast before the workday. Note that it is a good plan for Cooper to have breakfast

¹¹ Examples of trade-off objections to epistemic consequentialism can be found in Firth (1978), Fumerton (1995, p.12), Jenkins (2007), Greaves (2013), and Berker (2013a, 2013b). For a specific objection from which I am inspired, and which is directed at Schoenfield (2015), see Horowitz (2019).

before the workday as he is more prone to mistakes than usual when he is hungry. And it is not only a good plan to conform to but also a good plan to make for an agent who expects such shortcomings. However, although this looks like a good plan, it seems odd to expect epistemology to tell you to have breakfast.

The truth-goalers try to evade this objection by saying that the relevant plans must be purely epistemic. They may say that the plans should revolve around various epistemic considerations, e.g., what evidence one has, which inferential rules are applicable, and so on. However, this move does not suffice to rescue them from the challenge of explaining why the evidence of unreliability is governed by a distinctively epistemic norm. Consider the following case where Cooper makes a purely epistemic plan:

Crypto Coop

Suppose that Cooper has a cognitive predicament that his anxiety over financial matters weakens his performance in criminal investigations. For instance, when he puts money in the crypto market, he knows that a downtrend would make him anxious and more prone to mistakes in reasoning about the cases. Being aware of his predicament, Cooper formulates a rational rule (CRM) that whenever he is working on a case, he should ignore any evidence regarding the current state of the market. Cooper chooses to abide by CRM and ignores all evidence bearing on the current state of the market. So, his performance in criminal investigations remains strong, resulting in more true beliefs and fewer false beliefs in his general doxastic economy.

Now, on the one hand, it seems that for Cooper to have more true beliefs about the cases that he works on, it is good to ignore any evidence bearing on the current state of the market. On the other hand, it seems bad for Cooper to ignore evidence bearing on matters that are of interest to him. What is troubling about this case is that both of his options relate to an epistemic value. It seems epistemically good for Cooper to consider his expected shortcoming and ignore the evidence about the market so that he has more true beliefs about the criminal cases. It seems epistemically bad for Cooper to ignore evidence bearing on a matter that are of interest to him.

Let me put this problem in more detail. Now, on the one hand, there is a sense, let me call it epistemic₁, in which it is good to follow CRM and ignore the evidence

bearing on the current state of the market. According to epistemic1, ignoring the evidence about the market is good because it serves one to have more true beliefs and avoid false beliefs. On the other hand, there is a sense, let me call it epistemic2, in which it is bad to follow CRM. According to epistemic2, following (CRM) is bad because it condones epistemically unacceptable behavior, e.g., ignoring the evidence about matters of interest to one. My question is: which sense is weightier or the true sense of 'epistemic'? How can we settle this conflict between these two senses?

One possible response would be evading the question. One may respond that both senses are epistemic, but, on some occasions, these senses outweigh each other. That is, sometimes the circumstances determine which sense is more important. However, this response lets practical factors play a normative role in doxastic deliberation. I cannot think of a purely epistemic measure to determine which belief—about the market or the criminal cases—has more value. Some non-epistemic factors may rank them in order of importance instead. But letting such pragmatic encroachment seems costlier than defending one sense as intrinsically more epistemic.

Another response, hence, would be defending one as the truer sense of epistemic. For instance, truth-goalers tend to defend epistemic1 as the truer sense. It captures the sense in which it is good to take actions that helps one achieve the ultimate epistemic goal, i.e., ignoring the evidence about the market helps Cooper to avoid false beliefs. However, epistemic2 seems to capture an intuitive sense in which it is wrong, epistemically, to ignore evidence about a matter that is of interest to us. So, what do these two opposing groups have except their raw intuition to decide? Their respective intuitions regarding 'epistemic' are no reliable guide to settle the matter, let alone such intuitions being a final judge. So, the truth-goalers would still have a problem in giving a plausible explanation why following CRM is good in a distinctive/true epistemic sense.

I have shown that the truth-normers and the truth-goalers fail to account for how and why the evidence of unreliability makes a difference as to what it is rational to believe about the case. Their failure leads to a dilemma between two norms. On the one hand, according to the truth-norm, Cooper should believe that Diane is the murderer, for there is nothing wrong with believing a proposition for which one has evidence indicating its truth. If Cooper decides to ignore the evidence pointing to Diane's guilt, he exhibits an epistemically irrational behavior. On the other hand, according to the

instrumental norm created by the truth-goal, Cooper should consider his unreliability and withhold belief about who the murderer is. This is because following the evidence of unreliability serves Cooper best to have more true beliefs about the criminal cases while avoiding a false one in the current case. How can we resolve this dilemma? Which norm should Cooper abide by? Is there a scale to compare the normative relevance of these norms? In what sense could there be an objective measure as to which norm is epistemically weightier? If there is such a measure or a scale, what is it? There is no compelling answer to these questions. All Cooper can do is decide whether he should take the trade-off: whether to sacrifice the value/good of following the evidence pointing to Diane's guilt in exchange for the value/good of following the evidence of his unreliability? Should he prioritize having a true belief or avoiding a false one?

It seems that the prospects for both the truth-normers and the truth-goalers look bleak because neither has a decisive answer to these questions. Both may have arguments to motivate their views independent of the literature on the evidence of unreliability. However, I do not want to get caught in the tangle of this dialectical impasse. Instead, I will take 'knowledge' as the norm that sets the standard of justification for belief. I will embrace a line of knowledge-first theories of justification requiring 'being in a position to know' for justification. I will also buttress my account with anti-luck intuitions regarding epistemic justification and knowledge.

3.3 *Taking Stock Again: Some Explanatory Desiderata*

I have uncovered some desiderata to adequately explain why the evidence of unreliability makes a difference as to what it is rational to believe. Below is a list that can guide and constrain me as I present my proposal. Here are two desiderata regarding the nature of the evidence of unreliability:

- Desideratum 1. Since the evidence of unreliability does not bear on the propositional content of the corresponding first-order attitude, it provides a reason against believing the proposition without indicating that the proposition is false. For instance, when Cooper learns about the drug, he does not acquire any further evidence about who the murderer is.

