

Epistemic Peers, Disagreement, and Biased Judgment:

Why A Good Judge of Peerhood is Hard to Find

by

Adam G. Betts

A Dissertation Presented in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

Approved April 2023 by the  
Graduate Supervisory Committee

Nathan Ballantyne, Chair  
Peter Kung  
Ben Phillips  
Nathan King

ARIZONA STATE UNIVERSITY

May 2023

## ABSTRACT

The responses to idealized cases of peer disagreement given in the peer disagreement literature are presented as though those responses ought to be applied to real-world cases of disagreement. In order to apply the advice given in the literature to actual disagreement situations, one must first confidently identify one's epistemic peers. Previous work in the literature, especially by Nathan King, suggests that one cannot confidently identify one's epistemic peers in real-world cases of disagreement because it is unlikely that any two people will ever meet the idealized conditions of peerhood in real-world disagreements. I argue that due to the unconscious judgment-altering effects of certain cognitive biases, even if one could consciously meet the idealized conditions for epistemic peerhood as they are outlined in the peer disagreement literature, one should still not be confident that one has correctly identified others as one's epistemic peers. I give examples of how cognitive biases can affect one's judgments of one's own epistemic abilities and the epistemic abilities of others, and I conclude that the peer disagreement literature's prescriptions may not be suitable for, and are perhaps deleterious to, rational real-world disagreement resolution.

To my Grandma Dot

## ACKNOWLEDGEMENTS

I am indebted to the following people who provided inspiration, guidance, and support for my work on this dissertation. Bernie Kobes, Ángel Pinillos, and Maura Priest offered helpful advice early in the developmental stages of this project. Steve Reynolds and I talked through my ideas about this project and philosophy in general at length. Not only was his feedback and guidance indispensable to forming my project, his good humor and intellectual humility kept my spirits up and provided me with a model of how to be a better thinker.

My dissertation chair, Nathan Ballantyne, deserves more praise than I can reasonably fit here. I was inspired to write this dissertation by three journal articles, all written by philosophers I had never met. One of those three articles was written by Nathan. So, when he came over to Arizona State while I was working on this dissertation, I was overjoyed. Little did I know that he would also turn out to be an energizing force for my project. I have benefited greatly from his generosity, dedication to excellence, and expertise. Thank you, Nathan.

I am grateful for the encouragement and support I received from my other committee members at Arizona State, Peter Kung and Ben Phillips. I am also grateful that Nate King, who authored one of the other articles that inspired me to write this dissertation, agreed to participate in my project as an external committee member. The title of this dissertation is a direct reference to the title of his inspirational article, and I appreciate his willingness to review my work and participate in my defense.

Throughout the entire process of writing this dissertation, my parents, siblings, and graduate school colleagues provided me with nourishing support. Above all else, I wouldn't have completed this project without the constant caring support of my wife, Abigail. She centered me when I was flailing and cheered for me each time I achieved a small victory in the battle to complete this project.

Finally, I am grateful for the financial support provided by Arizona State's School of Historical, Philosophical, and Religious Studies (SHPRS). In addition to receiving the funding offered to teaching assistants, I was also fortunate to be awarded research grants from SHPRS that were crucial for my success.

## TABLE OF CONTENTS

CHAPTER	Page
1 INTRODUCTION TO THE PROJECT.....	1
1.1: Overview of Project.....	2
2 REVIEW OF PEER DISAGREEMENT LITERATURE.....	7
2.1: Early Foundations of the Peer Disagreement Problem.....	7
2.2: Conciliationism.....	18
2.3: Steadfastness.....	27
2.4: Conciliation and Steadfast Summary.....	32
2.5: Hybrid Positions.....	33
2.6: <i>Prima Facie</i> Objections to the Main Views.....	39
2.7: Peer Disagreement Response Strategies and the Puzzle as a Trilemma.....	44
2.7.1: Setting up the Peer Disagreement Trilemma.....	45
2.7.2: Peer Disagreement Trilemma.....	49
2.8: General Strategies of the Leading Responses.....	49
2.9: Summary and Conclusion of Chapter 2.....	53
3 ANALYSIS OF EPISTEMIC PEERHOOD.....	54
3.1: Epistemic Symmetry and Disagreement-Relevant Epistemic Asymmetries..	55
3.2: Definitions of Epistemic Peers in the Literature.....	60
3.2.1: Likelihood and Dispositional Definitions of Epistemic Peers.....	77
3.3: Extant Criticisms of Epistemic Peerhood.....	83

CHAPTER	Page
3.3.1: Concerns about Epistemic Peerhood.....	83
3.3.2: Criticisms of Epistemic Peerhood: Loose vs. Strict and Degrees of Peerhood.....	88
3.4: Summary and Conclusion of Chapter 3.....	105
4 KING’S CRITICISM OF EPISTEMIC PEERHOOD.....	107
4.1: King’s Argument.....	114
4.2: Summary and Conclusion of Chapter 4.....	114
5 BOUNDED RATIONALITY AND THE HEURISTICS AND BIASES PROGRAM.....	123
5.1: Overview of Simon’s Bounded Rationality.....	124
5.2: An Overview of the Heuristics and Biases Program.....	127
5.3: Anchoring, Availability, and Representativeness Heuristics .....	131
5.3.1: Representativeness Heuristic.....	132
5.3.2: Availability Heuristic.....	137
5.3.3: Anchoring Heuristic.....	141
5.4: Cognitive Biases.....	143
5.4.1: Cognitive Biases: Evaluating Our Own Abilities.....	145
5.4.2: Cognitive Biases: Evaluating the Abilities of Others.....	149
5.4.3: Cognitive Biases: Evaluating Evidence in General.....	154