Desideratum 2. Since the evidence of unreliability does not influence the objective probability that the target proposition is true, it leaves the evidential support relations on the first-order level intact. For instance, When Cooper learns that he is cognitively unreliable, he does not have to think that the evidence indicating that Diane is the murderer no longer points to her guilt. He can still have the evidence that Diane is the murderer, but he should not form a belief upon it.

Here are some desiderata regarding an account avoiding the problems that the truth-normers and the truth-goalers face:

Desideratum 3. A proper account can show that the evidence of unreliability acquires its prescriptive power from a distinctive epistemic norm without getting caught in the dialectical impasse between consequentialist and non-consequentialist views. That is, a proper account should specify a norm as setting the standard of justification that raises no doubt about its distinctive epistemic nature.

Desideratum 4. A proper account should explicitly put the disvalue that Cooper would exhibit if he chose to ignore his unreliability in deliberating about what to believe about the case.

Desideratum 5. A proper account does not lead to epistemic dilemmas, i.e., it can show how the evidence of unreliability overrides the prescriptive power of the evidence bearing on the case, or how these pieces of evidence do not make conflicting recommendations.



Chapter 4:

MY PROPOSAL: KNOWLEDGE NORM AND THE EVIDENCE OF UNRELIABILITY

So far, I have evaluated two potential explanations for why the evidence of unreliability seems to impose a substantive rational constraint on what Cooper should believe. I picked views that take truth as the ultimate source of normativity for belief. Firstly, I have shown that views taking truth as the norm setting the standard of correctness for belief exclude the evidence of unreliability out of the set of epistemic reasons that can put a normative demand on what to believe in first-order matters. According to the truth-normers, if Cooper decides to withhold belief upon learning that he is cognitively unreliable, he uses a reason that has no normative relevance to what it is rational to believe about the case. Secondly, I have shown that the views that take truth as the goal of belief can explain how the evidence of unreliability bears on Cooper's rational standing to the question of who the murderer is in a higher-order way. For the truth-goalers, the evidence of unreliability puts a rational constraint on what Cooper should believe about the case by bearing on how likely he is to hold a true or false belief about the matter. However, this view sanctions an intuitively problematic trade-off between the epistemic goodness of following reliable evidence about a matter of interest to Cooper and the overall goodness of his doxastic economy.

I plan to propose a better explanation not committing us to a problem about weighing the risk of having a false belief against the possibility of holding a true belief. According to my proposal, Cooper should not be worried about the risk of violating the truth-norm, nor should he be worried about the high risk of forming false beliefs. What he should be concerned is that he is at the high risk of failing to *know*. I argue that learning about his cognitive unreliability has a rational influence on what Cooper should believe by indicating that he is not likely to hold a belief about who the murderer is that would qualify as knowledge. Namely, the evidence of unreliability shows that Cooper is not in a position to know, which defeats his justification to hold a belief about the case.¹²

¹² Dutant and Littlejohn (2021) also suggest that the evidence of unreliability “could constitute evidence that the subject is not in a position to know (p. 238)”. To some extent, thus, my proposal is indebted to their account. However, my proposal is far from being exegetical of their theory of defeat. I

4.1 *Knowledge-Norm of Belief*

There is a growing number of epistemologists, let me call them knowledge-normers, who holds that ‘knowledge’ sets a standard of correctness for belief [Williamson, 2000; Sutton, 2007; Smithies, 2012; Gibbons, 2013; Engel, 2013; Littlejohn, 2013, 2017; Rosenkranz, 2018; Silva, 2017b; Simion, 2019; Kelp, 2016, 2017]. Knowledge-normers tend to conceive epistemic norms as concerned with knowledge and not merely with the truth. They accept the following norm:

Knowledge-norm of Correctness (K-Norm): A belief is correct if and only if it qualifies as knowledge.

According to the knowledge-normers, it does not suffice for a belief to be true; it must also meet the conditions qualifying it as knowledge. K-Norm puts a factive standard that excludes mere belief as normatively incorrect, i.e., only beliefs that qualify as knowledge can comply with it.

Note that K-Norm is a factive norm that is mainly about what things ought to be. In its current form, it is not an apt guide for how one should go about forming beliefs. To comply with this norm, an epistemic agent needs a prescriptive norm imposing a specific code of conduct. Such a norm would be:

Knowledge-norm of Belief (K-Rule): One should believe that *p* if and only if one knows that *p*.

To comply with the K-Norm, an epistemic agent must abide by the K-Rule. He should believe only what he knows. If his belief does not qualify as knowledge, he should avoid it not to violate any norms.

However, it is not much help to advise someone to believe only what he knows. It would be odd to say, “Earn more money!” to someone who seeks financial advice. One needs some derivative rules that can guide him to have beliefs when they amount to knowledge. There are various suggestions in the literature as to what such norms may

elaborate on the specifics about what prevents one to be in a position to know the target proposition upon learning about his cognitive unreliability.

be, and I do not intend to suggest a new one.¹³ But we can reasonably expect an epistemic agent to be sensitive to the factors bearing on whether his belief qualifies as knowledge. Borrowing a widely-held formulation from the literature [Bird, 2007; Ichikawa, 2014; Rosenkranz, 2018], I argue that an epistemic agent must be responsive to whether he is in a position to know the target proposition.

Knowledge-norm of Justification (K-Justification): One has a justification to believe that *p* if and only if one is in a position to know that *p*.

This rule is supposed to regulate the proper management of the risk of failing to comply with the K-Norm. However, what does it take for Cooper to be in a position to know who the murderer is? A good answer requires a characterization of what it takes to be in a position to know that *p*. There may be several characterizations as to what it takes to be in a position to know the target proposition. Some may postulate doxastic conditions, e.g., knowing that *p* in an acceptable way - whatever that may be. I do not claim that those conditions with their respective characterizations are false. However, I postulate an intuitive non-doxastic requirement that ‘one is in a position to know that *p* if were one to believe that *p*, he would know *p*.’ This non-doxastic condition for being in a position to know requires a non-accidental connection between believing a proposition and its truth. Also note that my formulation of ‘being in a position to know’ is about propositional justification. When Cooper is cognitively unreliable, he has no justification for holding a belief about who the murderer is regardless of whether he comes to hold a belief.

Based on this formulation, I argue that Cooper does not have a justification to hold a belief about who the murderer is in the *Drugged Coffee*. As I stated above, one prominent factor that bears on whether Cooper is in a position to know who the murderer is the condition that he has a non-accidental relation to his belief given it turns out to be true. He needs not to consciously check whether his belief meets all the standards and conditions for knowledge. But, still, he should be responsive to any consideration bearing whether these conditions are met. Learning that he has been

¹³ See Silva (2017b) for a survey of suggestions as to how epistemic justification can be explained in terms of knowledge. I do not plan to delve into how a suitable formulation may be, for it would not cause substantial changes to my proposal.

cognitively unreliable makes it salient that an essential condition will be unfulfilled. This is because when Cooper comes to hold a belief, an element of luck plays a role in his deliberation, rendering his cognitive success merely accidental.¹⁴

4.2 *Being in a Position to Know*

Before proceeding, let me make a brief remark regarding why luck prevents Cooper from being positioned to know. Virtually all epistemologists accept that a belief must not be true as a matter of luck if it is to count as knowledge. There must be a condition that distinguishes knowledge from mere accurate guesses. Knowing is a genuine cognitive achievement, and genuine achievements cannot be constituted as a matter of luck.¹⁵ So, if Cooper is in a situation where he will be lucky to hold a true belief, he is not positioned to acquire knowledge. He should see to it, then, that his belief would not be true by mere luck. Let me call this anti-luck requirement.

Keeping the anti-luck requirement in mind, reconsider the three possible ways that the events may turn out in the *Drugged Coffee*:

- (1) Cooper gets affected by the drug and holds a false belief about who the murderer is.
- (2) Cooper does not get affected by the drug and holds a true belief.
- (3) Cooper gets affected by the drug but manages to hold a true belief still.

As I stated in the previous section, these possibilities allow several interpretations. I have shown that the interpretations focusing on the possibility that Cooper is more likely to hold a false belief rather than a true belief are susceptible to objections against the consequentialist views in epistemology. I have also shown that they lead to dilemmas and fail to explicitly state from which epistemic norm the evidence of

¹⁴ Steglich-Petersen (2020) makes a similar point when he writes on whether luck excludes knowledge or certainty. However, he only remarks that “the knowledge account of luck must side with those that think that higher-order evidence can defeat knowledge” (p. 2395). I delve into exactly how an element of luck—and what kind of luck—plays a role in a deliberation where the agent is under the influence of a reason-distorting drug.

¹⁵ See Pritchard (2005), Riggs (2007), Coffman (2007), and many others for a detailed account of why knowledge is incompatible with the lucky true belief.

unreliability acquires its prescriptive force. I plan to focus on another potential interpretation bearing on whether Cooper is in a position to know who the murderer is in his current condition. According to my interpretation, Cooper is not in a position to know who the murderer is because there is no possible way for Cooper to have a true belief about the matter that would not be true as a matter of luck. I plan to support this claim by showing that in both (2) and (3), an element of luck plays in his reasoning, which prevents his cognitive success from counting as a genuine achievement—knowledge.

4.2.1 *Cooper is lucky to have a true belief in (2)*

Cooper's unreliable cognitive condition prevents him from securing some way of coming to form a belief about the case such that were he to use it, he would know who the murderer is. Consider a concrete case. Suppose that Cooper is at the shooting range honing his marksmanship and is under the influence of a drug impairing his hand-eye coordination. Is he in a position to hit bullseye? Given that hitting bullseye is a genuine achievement, it would be odd to claim that Cooper has been in a position to hit bullseye as he did so. Cooper seems to hit bullseye as a matter of luck, which undermines the achievement as merely accidental. Hence, we can say that Cooper has never been in a position to hit bullseye, for were he to shoot, he would be lucky to hit bullseye.

We can form an analogy with this case and Cooper's position in (2). Note that Cooper does not get affected by the drug and comes to hold a true belief without making any mistakes. Given that the drug is highly effective in distorting his cognitive abilities, it is highly probable that he will make a mistake in evaluating the evidence bearing on who the murderer is. However, Cooper turns out to be one of the very few who does not get affected by the drug and properly goes through his reasoning. So, it seems hard to deny that an element of luck plays a role in Cooper's unlikely success.

The moral: Cooper's belief could have easily been false. When he is highly unreliable, it is objectively not likely that he holds a true belief without making any mistakes in his reasoning. In the nearby modal environment, given how he formed his belief (without considering his condition), Cooper makes mistakes, thereby, holds false beliefs. Precisely, in most nearby possible worlds where Cooper forms a belief about who the murderer is in the same way as he does in the actual world, his belief would be

false. The kind of luck that is in play in such situations is called ‘environmental luck’ in the anti-luck epistemology literature.

4.2.2 *Cooper is lucky to have a true belief in (3)*

In (3), Cooper gets affected by the drug but manages to hold a true belief still. Can we say that he knew that Diane is the murderer? I argue that we cannot. To illustrate, consider a case where Cooper is honing his marksmanship at the shooting range. He pulls the trigger, and an electromagnetic wave deviates the bullet from its trajectory, but another wave reorients it back to its course, resulting in the bullet hitting the target bullseye. Was he lucky in exercising a successful shot? It is hard to deny that Cooper was lucky to hit the target because the success cannot be attributable to the manifestation of his shooting skills.

By analogy, Cooper makes one mistake that deviates him away from the correct verdict about the case but then makes another one that brings him back to hold a true belief about who the murderer is. When the mistakes in his reasoning inadvertently lead him to hold a true belief, Cooper is lucky to hold a true belief, for this cognitive success cannot be attributable to the manifestation of his full cognitive agency (Pritchard, 2005; Riggs, 2007). In this sense, the evidence of unreliability gives him a good reason to think that he would be lucky to hold a true belief.

An objector may say that Cooper can anticipate the waves and shoot by considering the possible impact they may have on his shot. If Cooper anticipates the possible effects of the waves and shoots accordingly, and the bullet hits the bullseye, the success can be attributed to the manifestation of Cooper’s full agency. In this situation, the bullseye would not be a matter of mere luck but a product of competent performance, for anticipation eliminates the element of luck [Greene, 2013].