CHAPTER	Page
5.4.4: Cognitive Biases: Introspective Bias Detection Failure.....	158
5.5: The Cognitive Bias Problem.....	161
5.6: Summary and Conclusion of Chapter 5.....	162
6 RELATIVE EPISTEMIC PEER EVALUATIONS AND THE COGNITIVE BIAS PROBLEM.....	163
6.1: Relative Epistemic Status Evaluations (RESE).....	164
6.2: The Relationship Between the Evidence and Virtue Conditions in RESEs.	170
6.3: Epistemic Peer Calibration Cases.....	174
6.3.1: Peerhood Case 1: COMPUTERS.....	175
6.3.2: Peerhood Case 2: POLICY.....	178
6.3.3: Peerhood Case 3: RESTAURANT (RETOLD).....	181
6.4: Subjective Confidence Estimation Biases and Epistemic Peer Evaluations.	185
6.5: Halo and Horns Effects and Epistemic Peer Evaluations.....	192
6.6: Confirmation Bias in Epistemic Peer Evaluations with Track Records.....	196
6.7: Conclusions of Applying CBP and Specific Cognitive Biases to Epistemic Peer Evaluations.....	199
6.8: Main Argument.....	201
6.8.1: Sketch of Main Argument.....	201
6.8.2: Valid Main Argument with Comments.....	203
6.9: Summary and Conclusion of Chapter 6.....	216
7 CONCLUSION TO THE PROJECT.....	217



CHAPTER	Page
BIBLIOGRAPHY.....	221

## CHAPTER 1: INTRODUCTION TO THE PROJECT

Let's say that you and I disagree about some topic that we care about deeply. As we discuss and defend our differing points of view, we discover there is one particular claim that we disagree about. I believe that claim is true, and you believe it is false. Now what?

Since we're both reasonable people with vast libraries of information available at our fingertips, we search the internet using the keywords "rational responses to disagreement". The top results lead us to an interesting project in professional academic epistemology: the epistemic significance of peer disagreement. This seems like a promising lead. As we read on, we discover that many smart philosophers have been diligently working on this project. What we learn from their work might help us decide how to respond to our disagreement. However, we discover another disagreement in the literature: some authors argue that we can remain steadfast in our views while others insist that, *if we are epistemic peers*, then we ought to conciliate with each other by revising our disputed beliefs. So, now it is up to us to decide if we are epistemic peers, or if it is at least reasonable for us to believe we are.

In what follows, I'm going to dig deeper into the peer disagreement literature because I want to discover what it means for us to be epistemic peers and why that is important for properly responding to our disagreement. I invite you to look over my shoulder as I investigate the topic of peer disagreement and the problems that we might face when we attempt to determine if we are peers or not. After all, we need to get this

right: our disagreement is on a topic that we care about deeply, so we'll need to make sure that what these philosophers say about disagreements applies to our situation.

### **1.1: Overview of Project**

Here is the driving question in the peer disagreement literature: How ought we respond when we discover that we disagree with a so-called epistemic peer? Several responses have been proposed, and the literature has primarily focused on these two: (1) when we discover that we disagree with an epistemic peer, we ought to change our beliefs about the matter being disagreed upon in such a way that we move our beliefs “closer” to those of our disagreeing interlocutor (this response is called conciliationism), or (2) when we discover that we disagree with an epistemic peer, we are permitted, rationally, to retain the beliefs that we had prior to discovering the disagreement (this response is called steadfastness). There are other responses to the peer disagreement question, namely, Thomas Kelly's (2010) Total Evidence View and Jennifer Lackey's (2007) Justificationist View. In Chapter 2, I provide a brief review of the arguments for these positions.

I will note here and elsewhere that there are many good reviews of peer disagreement literature available (Frances 2014, and Matheson 2014, Frances and Matheson 2018, and Loughheed 2020 stand out). However, my review is not just an additional survey; I also discuss the literature's development and demonstrate that, even though the cases debated in the literature are purposefully idealized, the intent of the authors is to offer advice on how we ought to respond to real-world disagreements. If the prescriptions offered in the peer disagreement literature are intended to apply to real-

world disagreements, then my thesis, given below, finds a foothold in my review of the literature's development and purported applications. Remember, I am interested in learning how the peer disagreement project can guide responses to real-world disagreements about our cherished beliefs.

Central to the peer disagreement literature is the idea that we have—or could have—**epistemic peers**, a concept that I analyze in Chapter 3. Epistemic peers are people who have the same evidence and reasoning capabilities as us regarding some topic or, more formally, regarding the veracity of some proposition—they are in a state of epistemic symmetry with us. In contrast to the epistemic symmetry between epistemic peers proposed in the literature, we also have epistemic *inferiors* and *superiors* on certain topics or, again, on certain propositions. According to the literature, an epistemic superior is someone who has better disagreement-relevant evidence or better disagreement-relevant epistemic virtues than their inferiors—they are in epistemically *asymmetrical* positions relative to some proposition. I explain what “disagreement-relevant” means in Chapter 3. For now, when disagreement-relevant epistemic asymmetries are present among interlocutors, this state of affairs purportedly guides how we ought to respond to a disagreement—inferiors ought to defer to the opinions of their superiors, as the labels suggest. That is, epistemic asymmetries in evidence or epistemic virtues between two people explain *why* they disagree and, in the case where one person is an epistemic inferior, the normative force of disagreement demands that they revise their attitude about the disputed proposition. However, when two (or more) people lack any disagreement-relevant epistemic asymmetries, they are epistemic peers. In these cases, which are highly

idealized, we are faced with a puzzle: Is disagreement with an epistemic peer epistemically significant (i.e., should we worry that we've gotten something wrong)? And what good reasons do we have to maintain our beliefs in the face of peer disagreements, if we ever do? These are the questions motivating the literature.

However, as I demonstrate in Chapter 3, if we are going to apply the lessons from the literature to real-world disagreements, we also need to perform epistemic peer evaluations. That is, according to the conceptual framework of the literature, we must first perform an epistemic peer evaluation to determine the relative epistemic status (peer, superior, inferior) of our real-world interlocutors before we can apply the advice given in the literature. I argue that, due to psychological constraints such as the presence of cognitive biases, we won't be able to reliably perform those evaluations accurately.