As a response, I should add one caveat to the shooting range example. Suppose that Cooper is incapable of factoring in the possible impact of the waves when he shoots. He cannot foresee all kinds of factors that may influence his shot. And, although he sometimes foresees what may influence his shot, he cannot properly anticipate how and when it may do so. This caveat highlights Cooper’s position with respect to an introspectively indiscernible cognitive impairment. He is not some ideally perfect epistemic agent who can always see his shortcomings from an objective point of view.

If anything goes wrong in his reasoning due to the effect of the drug, it will go unnoticed, for his reasoning is not reflectively transparent.

4.3 *The Evidence of Unreliability as a Defeater*

Now we can see how the evidence of unreliability puts a normative demand on what to believe about who the murderer is by indicating that a condition for justification is not met, or the subject is not in the right circumstances to have a justified belief.

According to the K-Justification, Cooper has justification to believe that Diane is the murderer if and only if she is in a position to know that she is. There are some enabling conditions for him to be positioned to know who the murderer is. And Cooper is in a position to know who the murderer is if these conditions are met. However, some factors play disabling roles. For instance, cognitive unreliability hinders one from successfully implementing the relevant procedure that brings about the desired state - knowing. It puts one in a position in which one would be merely lucky to hold true beliefs. For instance, in the *Drugged Coffee*, Cooper would be lucky to hold a true belief about who the murderer is. His unreliable cognitive condition prevents him from having a non-accidental relation to the successful doxastic attitude that he takes toward the issue. His true belief either could have easily been false, as in (2), or cannot be attributed to the full manifestation of his cognitive agency, as in (3). Hence, Cooper should always be responsive to the evidence bearing on whether these conditions are met. Learning about his cognitive unreliability is one such piece of evidence indicating that he is unable to properly carry on with the relevant procedure, i.e., he is not in a position to know who the murderer is just by considering the evidence pointing to several suspects.

So, before learning about the drug, Cooper is justified to go about the relevant procedure by following the evidence pointing to Diane's guilt and come to believe that she is the murderer. However, learning about the drug constitutes evidence that Cooper is at the high risk of failing to hold a belief that can qualify as knowledge. Namely, he acquires evidence that he is not in a position to know who the murderer is regardless of his particular belief about who it may be. Now let us reconsider his total evidence for the last time. A part of the total evidence still indicates that it is highly probable that Diane is the murderer. However, I have shown that the other part of his total evidence—

the evidence of unreliability—indicates that he is not in a position to know who the murderer is. Hence, although he has evidence for Diane's guilt, he does not have a justification to hold a belief about who the murderer is, given that he has evidence indicating that he is not in a position to know who the murderer is. This is how the evidence of unreliability overrides the prescriptive power of the evidence pointing to Diane's guilt and defeats Cooper's justification to stick to his belief.

4.3.1 *Reply to Some Objections*

Objection 1. One may ask about the cases in which Cooper is misleadingly informed about his cognitive condition. Would Cooper be justified to withhold belief if he is misinformed that he is cognitively unreliable, although he is not at a high risk of failing to know who the murderer is? Can the evidence of unreliability preserve its defeating power where there is not objectively a risk regarding how likely Cooper is to know who the murderer is?

To respond to this objection, I need to consider a distinction between what might be called subjective as opposed to objective risk.¹⁶ Suppose that Cooper wishes to evaluate how risky it is to use his firearm in fieldwork. Given the track record of prior incidents, he acknowledges the risk of endangering uninvolved persons in using his firearm. He judges that the possibility of the risk event is modally close, and hence using a firearm in fieldwork is a high-risk activity. Usually, Cooper's judgment may correspond to the facts, i.e., it is objectively risky for him to use his firearm. However, this need not always be the case. Suppose that past use has inflicted damage on the firearm, and it is not functioning properly. Objectively, then, there is no risk involved in Cooper's attempting to use his firearm in fieldwork, although he is justified to subjectively judge that there is, given that the firearm is non-functioning unbeknownst to him.

If I apply this distinction to the epistemic risk, it seems that from a subjective point of view, Cooper should judge that there is high epistemic risk involved in forming a belief under current circumstances even though, objectively, there is not. Based on what he learns from the informant, he is right to judge that he is cognitively unreliable, and hence it is highly improbable—seems impossible in the Drugged Coffee—for him to

¹⁶ For a detailed account of this distinction, see Bondy (2018). For a closely related distinction between veritic epistemic luck and reflective epistemic luck, see Pritchard (2005).

hold a belief qualifying as knowledge. For this reasons, the fact that the evidence of unreliability is misleading does not deprive it of its defeating power.

Consider a concrete case. Suppose that as Cooper is traveling the countryside, he sees a barn and comes to believe that there is a barn ahead. Later he encounters a local who falsely informs Cooper that he is in a fake barn country, where locals put up fake barn façades along the countryside. Although Cooper has been misleadingly informed about his surroundings, he *still* has an undercutting defeater indicating that his perceptual evidence is not a reliable means to determine whether there is, in fact, a barn ahead. Similarly, in the *Drugged Coffee*, one can misleadingly inform Cooper that he has been cognitively unreliable throughout the case. However, as long as he does not have a good reason to doubt the informant's credibility, e.g., the informant is lying or has deceived Cooper into thinking that he is a trustworthy person, the evidence of unreliability preserves its defeating power.

Lastly, note that my response fits to the framework I have drawn to explain how the relevant norms can guide Cooper in forming beliefs. I have stated earlier that we cannot expect Cooper to perfectly conform to (K-Norm) the factive norm of having beliefs when they qualify as knowledge. This is impossible for a fallible agent who does not have a direct access to the facts. However, we can reasonably expect Cooper to try his best to approximate (K-Rule) this ideal of believing only what one knows. And the derivative rule (K-Justification) that puts a rational constraint on what Cooper should believe does not require objective certainty; instead, it governs what one should do, given objective uncertainty. He needs not to consciously check whether his belief meets all the standards and conditions for knowledge. He is, then, held accountable only for his responsiveness and sensitivity to the considerations and factors bearing on whether he is in a position to know.

So, learning that he has been cognitively unreliable makes it salient that an essential condition will be unfulfilled. This is because when Cooper comes to hold a belief, an element of luck plays a role in his deliberation, rendering his cognitive success merely accidental. Without any reason to doubt the truth of what he learns about the possible effect of the drug, he is justified to withhold belief upon recognizing the epistemic risk of forming belief under unreliable cognitive conditions.

I can concede that there is a sense in which Cooper's attitude falls short of having a certain level of epistemically positive status. He refrains from holding a belief when he

had the justification to do so. This puts him in a somewhat bad position. I have some possible options to account for this badness. Firstly, I may resort to a different evaluative terms, for instance, I can hold that Cooper is not blameless or uncriticizable for his attitude although it does not comply with the factual norm governing how he should form beliefs. Secondly, I may prefer to use a looser sense of the term justification, e.g., I can say that Cooper is justified but in a subjective and looser sense that does not require conformity to the factual norms and objective facts. For instance, he may be excusable for his attitude although it objectively fails to conform to the factual knowledge-norm of belief [Williamson, forthcoming; Littlejohn, forthcoming].

¹⁷ However, I do not plan to adopt these options as they require rigorous elaboration exceeding the scope of this paper. Instead, I believe that the distinction between objective and subjective risk suffice to explain why Cooper would be justified to withhold belief upon acquiring undefeated evidence that he is cognitively unreliable. If this distinction does not convince the objector, the burden of argument is on her/him to respond to the following question: Why must the evidence of unreliability give more than a subjective viewpoint about one's epistemic position, e.g., how risky one's situation is in a certain situation? Why should we hold someone responsible for the things over which he does not have infallible access and control, e.g., complete conformity with the facts, objective probabilities, and the active norms?

Objection 2. According to the normative framework—knowledge-first epistemology—upon which I build my proposal, ‘evidence must be true.’ They have a non-psychological conception of evidence, i.e., evidence must not consist of non-factive mental states but of propositions corresponding to the facts. This entails that one's evidence consists of what one knows. So, considering Cooper's epistemic position in the Drugged Coffee, can we say that Cooper knows that he is cognitively unreliable?

This objection resembles the previous one in that Cooper fails to know what he takes to be evidence. Cooper could not know that he is cognitively unreliable in the previous objection because it was false. In this objection, Cooper cannot know that he is cognitively unreliable because his condition may prevent him to be in a position to know whether he is.

¹⁷ For a functionalist account, see Simion, M., Kelp, C. & Ghijsen, H. (2016).

I believe that the answer to this objection resides in how the case is told. What Cooper learns from the reliable informant is normally qualifies as an instance of testimonial knowledge. However, supposing that the drug influences him in a way that he cannot trust his capacities required for acquiring testimonial knowledge, e.g., hearing, memory, etc., Cooper is in no position to know whether he is cognitively unreliable. This is not the case in the *Drugged Coffee*. But if I were to account for a case where Cooper is incapacitated to that extent, I would exempt Cooper from epistemic evaluation. That is why his position in such a case is irrelevant to what he should do in the *Drugged Coffee*.

Objection 3. According to my objection to the instrumentalists (truth-goalers), allowing a trade-off in a normative theory is an undesirable outcome. But it seems that my proposal may also be susceptible to a similar objection. Learning that he has been cognitively unreliable, Cooper has evidence that he is not in a position to know who the murderer is. However, he also still has the evidence pointing to Diane's guilt. So, if he chooses to refrain from holding a belief to avoid violating the knowledge-norm, he ignores the evidence that can help him to come to hold a true belief about who the murderer is (or believe that Diane is the murderer when it is true that Diane is the murderer). That is, Cooper chooses to comply with the knowledge-norm at the expense of violating the truth-norm. Isn't this a kind of trade-off?¹⁸

I do not have a conclusive response to this objection, but I can emphasize a crucial difference that, I believe, will reverse the burden of proof. Let us focus on the details of the trade-off sanctioned by the truth-goalers. Note that when Cooper refrains from holding a belief about the case, he avoids the risk of holding a false belief. In refraining from holding a belief about who the murderer is, he sacrifices the value/good of holding a true belief in exchange for the value/good of avoiding a false belief. The trade-off is between the value/good of 'holding a true belief' and 'avoiding a false belief.' In §3.2, I claim that there is no compelling argument to determine whether holding a true belief or avoiding a false belief is epistemically more valuable. So, in this sense, the trade-off comes out as objectionable or undesirable.

However, when Cooper refrains from holding a belief as he is not in a position to know who the murderer is, he avoids holding a belief that would not qualify as

¹⁸ Thanks to Baç for raising this telling objection.

knowledge. Namely, in refraining from holding a belief, Cooper sacrifices the value/good of holding a true belief at the expense of avoiding a non-knowlegeable mere true belief. Contrary to the trade-off in §3.2, the exchange values have an almost universally accepted hierarchy. It is virtually a universal acceptance in epistemology that knowledge is epistemically more valuable than mere true belief. So, in this sense, this trade-off between knowledge and mere true belief does not come out as objectionable, if not permissible. My response may remain open to discussion, but I believe that it suffices to reverse the burden of argument to the objectors.





Chapter 5:

CONCLUSION

Here I wrap up this paper by showing how my proposal meets the explanatory desiderata listed in §3.3 and making some concluding remarks.

5.1 *Meeting the Explanatory Desiderata*

Now I want to show how my proposal meets the explanatory desiderata. Firstly (Desideratum 1), according to my proposal, when Cooper learns about the drug, he does not acquire further evidence bearing on the case, which speaks to the question who the murderer is. Secondly (Desideratum 2), when Cooper has the evidence that he is unreliable, the evidence of his unreliability does not influence the objective evidential support relations between the evidence pointing to Diane's guilt and the proposition that Diane is the murderer.

Thirdly (Desideratum 3), it is almost a universal acceptance in epistemology that true belief has an epistemic value. And virtually all epistemologists, including the ones I grouped as the truth-normers and the truth-goalers, accept that knowledge is no less than true belief if it is not more. If having a true belief is an epistemically valuable state, it would be odd to deny that it is also epistemically valuable to hold a true belief that is more than mere guesswork. This is why I believe that the norm that prescribes holding beliefs that qualify as knowledge promotes a distinctive epistemic value.

Fourthly (Desideratum 4), if Cooper ignores the evidence that he is highly likely to hold a true belief as a matter of luck, he remains unresponsive to the high risk of holding a belief that does not qualify as knowledge. This is blatantly irresponsible behavior for an epistemic agent who aims to abide by the norm of believing only what he knows. Such irresponsibility highlights the epistemic disvalue that he would exhibit in excluding the evidence of unreliability as he deliberates about what to believe about the case. Cooper does not consciously do his best to ensure that he conforms to the knowledge-norm, i.e., he does not care whether his doxastic attitude is in a normatively correct epistemic standing.