My thesis in this dissertation, then, is the following: *Cognitive biases can affect our real-world epistemic peerhood evaluations in disagreements about our cherished beliefs to such an extent that we should not be confident that we can perform those evaluations in a way that satisfies the requirements for peerhood given in the peer disagreement literature.* I argue that while the peer disagreement project's idealized cases help us understand how we rationally ought to weigh the opinions of others, the conceptual framework of the peer disagreement project—especially with regard to epistemic peer evaluations—is not amenable to real-world application. So, my focus here is on epistemic peer evaluations in real-world contexts. Recall that I am trying to determine how to respond to our *actual* disagreement.

While my specific concerns about the problems with epistemic peerhood evaluations are original, I am not the first to notice the literature's lack of attention to how we can move from our intuitions about idealized cases to real-world application. My work in this dissertation is partially an extension of Nathan King's 2012 article "Disagreement: What's the Problem? *or* A Good Peer is Hard to Find". King's article is critical of the conceptual framework assumed in the literature. He argues that the necessary and sufficient conditions of epistemic peerhood are—even on a charitable interpretation of the literature's standard definitions—so obviously unlikely to obtain that we shouldn't expect to find many if any actual epistemic peers. So, King concludes that the likelihood of encountering actual epistemic peers is so miniscule that the peer disagreement framework cannot be applied to real-world disagreements. I discuss King's views further in Chapter 4.

In Chapter 5, I turn my discussion towards empirical research on cognitive biases. I start with an overview of Herbert Simon's theory of bounded rationality and discuss the benefits of an "ought implies can" approach to designing advice that will be useful in real-world situations. Next, I introduce the heuristics and biases program. I discuss Daniel Kahneman and Amos Tversky's (among others) early work on heuristics and cognitive biases, which is the focus of my attention in Chapter 5. Kahneman and Tversky's discovery of the anchoring, availability, and representativeness heuristics (Kahneman and Tversky 1982, and Gilovich, Griffin, and Kahneman 2002) laid the groundwork for the heuristics and biases program, which continues to this day. These three heuristics have been demonstrated in a multitude of replicated empirical studies and

are sometimes a source of bias. To put it roughly, a cognitive bias is a systematic error in thinking produced by, among other things, a well-functioning cognitive heuristic that sometimes outputs wrong “answers” to certain questions. However, and importantly, not only do the anchoring, availability, and representativeness heuristics often output the wrong answers to certain questions, we also cannot stop them from giving us these wrong answers. Additionally, we cannot introspectively examine the fidelity of their processes to determine when they are likely to give us reasonable answers. After my review of the heuristic and biases program, I present studies of specific biases that can—and probably furtively do—affect our judgments during epistemic peer evaluations. I close Chapter 5 by defining what I call the Cognitive Bias Problem, which I apply in Chapter 6 to the task of evaluating peerhood.

In Chapter 6, I pull together the threads of my main argument to defend my thesis. In the first part of the chapter, I discuss the literature’s requirements for performing an epistemic peer evaluation. I also discuss the relationship between the evidence and virtue conditions for peerhood. Next, I select a case from the literature that best represents a strong, paradigmatic case of peer disagreement. Then, I deploy the Cognitive Bias Problem in that paradigmatic case to test my thesis. I conclude that cognitive biases can negatively influence the accuracy of peer evaluations in disagreements over cherished beliefs. I close Chapter 6 with my main argument that demonstrates my thesis.

In Chapter 7, I conclude my project by returning to our original question about how we ought to respond to our disagreement. I contend that, despite our good intentions

to use the peer disagreement literature to help us respond to our disagreement, we will be better off looking for advice outside of the peer disagreement framework.

## **CHAPTER 2: REVIEW OF PEER DISAGREEMENT LITERATURE**

The following review of the peer disagreement literature serves three functions for my project. The first is to familiarize the reader with the purpose, positions, and arguments in the literature. The development of the problem of peer disagreement and proposed responses are presented here in (roughly) chronological order, showing how the debate evolved. The second function is to establish that philosophers contributing to the literature are concerned with how we ought to rationally respond to actual cases of disagreement. I'll show that the literature is, in part, intended to be used as normative advice for resolving real-world disagreement situations. The third function of this review is to sketch the strategies used in the literature to resolve the problem of peer disagreement. I pose these strategies as solutions to a trilemma in the final section of this chapter. But first, let's take a look at the problem under discussion.

### **2.1: Early Foundations of the Peer Disagreement Problem**

The problem of peer disagreement arises when we try to answer the following question: How ought we rationally respond when we discover that an epistemic peer disagrees with us about some proposition? Put differently, the core of inquiry in the literature asks us whether peer disagreements are epistemically significant in terms of how we should doxastically respond when we discover that an epistemic peer disagrees with us about some proposition. Below, I trace the inception and development of the problem and discuss the proposals that philosophers have made to address it. There are



several good reviews already available (e.g., Frances 2014, Matheson 2014, and Frances and Matheson 2018). However, my review is unique for two reasons. First, I am searching for evidence to support my claim that the literature is indeed intended to be used in real-world cases. Second, I provide a new way to conceptualize the puzzle<sup>1</sup> as a trilemma.

The peer disagreement project became popular through Richard Feldman (2006), Thomas Kelly (2005), and David Christensen (2007).<sup>2</sup> Richard Feldman proposed the problem of peer disagreement in his “Epistemological Puzzles about Disagreement” (2006). Feldman observed that his students were often willing to agree to disagree about certain issues, religion being a salient example, and he wondered whether agreeing to disagree was a rational response to disagreements in general. In his words:

It is a familiar fact that there is widespread and robust disagreement about many of the most prominent issues in our intellectual lives. This is quite obviously true in epistemology itself, as well as in philosophy more generally. There is similar disagreement about religious matters, many scientific topics, and many issues of

---

<sup>1</sup> I use “puzzle” and “problem” interchangeably throughout this work. For those who believe that there is an important distinction between the senses of these words, I implore you to set down your fire poker and temporarily suspend that belief as you read.