Lastly (Desideratum 5), my proposal does not lead to an epistemic dilemma. Note that, according to my proposal, when Cooper deliberates about what to believe about the case, he is not merely aiming for a true belief or ultimately acting in ways that comply

with the norm ‘believe that p just in case p is true. Instead, the knowledge-norm of belief governs his deliberation, i.e., he aims to hold a belief if and only if the belief qualifies as knowledge. It does not suffice for Cooper’s belief merely to be true for such a norm. He must know what he believes.

Keeping these in mind, let us see whether Cooper’s evidential situation gives rise to a genuine epistemic dilemma. An epistemic dilemma is an evidential situation in which the subject cannot take an attitude or an action without violating some rational ideal. One faces an epistemic dilemma when his total evidence makes rationally incompatible recommendations. Reconsider Cooper’s total evidence again then. On the one hand, he has evidence indicating that Diane is probably the murderer. This set of evidence permits him to believe that she is the murderer. On the other hand, he has the evidence that he is not in a position to know who the murderer is. This piece of evidence does not permit him to hold a belief about who the murderer may be. It seems that Cooper is caught in an epistemic dilemma where he is both permitted and not permitted to believe that Diane is the murderer.

However, this bind is nothing but a resolvable conflict when one pays attention to the overarching norm governing what Cooper should do in deliberating about what to believe. According to the knowledge-norm, his ultimate aim is to hold beliefs when they qualify as knowledge. In learning about the drug, he acquires evidence that he cannot know who the murderer is regardless of the evidence speaking to who the murderer is. The presence of the evidence indicating that it is highly probable that Diane is the murderer does not rescue Cooper from his situation in which he cannot hold a true belief that would not be a matter of luck. But the evidence of his unreliability has nothing to do with who the murderer is; it indicates that Cooper cannot abide by the knowledge-norm of belief, i.e., in his current cognitive condition, he is incapable of holding a belief when it qualifies as knowledge. In that sense, the evidence of unreliability overrides the prescriptive power of the evidence pointing to Diane’s guilt.

5.2 *Summing Up*

The *Drugged Coffee* stages a scenario where Cooper should not believe that Diane is the murderer. Why does he seem unjustified to believe that Diane is the murderer? I began by noting that this is not an easy question. However, focusing on his evidential

situation provided a starting point. Cooper has two sets of evidence: (1) the evidence indicating that Diane is most probably the murderer, (2) the evidence that Cooper has been cognitively unreliable throughout his deliberation as to who the murderer is. So, a good explanation resides in how these two sets of evidence interact.

I have shown that some epistemologists, who take truth as the norm that sets the standards of justification for belief, fail to give an explanation because, due to their evidentialist constraint on epistemic reasons, the evidence of unreliability becomes normatively irrelevant. Their view entails that only evidential considerations can serve as an epistemic reason to take a doxastic attitude. But learning about his cognitive unreliability does not constitute evidence bearing on who the murderer is. That is why the evidence of unreliability cannot serve as a reason for Cooper to take a particular doxastic attitude towards who the murderer may be.

I continued by considering another way to integrate truth into a normative theory of epistemic rationality. Some epistemologists—the truth-goalers—take epistemic rationality as a part of instrumental rationality and hold that it is about achieving certain goals and ends. According to their view, learning about his cognitive condition can provide a reason that has a higher-order bearing on Cooper's epistemic position, i.e., how likely he is to achieve the ultimate epistemic goal of having true beliefs and avoiding false ones. The evidence of unreliability can show Cooper that he is at a high risk of holding a false belief, which falls short of promoting this ultimate epistemic end. In that sense, it can serve as a reason for refraining from holding a belief about who the murderer is. However, I have shown that this suggestion is susceptible to trade-off objections raised against the consequentialist and instrumentalist views in epistemology and leads to an epistemic dilemma, which causes a critical obstacle to an explanation for a systematic kind of epistemic defeat.

Although this suggestion brings out some undesirable costs, they are on to something about how the evidence of unreliability has a higher-order bearing on Cooper's epistemic position. The truth-goalers focus on how likely it is for Cooper to hold a true or false belief. However, learning about his cognitive unreliability is, instead, about how likely Cooper is to hold a non-accidental true belief. In other words, the evidence of unreliability speaks to how likely Cooper is to know who the murderer is. Learning about his unreliable cognitive condition defeats Cooper's justification to believe that Diane is the murderer, for it constitutes evidence that he is not in a position

to know who the murderer is, which is required for justifiably holding a belief about who the murderer may be, including a belief that Diane is.

This shift from ‘truth’ to ‘knowledge’ shows that higher-order defeat carves out another tier of epistemic value other than mere true belief. But note that even though the truth-centered views turned out to be incapable of properly accounting for higher-order defeat, I do not commit myself to a greater claim that they are untenable in explaining any other epistemic phenomena. However, I have shown in this paper that those who explain justification in terms of knowledge—or who take knowledge as the norm which sets the standard of justification for belief—get the edge on accounting for an essential kind of our fallibility, which also sheds light on the rational underpinnings of higher-order evidence and defeat.





BIBLIOGRAPHY

- [Adler, 2002] Adler, J. E. (2003). *Belief's Own Ethics*. Cambridge, MA: MIT Press.
- [Berker, 2013a] Berker, S. (2013). Epistemic Teleology and the Separateness of Propositions. *The Philosophical Review*, 122: 337-393.
- [Berker, 2013b] Berker, S. (2013). The Rejection of Epistemic Consequentialism. *Philosophical Issues*, 23: 363-387.
- [Bird, 2007] Bird, A. (2007). Justified Judging. *Philosophy and Phenomenological Research*, 74(1): 81-110.
- [Boghossian, 2003] Boghossian, P. A. (2003). The Normativity of Content. *Philosophical Issues*, 13: 31-45.
- [Bondy, 2018] Bondy, P. (2018). Propositional epistemic luck, epistemic risk, and epistemic justification. *Synthese*, 195: 3811-3820
- [Christensen, 2007] Christensen, D. (2007). Epistemology of Disagreement: The Good News. *The Philosophical Review*, 116(2): 187-217.
- [Christensen, 2010] Christensen, D. (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research*, 81(1): 185-215.
- [Christensen, 2011] Christensen, D. (2011). Disagreement, question-begging, and epistemic self-criticism. *Philosopher's Imprint*, 11: 1-22.
- [Christensen, 2013] Christensen, D. (2013). Epistemic Modesty Defended. In D. Christensen and J. Lackey (Eds.), *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.
- [Christensen and Lackey, 2013] Christensen, D. and Lackey, J. (2013). *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.
- [Christensen, 2016] Christensen, D. (2016). Conciliation, uniqueness, and rational toxicity. *Noûs*, 50: 584-603.