<sup>2</sup> I’m piecing together a chronology of the development of peer disagreement literature based on *publication* dates rather than from the first presentations of the problem. Feldman’s *Reasonable Religious Disagreements* was published as a chapter in Louise Antony’s (ed.) 2007 *Philosophers Without Gods: Meditations on Atheism and the Secular Life*, but Feldman had been presenting and discussing that work as early as 2004, and he eventually borrowed Kelly’s use of “epistemic peer” to employ in presentations that predated his “Epistemological Puzzles About Disagreement,” which was published in 2006. Additionally, David Christensen’s “Epistemology of Disagreement: The Good News” was in the works as early as 2004 as well, though it was not published until 2007 in *Phil Review* (see Elga 2007 for comments on development). This is worth noting because my review of the literature gives a sort of narrative of how the peer disagreement literature developed for the wider philosophical community rather than how it developed within the group of seminal authors working on the project in its nascent, pre-published stages.

public policy. In all these areas, informed and intelligent people disagree with one another. To make it more personal, on many of these issues about which you have a belief, informed and intelligent people disagree with you. The question I will raise concerns the reasonableness of maintaining your point of view in the light of such disagreements. (Feldman, 2006, p. 217)

Subsequently, Feldman notes many specific examples of disagreements from each of the domains of inquiry mentioned in the quote above; these include disagreements over law and science, politics, philosophy, and religion (2006, pp. 217-219). Already, we can see that real-world topics of debate were intended to be included in the peer disagreement project.

Feldman was concerned about the appropriate rational response to cases of disagreement among people who share **epistemic symmetry**. In these cases, one party to a disagreement believes that proposition  $p$  and the other party believes that not- $p$ . Of course, when epistemic symmetry does not obtain, disagreements aren't all that puzzling; people disagree all of the time and, as one standard response goes, the person who is epistemically disadvantaged should defer to the epistemically advantaged party. The puzzle of peer disagreement only arises when both parties to a disagreement have some sort of epistemic symmetry—spelling out what that involves is part of my project below (esp. in Chapter 3). Even though Feldman did not use the phrase “epistemic peer” in his foundational work, he describes what will eventually become a prototype for epistemic peerhood:

[C]onsider those cases in which the reasonable thing to think is that another person, every bit as sensible, serious, and careful as oneself, has reviewed the same information as oneself and has come to a contrary conclusion to one's own. And, further, one finds oneself puzzled about how that person could come to that conclusion. An honest description of the situation acknowledges its symmetry. (Feldman, 2006, p. 235)

For now, we should take note of the various requirements for epistemic symmetry that Feldman identifies. When we reasonably believe that our interlocutors are “as sensible, serious, and careful as oneself” and have “reviewed the same information as oneself” about *p*, we can consider those interlocutors as sharing epistemic symmetry with us regarding judgments about *p*.

To generate the philosophical problem of peer disagreement, Feldman rules out some important and relevant initial objections and concerns. First, the disagreeing interlocutors must be engaged in a **genuine disagreement**—one in which the parties are not having a mere verbal disagreement. Second, some version of the **Uniqueness Thesis** must hold. The Uniqueness Thesis (also just called “uniqueness” here) posits that any given body of evidence can support at most one truth value for any given proposition (uniqueness is discussed in more detail below). As I discuss later, these three conditions—epistemic symmetry (i.e., epistemic peerhood), genuine disagreement, and uniqueness—generate a trilemma that represents the peer disagreement puzzle. For now, I'll focus on the development of the puzzle in the literature. To take an example, here is

Feldman's paradigmatic case (one of several, this being the most basic) that illustrates a peer disagreement:

QUAD:

Suppose that you and I are standing by the window looking out on the quad. We think we have comparable vision and we know each other to be honest. I seem to see what looks to me like a person in a blue coat in the middle of the quad.

(Assume that this is not something odd.) I believe that a person with a blue coat is standing on the quad. Meanwhile, you seem to see nothing of the kind there. You think that no one is standing in the middle of the quad. We disagree. In isolation—before we talk to each other—each of us believes reasonably. But suppose we talk about what we see and we reach full disclosure. At that point, we each know that something weird is going on, but we have no idea which of us has the problem. Either I am 'seeing things' or you are missing something. I would not be reasonable in thinking that the problem is in your head, nor would you be reasonable in thinking that the problem is in mine. (Feldman, 2006, p. 8)

Feldman's use of "full disclosure" in the above quoted example refers to a type of disagreement where the parties not only realize that others disagree with them, but they are actually having a disagreement with someone with whom they *believe* they share epistemic symmetry with.<sup>3</sup> Furthermore, Feldman gives this example to defend against the idea that parties to a disagreement may have a special, private (subjective) kind of evidence that gives one party an epistemic advantage over the other party. It is also worth

---

<sup>3</sup> Feldman distinguishes between the following stages or types of disagreement: isolated, full disclosure, and mutually recognized reasonable disagreement.

noting that Feldman devised QUAD as a response to the idea that intuitions or other forms of private evidence break the epistemic symmetry. That is, Feldman's reason for presenting us with QUAD is to suggest that there are disagreements where private evidence doesn't seem to matter (in QUAD, exactly as it is stipulated, it is difficult to imagine what kind of private evidence a person *could* have that would change the conclusion that Feldman draws).

Appealing to QUAD (and other examples), Feldman arrives at what is ultimately his conclusion for how we ought to handle cases of disagreement when epistemic symmetry obtains: if we have no good reason to believe that our disagreeing interlocutor is wrong in a situation where the public evidence is simple, obvious, and shared, then we ought to suspend judgment on our own belief(s) about the matter that is being disagreed upon. Thus, the resolution to the problem of peer disagreement, according to Feldman, is that we ought to suspend judgment when faced with a disagreeing peer.

At first, Feldman's view might appear innocuous. After all, it seems reasonable to initially endorse a principle along these lines: when we are faced with a genuine disagreement where epistemic symmetry obtains, we ought to suspend judgment about the matter under dispute. However, the motivation to discuss the problem of peer disagreement any further comes from the observation that these disagreements are *common*. Feldman's impetus for investigating these types of disagreements isn't that they are philosophical oddities that crop up in a few weird cases, but instead that they show up all over the place. Such disagreements are present in law, science, politics, philosophy, religion, and so on. If these disagreements are common, then, following Feldman's

intuitions from QUAD, we rationally ought to find ourselves *suspending judgment regularly* in our daily lives much more than we typically do. That is the original problem of peer disagreement, and Feldman’s solution to that problem is that we ought to suspend judgment on the matter being disagreed upon. This position—one that encourages us to suspend judgment—was subsequently labeled **conciliationism**.

Of special interest here, notice that Feldman is both motivated by and concerned about disagreements that we—actual people in real-world situations—face often. That is, Feldman is not focused on formalizing theoretical models of disagreement resolution; he is concerned with discovering principles that guide how we *actually ought to respond to real-world disagreements*. His concern carries through to the works of other seminal authors in the literature, including the early work of Thomas Kelly.

In his 2005 “The Epistemic Significance of Disagreement”, Thomas Kelly discusses several issues surrounding the problem of peer disagreement.<sup>4</sup> For instance, Kelly (2005) compares merely possible vs. actual disagreement and discusses the necessity of disagreeing on matters of fact for a genuine disagreement to take place. However, there are two parts of Kelly’s 2005 article that are relevant for us here: the required conditions for epistemic peerhood and Kelly’s first argument against conciliationism.

---

<sup>4</sup> Kelly (2005) also briefly traces the history of the epistemic issues with disagreement from Sextus Empiricus to Henry Sidgwick, which is illuminating in terms of how philosophers have considered the significance of disagreement with “peers” throughout the history of philosophy. Additional significant precursors to the current peer disagreement discussion are found in van Inwagen 1996 and Richard Foley 2001. Also, it should be noted that Kelly’s view changes somewhat from his 2005 chapter to his 2010 chapter, where he advocates for the Total Evidence View, discussed later in this chapter.

Kelly's definition of epistemic peerhood is more formal and precise than Feldman's. Here is how Kelly, borrowing from Gary Gutting,<sup>5</sup> defines epistemic peerhood:

Let us say that two individuals are **epistemic peers** with respect to some question if and only if they satisfy the following two conditions:

- (i) they are equals with respect to their familiarity with the evidence and arguments which bear on that question, and
- (ii) they are equals with respect to general epistemic virtues such as intelligence, thoughtfulness, and freedom from bias (2005, pp. 174-175)

An extensive analysis of epistemic peerhood follows in the next chapter. There, I discuss Kelly's definition of epistemic peerhood in detail. For now, note that Kelly proposes two distinct conditions for epistemic peerhood to obtain. First, for interlocutors to be epistemic peers, they must be equally familiar with the **evidence and arguments** that bear on the proposition being disagreed upon. So, for example, if you and I disagree about  $p$ —let's say I believe that  $p$  and you believe that not- $p$ —one of the two conditions that must be met for us to be engaged in a *peer* disagreement is that we are equals with respect to the  $p$ -relevant evidence and arguments. I'll refer to this condition as the "**evidence condition**". As Kelly explains, consistent with Feldman's epistemic symmetry

---

<sup>5</sup> From Kelly (2005): "I owe the term 'epistemic peer' to Gutting (1982). Gutting uses the term to refer to those who are alike with respect to 'intelligence, perspicacity, honesty, thoroughness, and other relevant epistemic virtues' (p.83). I will use the term in a somewhat extended sense. As I will use the term, the class of epistemic peers with respect to a given question are equals, not only with respect to their possession of the sort of general epistemic virtues enumerated by Gutting, but also with respect to their exposure to evidence and arguments which bear on the question at issue." (p. 3, fn 3)

requirement, being equals with respect to the evidence condition means that you and I have reviewed the same body of evidence that bears on  $p$  and we have spent qualitatively equal time and care studying that body of evidence. For instance, if I merely glanced at the body of evidence regarding  $p$  and you studied it carefully, Kelly's first condition for epistemic peerhood would not be met since you studied the evidence more carefully than I did.

To meet the second condition, you and I must possess and exercise similar **epistemic virtues**, such as intelligence and freedom from bias, as each of us evaluated the evidence that bears on  $p$ . I will refer to this as the "**virtue condition**". A standard example of how two disagreeing parties may meet the evidence condition but not meet the virtue condition involves two parties judging the performance of a child at some task (say, a recital). The parent of the child will often find it difficult to be objective about their child's performance while a stranger can often more objectively assess the child's performance (this example is from Christensen 2014). Should the parent and stranger come to disagree about the quality of the child's performance, the stranger may question the parent's ability to remain unbiased when judging the quality of that performance. Since "freedom from bias" is part of the virtue condition, the parent and the stranger may not be on equal ground in this case. Understanding these conditions clarifies what precisely is being compared among interlocutors for peerhood to obtain.

With this more formalized notion of epistemic peerhood in place, Kelly argues that there can be cases where epistemic peerhood obtains (perhaps only initially) but that the discovery of disagreement does not rationally require the peers to suspend their



beliefs. According to Kelly 2005, discovering that someone disagrees with you—whether they are a peer or not—does not change the original, first-order evidence that bears on the proposition under dispute. For Kelly, first-order evidence is the evidence that bears on whatever is being disagreed upon, and higher-order evidence is evidence about the factors involved that surround the disagreement. For instance, the discovery of disagreement itself counts as higher-order evidence. The discovery of disagreement, according to Kelly, should not always be considered *informative* higher-order evidence about the first-order evidence that bears on whatever is being disagreed upon. That is, discovering disagreement need not always demand that we change how we evaluate our attitudes towards the *original evidence* about what is being disagreed upon. Because of this, for Kelly, the epistemic significance of peer disagreement is limited. I discuss his reasons for this conclusion below.

Kelly 2005 argues that potential peers may *generally* meet the evidence condition and the virtue condition, but, in cases when we disagree on a *specific* proposition, that potential peer may turn out to not be our peer after all. Here is Kelly's analogy: "Two chess players of equal skill do not always play to a draw; sometimes, one or the other wins, perhaps even decisively" (2005, p. 16). The idea here is that the two chess players might be generally equally skilled at playing chess, but in any particular chess match, one player might have made less mistakes than the other player or noticed an opportunity that the other player did not. These mistakes and missed opportunities explain why one player wins and the other does not, despite the fact that they are otherwise equally ranked.