- [Coates, 2012] Coates, A. (2012). Rational epistemic akrasia. *American Philosophical Quarterly*, 49: 113-124.
- [Coffman, 2007] Coffman, E. J. (2007). Thinking about luck. *Synthese*, 158: 385-398.
- [Cowie, 2014] Cowie, C. (2014). In Defence of Instrumentalism about Epistemic Normativity. *Synthese*, 191: 4003-4017.
- [DiPaolo, 2018] DiPaolo, J. (2018). Higher-Order Defeat is Object-Independent. *Pacific Philosophical Quarterly*, 99: 248-269.
- [Dutant and Littlejohn, 2021] Dutant, J. & Littlejohn, C. (2021). Defeaters as Indicators of Ignorance. In J. Brown & M. Simion (Eds.), *Reason, Justification, and Defeat*. Oxford: Oxford University Press.
- [Elga, 2007] Elga, A. (2007). Reflection and Disagreement. *Noûs*, 41: 478-502.
- [Elga, 2010] Elga, A. (2010). How to disagree about how to disagree. In T. Warfield & R. Feldman (Eds.), *Disagreement*. Oxford: Oxford University Press.
- [Engel, 2013] Engel, P. (2013). In Defence of Normativism about the Aim of Belief. In T. Chan (Eds.), *The Aim of Belief*. Oxford: Oxford University Press.
- [Feldman, 2005] Feldman, R. (2005). Respecting the Evidence. *Philosophical Perspectives*, 19(1): 95-119.
- [Feldman, 2006] Feldman, R. (2006). Epistemological Puzzles about Disagreement. In S. Heatherington (Eds.), *Epistemology Futures*. Oxford: Oxford University Press.
- [Feldman, 2007] Feldman, R. (2007). Reasonable Religious Disagreements. In L. M. Antony (Eds.), *Philosophers Without Gods: Meditations on Atheism and the Secular Life*. New York: Oxford University Press.
- [Feldman, 2009] Feldman, R. (2009). Evidentialism, Higher-Order Evidence, and Disagreement. *Episteme*, 6(3): 294-312.
- [Feldman and Warfield, 2013] Feldman, R. and Warfield, T. (2013). *Disagreement*. Oxford: Oxford University Press.

- [Firth, 1978] Firth, R. (1978). Are Epistemic Concepts Reducible to Ethical Concepts? In A. Goldman & J. Kim (Eds.), *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt*. Dordrecht: Kluwer Academic Publishers.
- [Foley, 1987] Foley, R. (1987). *Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press.
- [Fumerton, 1995] Fumerton, R. (1995). *Metaepistemology and Skepticism*. Lanham, MD: Rowman & Littlefield.
- [Gibbard, 2005] Gibbard, A. (2005). Truth and Correct Belief. *Philosophical Issues*, 15(1): 338-350.
- [Gibbons, 2013] Gibbons, J. (2013). *The Norm of Belief*. Oxford University Press.
- [Greaves, 2013] Greaves, H. (2013). Epistemic Decision Theory. *Mind*, 122: 915–952.
- [Greene, 2013] Greene, P. (2013). When is a belief true because of luck? *Philosophical Quarterly*, 63(252): 465-475.
- [Hazlett, 2012] Hazlett, A. (2012). Higher-Order Epistemic Attitudes and Intellectual Humility, *Episteme*, 9(3): 205-223.
- [Hieronymi, 2005] Hieronymi, P. (2005). The Wrong Kind of Reason. *Journal of Philosophy*, 102(9): 437-457.
- [Horowitz, 2014] Horowitz, S. (2014). Epistemic Akrasia. *Noûs*, 48: 718–744.
- [Horowitz, 2019] Horowitz, S. (2019). Predictably Misleading Evidence. In Mattias Skipper and Asbjørn Steglich-Petersen (Eds.), *Higher-Order Evidence: New Essays*. Oxford University Press.
- [Ichikawa, 2014] Ichikawa, J. J. (2014). Justification Is Potential Knowledge. *Canadian Journal of Philosophy*, 44(2): 184–206.
- [Jenkins, 2007] Jenkins, C. S. (2007). Entitlement and Rationality. *Synthese*, 157: 25–45.

- [Kelp, Simion, and Ghijsen, 2016] Kelp, C., M. Simion, H. Ghijsen. (2016). Norms of Belief. *Philosophical Issues*, 16: 374-92.
- [Kelp, 2017] Kelp, C. 2017. Knowledge First Virtue Epistemology. In A. Carter, E. Gordon, and B. Jarvis (Eds.), *Knowledge First: Approaches in Epistemology and Mind*. Oxford: Oxford University Press.
- [Kelly, 2005] Kelly, T. (2005). The Epistemic Significance of Disagreement. *Oxford Studies in Epistemology*, 1: 167-96.
- [Kelly, 2010] Kelly, T. (2010). Peer disagreement and higher order evidence. In A. I. Goldman & D. Whitcomb (Eds.), *Social epistemology: Essential readings*. Oxford: Oxford University Press.
- [Lasonen-Aarnio, 2014] Lasonen-Aarnio, M. (2014). Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research*, 88(2): 314-345.
- [Lasonen-Aarnio, 2020] Lasonen-Aarnio, M. (2020). Enkrasia or evidentialism? Learning to love mismatch. *Philosophical Studies*, 177: 597-632.
- [Littlejohn, 2013] Littlejohn, C. (2013). The Russellian Retreat. *Proceedings of the Aristotelian Society*, 113(3): 293–320.
- [Littlejohn, 2017] Littlejohn, C. (2017). How and Why Knowledge is First. In A. Carter, E. Gordon & B. Jarvis (Eds.), *Knowledge First*. Oxford: Oxford University Press.
- [Littlejohn, forthcoming] Littlejohn, C. (forthcoming). A plea for epistemic excuses. In F. Dorsch and J. Dutant (Eds.), *The New Evil Demon: New Essays on Knowledge, Rationality and Justification*. Oxford: Oxford University Press.
- [Lord and Sylvan, 2021] Lord, E. & Sylvan, K. (2021). Reasons to Suspend, Higher-Order Evidence, and Defeat. In M. Simion & J. Brown (Eds.), *Reasons, Justification, and Defeat*. Oxford University Press.
- [Lynch, 2009] Lynch, M. (2009). *Truth as One and Many*. Oxford: Oxford University Press.