Kelly's analogy connects back to the problem of peer disagreement in the following way:

sometimes our peers have made a momentary mistake or aren't as engaged in the debate as we are at the time of the disagreement. They may be generally as smart and thoughtful as we are and they may have access to the same set of evidence that we have, but we can't expect that equality to remain consistent during every single disagreement. So, for Kelly, the discovery of disagreement need not always count as relevant higher-order evidence that our original first-order evidence is faulty. Instead, the discovery of disagreement may actually serve as higher-order evidence that our interlocutor is *not* an epistemic peer regarding the disputed proposition. By demoting one's potential epistemic peer to epistemic inferior, one is permitted to retain their original doxastic attitude towards whatever is being disagreed upon. Clearly, this position conflicts with Feldman's conciliationist view. This is called the **steadfast** view. Supporters of the steadfast view claim that even in the face of disagreement with a potential peer, we are permitted to retain our controversial attitudes in some if not most cases.

So, two key positions appeared early in the literature, almost simultaneously:

- (1) Conciliationism: When one discovers that one's epistemic peer disagrees with them about  $p$ , one should rationally suspend judgment about whether  $p$ .
- (2) Steadfastness: When one discovers that one's putative epistemic peer disagrees with them about  $p$ , one can demote their epistemic peer to an epistemic inferior about  $p$  and maintain one's original belief about  $p$ .

Although there are other perspectives on the peer disagreement problem presented in the literature, these are the two main positions debated by philosophers.<sup>6</sup> Below, I summarize the development of the two main views.

## **2.2: Conciliationism**

After Feldman's initial proposal that suspending judgment is the appropriate response to a peer disagreement, two additional philosophers, David Christensen and Adam Elga, championed the conciliatory view. I briefly review their defense of that view below.

To start, here is Christensen's (2007) definition of epistemic peers and his formulation of the peer disagreement problem:

Suppose I find out that my friend disagrees with me about P: she has moderately high confidence that it's true, and I have moderately high confidence that it's false. But to the best of my knowledge, my friend is just as well informed as I am—in fact, we may suppose that my friend and I have had long discussions in which we share every bit of evidence we can think of that's relevant to P. And suppose further that I have good reason to believe that my friend and I are equally intelligent and rational, and that I know of no general reason (like the fact that people tend to be biased toward their children) to think either of us is especially likely to be particularly good, or bad, at reacting to evidence on this particular

---

<sup>6</sup> There are several other interesting discussions not focused on defending or refuting the two core views in the literature. However, the bulk of the literature thus far is aimed at defending one of the primary views—conciliationism or steadfastness—or denying both views in favor of a different response to peer disagreement altogether. Also, the disagreement literature is still relatively new in terms of the history of philosophy, so conceivably the project could mature and move away from its focus on those two main positions.

topic—no reason, that is, aside from the fact that my friend disagrees with me about P. In other words, my friend seems to be what some have called an “epistemic peer.” In this sort of case, should I revise my belief? (2007, pp. 188-189)

Notice here that Christensen is not focused on a general notion of epistemic peerhood, but on evaluations of epistemic peerhood made during disagreements about particular propositions. That is, *epistemic peers are peers on p*, not on a general topic that includes *p*. This is a deviation from Kelly’s claim that two parties can be epistemic peers in general but perhaps not peers in cases where the dispute is over a particular proposition (recall that Kelly’s definition of peerhood is phrased as two people being peers “with respect to some question”. (2005, p. 16) rather than a specific proposition). So, by Christensen’s lights, epistemic peers are peers relative to specific propositions rather than general domains. This adjustment is important to notice because, as I’ll argue below, this version of peerhood introduces problems with how peers evaluate themselves and each other with respect to a proposition; the definition of peerhood is narrowing. For now, with this more precise definition of peerhood in mind, let’s examine Christensen’s conciliationist approach to resolving the problem of peer disagreement.

To defend conciliationism, Christensen (2011) develops the **Independence Principle** (IP), which he states as follows: “In evaluating the epistemic credentials of another’s expressed belief about P, in order to determine how (or whether) to modify my own belief about P, I should do so in a way that doesn’t rely on the reasoning behind my initial belief about P” (2011, pp. 1-2). In other words, when deciding how to respond to a

peer disagreement about  $p$ , we should make that decision using evidence that is independent of the evidence and reasons we had for holding that  $p$ .

The IP is intended to give us an intuitively appealing result because, as Christensen argues, not following the IP would lead to problems in how we evaluate the epistemic status of others. The central worry is that if I discover that a peer disagrees with me about  $p$ , it doesn't seem rational for me to shrug off the epistemic pressure to revise my belief about  $p$  on the assumption that my original reasoning and evidence regarding  $p$  is solid. If that were the case, the disagreement would be of little epistemic consequence for my belief about  $p$ . Christensen argues that appealing to our original reasoning and evidence regarding the proposition under dispute is question-begging. For example, if I assume my original evidence regarding  $p$  is solid and I therefore think that the discovery of disagreement is epistemically inconsequential, I could use my confidence in my original evidence and reasoning to become extremely epistemically dogmatic. Consider: if one thousand experts are in consensus about  $p$  and disagree with me about  $p$ , without the IP, I could justify not changing my mind about  $p$  merely by citing my (probably bad) original evidence. In that case, I would not be under any pressure to change my mind about  $p$ . This is a bad result given that the norms of rationality usually require reasonable doxastic revision in the face of disagreement with one thousand experts. According to the IP, to maintain a rational view about  $p$  after discovering disagreement about  $p$ , I would need to find evidence and reasoning *independent of my original evidence and reasoning* to overcome the epistemic pressure to revise my belief.

Notice that Christensen's IP directly contradicts Kelly's claim that we can remain steadfast during a peer disagreement by citing our original, first-order evidence as good reason to not revise our beliefs in the face of disagreement. This would permit us to ignore the discovery of disagreement as though it was epistemically *insignificant*. That is, according to Kelly's 2005 view, we are not rationally required to take the discovery of disagreement over  $p$  as higher-order evidence that we may be wrong about  $p$ ; on that view, there are some cases where our first-order evidence concerning  $p$  is sufficient to fend off the resulting epistemic pressure to revise our beliefs in the face of disagreement.

Since the IP limits the scope of what evidence we are permitted to employ to defend our position in a disagreement, the only remaining rational option, on Christensen's proposal, is to take the discovery of disagreement about  $p$  among peers as evidence that at least one of the parties is wrong about  $p$ . Given the epistemic symmetry between the parties, we don't have a principled way to tell which (if not both) of the parties is mistaken, so the pressure to conciliate to some degree (but perhaps not equally) falls on both parties in the disagreement.

The question of how much conciliation is rationally warranted in any particular disagreement is also an ongoing matter of debate. The prevailing view among advocates of conciliationism is that we ought to "split the difference" in terms of our belief credences toward the proposition at issue. Suppose we are epistemic peers and I believe that  $p$  with a credence of 0.8 and you believe that not- $p$  with a credence is 0.2. If we disagree about  $p$ , the advice to split the difference suggests that we should both end up at a 0.5 credence in our respective beliefs about  $p$ . However, without strong evidence of

epistemic symmetry between us (for example, in cases where it is reasonable for both of us to believe that we might be epistemic peers on  $p$  but we wouldn't bet our life savings on that), it could turn out that in the previous case, it is reasonable for me to arrive at 0.6 and you to arrive at 0.4. However, it is unclear if this still counts as splitting the difference in the appropriately conciliatory way. The prevailing view of conciliationism leans towards splitting the difference with exact averages of credences for all contributing parties involved in the disagreement.<sup>7</sup>

Christensen further proposes that we ought to treat disagreements with those whom we judge to be our peers to be moderated by the **Equal Weight View** (EWV).<sup>8</sup> The EWV is motivated by Christensen's Restaurant Case (or Mental Math):

RESTAURANT:

Suppose that five of us go out to dinner. It's time to pay the check, so the question we're interested in is how much we each owe. We can all see the bill total clearly, we all agree to give a 20 percent tip, and we further agree to split the whole cost evenly, not worrying over who asked for imported water, or skipped dessert, or drank more of the wine. I do the math in my head and become highly confident that our shares are \$43 each. Meanwhile, my friend does the math in her head and becomes highly confident that our shares are \$45 each. How should I react, upon learning of her belief?

---

<sup>7</sup> See Christensen 2007, pp. 210-214 for a brief discussion on balancing credences, and Matheson 2015, pp. 74-79 for a fuller explanation of splitting the differences with credences, especially in relation to how all-or-nothing belief models might map onto various credence adjustments.

<sup>8</sup> Elga 2007 terms Christensen's view the "Equal Weight View" retroactively in publication.

I think that if we set the case up right, the answer is obvious. Let us suppose that my friend and I have a long history of eating out together and dividing the check in our heads, and that we've been equally successful in our arithmetic efforts: the vast majority of times, we agree; but when we disagree, she's right as often as I am. So for the sort of epistemic endeavor under consideration, we are clearly peers. Suppose further that there is no special reason to think one of us is particularly dull or sharp this evening—neither is especially tired or energetic, and neither has had significantly more wine or coffee. And suppose that I didn't feel more or less confident than usual in this particular calculation, and my friend reports that she didn't either. If we set up the case in this way, it seems quite clear that I should lower my confidence that my share is \$43 and raise my confidence that it's \$45. In fact, I think (though this is perhaps less obvious) that I should now accord these two hypotheses roughly equal credence. (Christensen, 2007, p. 193)

Before discussing the details of RESTAURANT and how it relates to the EWV, I'd like to pause to make a special note about Christensen's stated goals for designing this now-famous case in the way that he did. Christensen acknowledges that RESTAURANT is "somewhat idealized" (2007, p. 189) and intentionally simple: "The restaurant case is designed to be simple in two ways: in the evidential situation and in the evaluation of the general capacities my friend and I exercise in reacting to that sort of evidential situation. This makes our intuitions about the case particularly clear. But the same lessons emerge, I think, from cases involving a bit more complexity" (2007, p. 193). After discussing



RESTAURANT as his initial test case, Christensen then offers a more complex case involving meteorologists who disagree about the chances of rain after examining their respective forecasting models. One of the goals of presenting this more complex case, according to Christensen, is “[t]o make this less like the restaurant case and more like many cases of real-life disagreement...” (2007, p. 193). That is, Christensen is concerned with building test cases that resemble “real-life” cases of disagreement. So, in addition to Feldman’s concerns about how disagreements among interlocutors with epistemic symmetry in domains such as politics, law, science, and so on ought to be resolved, Christensen is also concerned with how to transfer the intuitions elicited by his cases to real-world disagreements. This serves as further evidence to support my claim that the peer disagreement literature was and is aimed at giving us advice about how we ought to respond to and resolve real-world disagreements. This point is crucial for understanding the spirit of my project, so I recommend flagging this paragraph as support for a premise in my argument in Chapter 6. For now, we can turn our attention back to Christensen’s RESTAURANT case.

Returning to RESTAURANT and the EWV: Since the two diners are epistemic peers in this case, our reaction is supposed to be that we can find no good reason to claim that one party is correct while the other is not. Because of this, according to Christensen, the two disagreeing diners are rationally required to lower their confidence in their respective beliefs about how to split the bill. Recall that the IP dictates that the diners are not permitted to rely on their original reasoning to avoid this mandated belief revision. That is, given the IP, the parties involved in RESTAURANT ought not defend their

respective position on the divided total simply by claiming that they performed the arithmetic correctly.

RESTAURANT intends to show that when the parties to a disagreement are epistemic peers and the IP is followed, those parties ought to treat each other's opinions as though they are "weighted" equally. Furthermore, the fact their opinions should—due to the evidential and virtue equality or parity required for peerhood—be equally weighted bears directly on how much each party should revise their original view. This explains the reasoning behind endorsing a basic version of the EWV: if you and I disagree about  $p$ , and neither of us had any good *antecedent* reason to doubt that either of us have epistemic advantage regarding  $p$  (i.e., we are epistemic peers on  $p$ ), then we ought to treat each other's respective opinions as just as plausible as our own differing opinions regarding  $p$ .