- [McGrath, 2021] McGrath, M. (2021). Undercutting Defeat: When It Happens and Some Implications for Epistemology. In J. Brown and M. Simion (Eds.), *Reasons, Justification, & Defeat*. Oxford: Oxford University Press.
- [Pollock, 1986] Pollock, J. (1986). *Contemporary Theories of Knowledge*. Savage, MD: Rowman and Littlefield.
- [Pollock and Cruz, 1999] Pollock, J & Cruz, J. (1999). *Contemporary Theories of Knowledge*. Lanham, MD: Rowman and Littlefield.
- [Pritchard, 2005] Pritchard, D. (2005). *Epistemic luck*. Oxford: Oxford University Press.
- [Raz, 2011] Raz, J. (2011). *From Normativity to Responsibility*. Oxford: Oxford University Press.
- [Riggs, 2007] Riggs, W. (2007). Why Epistemologists Are So Down on their Luck. *Synthese*, 158: 329-344.
- [Rosenkranz, 2018] Rosenkranz, S. (2018). The Structure of Justification. *Mind*, 127(506): 309–338.
- [Shah, 2003] Shah, N. (2003). How Truth Governs Belief. *Philosophical Review*, 112: 447-482.
- [Shah, 2006] Shah, N. (2006). A New Argument for Evidentialism. *Philosophical Quarterly*, 56: 481-498.
- [Silva, 2017a] Silva, P. (2017). How Doxastic Justification Helps Us Solve the Problem of Misleading Higher-Order Evidence. *Pacific Philosophical Quarterly*, 98(1): 308-328.
- [Silva, 2017b] Silva, P. (2017). Knowing How to Put Knowledge First in the Theory of Justification. *Episteme*, 14(4): 393-412.
- [Simion, 2019] Simion, M. (2019). Knowledge-first functionalism. *Philosophical Issues*, (29)1: 254-267.

- [Smithies, 2012] Smithies, D. (2012). The Normative Role of Knowledge. *Noûs*, 46(2): 265-288.
- [Steglich-Petersen, 2011] Steglich-Petersen, A. (2011). How to Be a Teleologist about Epistemic Reasons. In A. Steglich-Petersen and A. Reisner (Eds.), *Reasons for Belief*. Cambridge University Press.
- [Steglich-Petersen, 2020] Steglich-Petersen, A. (2020). Does luck exclude knowledge or certainty? *Synthese*, 197(6): 2387-2397.
- [Steup, 2022] Steup, M. (2022). Defeaters, Higher-Order Evidence, and Epistemic Dilemmas. In K. McCain, S. Stapleford, and M. Steup (Eds.), *Epistemic Dilemmas: New Arguments, New Angels*. New York: Routledge.
- [Sturgeon, 2014] Sturgeon, S. (2014). Pollock on defeasible reasons. *Philosophical Studies*, 169(1): 105–18.
- [Sutton, 2007] Sutton, J. 2007. *Beyond Justification*. Cambridge: MIT Press.
- [Thomson, 1985] Thomson, J.J. (1985). The trolley problem. *The Yale Law Journal*, 94(6): 1395-1415.
- [Titelbaum, 2015] Titelbaum, M. (2015). Rationality's Fixed Point (Or: In Defence of Right Reason). *Oxford Studies in Epistemology*, 5: 253-294.
- [van Wietmarschen, 2013] van Wietmarschen, H. (2013). Peer disagreement, evidence, and well-groundedness. *Philosophical Review*, 122(3): 395–425.
- [Way, 2012] Way, J. (2012). Transmission and the Wrong Kind of Reason. *Ethics*, 122(3): 489–515.
- [Weatherson, 2019] Weatherson, B. 2019. *Normative Externalism*. Oxford: Oxford University Press.
- [Wedgwood, 2007] Wedgwood, R. (2007). *The Nature of Normativity*. New York: Oxford University Press.
- [Wedgwood, 2012] Wedgwood, R. 2012. Justified inference. *Synthese*, 189: 273–23.

- [Wedgwood, 2013] Wedgwood, Ralph. 2013. Akrasia and Uncertainty. *Organon F*, 20(4): 484–506.
- [Whiting, 2019] Whiting, D. (2019). Whither Higher-Order Evidence? In M. Skipper and A. Steglich-Petersen (Eds.), *Higher-Order Evidence: New Essays*. Oxford University Press.
- [Whiting, 2020] Whiting, D. (2020). Higher-Order Evidence. *Analysis*, 80(4): 789-807.
- [Williamson, 2000] Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford UP.
- [Williamson, 2011] Williamson, T. (2011). Improbable knowing. In T. Dougherty (Eds.), *Evidentialism and its discontents*. Oxford: Oxford University Press.
- [Williamson, 2014] Williamson, T. 2014. Very improbable knowing. *Erkenntnis*, 79: 971–99.
- [Williamson, forthcoming] Williamson, T. (forthcoming). Justification, excuses, and sceptical scenarios. In F. Dorsch and J. Dutant (Eds.), *The New Evil Demon: New Essays on Knowledge, Rationality and Justification*. Oxford: Oxford University Press.
- [Worsnip, 2018] Worsnip, A. (2018). The Conflict of Evidence and Coherence. *Philosophy and Phenomenological Research*, 96: 3-44.
- [Ye, 2020] Ye, R. (2020). Higher-order defeat and intellectual responsibility. *Synthese*, 197(12): 5435-5455.