To illustrate, here is an example from Adam Elga (2007) that he believes leads us to the intuition that EWV is the correct way to handle peer disagreement cases:

HORSE RACE:

To see the correctness of the equal weight view, start with a case of perceptual disagreement.

You and a friend are to judge the same contest, a race between Horse A and Horse B. Initially, you think that your friend is as good as you at judging such races. In other words, you think that in case of disagreement about the race, the two of you are equally likely to be mistaken. The race is run, and the two of you

form independent judgments. As it happens, you become confident that Horse A won, and your friend becomes equally confident that Horse B won.

When you learn of your friend's opposing judgment, you should think that the two of you are equally likely to be correct. For suppose not—suppose it were reasonable for you to be, say, 70% confident that you are correct. Then you would have gotten some evidence that you are a better judge than your friend, since you would have gotten some evidence that you judged this race correctly, while she misjudged it. But that is absurd. It is absurd that in this situation you get any evidence that you are a better judge. (Elga, 2007, p. 486).

Elga discusses further reasons why not equally weighting the opinion of your friend would be “absurd.” If the HORSE RACE scenario were extended to a series of races, you—with your 70% confidence that you judged the winner correctly each time—would build up immense confidence that you were correct without, as stipulated in the case, any antecedent reason to justify that confidence. According to Elga, reasoning this way would amount to illicitly bootstrapping your confidence about which horse won the race. Elga concludes: “Here is the bottom line. When you find out that you and your friend have come to opposite conclusions about a race, you should think that the two of you are equally likely to be correct” (2007, p. 487). Hence, we should treat our peer's opinions as equally weighted to our own in cases of disagreement.

Conciliationism, then, is supported by the Independence Principle and the Equal Weight View.<sup>9</sup> If you and I believe each other are epistemic peers, and we disagree about

---

<sup>9</sup> As stated in a previous footnote, Elga (2007) calls Christensen's view the Equal Weight View. This could lead one to think that “Equal Weight View” is synonymous with “Conciliationism.” However, to capture

*p*, and I cannot find a reason independent of my original reasoning for my belief to defeat your position, and I must treat your opinion as equally (epistemically) weighted to my own, it seems the only rational option left for me is to conciliate and move closer to your view. That is the core defense of conciliationism in the literature.

Those who deny that conciliating is the only rational response to peer disagreement will have to either undermine or refute these principles or reframe the problem. We turn now to those attempts by the supporters of the steadfast view in the literature.

### **2.3: Steadfastness**

Most of us have had the experience of being stubborn about our position during a disagreement. Sometimes we were justified to be stubborn, and sometimes we weren't. Deciding when it is rational to remain stubborn about our position isn't always easy—we are often committed to our doxastic positions by way of inference from other cherished beliefs, and sometimes we are just too emotionally wrapped up in our views to detect the errors in our reasoning. However, there are times when deciding to remain steadfast is easy. If a typical child disagrees with us about the solution to a complex math problem and we are sure they have not been taught how to solve it (but we have), then there is no pressure to reevaluate our solution once we discover the child's dissenting opinion. But, when it isn't clear that we have an epistemic advantage over the other party, remaining

---

how some authors use the EWV (esp. Kelly 2010) in several modified versions, I distinguish EWV from Conciliationism in the following way: EWV tells us that we ought to give our disagreeing peer's opinions the same weight as our opinions, full stop. Conciliationism describes the appropriate rational response ("splitting the difference") when we give equal weight to the opinions of our peers *and* have no good reason to remain steadfast in our views. Hence, I see the EWV as a component of conciliationism rather than as just another name for conciliationism. If this idiosyncratic, I've at least explained why here.

steadfast in our position requires justification. Kelly (2005) proposed that we can remain steadfast in our position because our peers may turn out to only be peers generally rather than on a specific proposition or on a specific occasion (recall his chess players analogy above). However, there may be times when we are rationally permitted to remain steadfast even when we disagree with those whom we initially believe to be our epistemic peers. As I discuss below, Peter van Inwagen and Richard Fumerton take similar but distinct approaches to providing ways that we might be able to justify remaining steadfast in our views when faced with peer disagreements.

Van Inwagen (2010) suggests, hesitantly, that it can be rational for two parties to hold contradictory views (i.e., one accepts a proposition and the other denies it) even when they possess the same *public* evidence. At first blush, that suggestion would seem to violate the Uniqueness Thesis (recall that uniqueness posits that any given body of evidence can support at most one truth value for a proposition). How could it be rational for two people to reach incompatible conclusions when they have the same evidence and the same epistemic virtues? When this happens, something has apparently gone wrong. Either uniqueness does not hold (and I'm stipulating here that it does hold), or another solution needs to be presented. Van Inwagen takes the latter route and proposes that similarly intelligent and thoughtful people can have access to the same *public* evidence and still rationally disagree with each other without violating uniqueness. According to van Inwagen, we often have unexportable, ineffable evidence that is *private and cannot be made public*. Here is his example of this type of evidence:

I sometimes know that my wife is angry when no one else does, for example, and I cannot explain to anyone how I know this—I cannot give what Plato would call an “account” of what underlies my conviction that she is angry. It seems to me to be plausible to say that in such cases my belief that my wife is angry is grounded in some body of evidence, evidence that lies entirely within my mind and that I cannot put into words. (van Inwagen, 2010, p. 25)

Van Inwagen gives two additional examples of ineffable private evidence: chicken sexers and mathematicians who can “see” a proof but cannot prove it at first (2010, p. 25).

Given these examples, it seems that there may be cases of “peer” disagreement in which there is evidence that we simply cannot share with others. So, if it is the case that we have private evidence concerning a disputed proposition  $p$ , it may be rational for two people to disagree about  $p$  without feeling the pressure to be conciliatory. That is, If you and I have the same public evidence (we meet the evidence condition in terms of *public* evidence) that bears on  $p$  and we have the same epistemic virtues (we meet the virtue condition for epistemic peerhood), we can remain steadfast in our positions on  $p$  because we each also have private, unexportable evidence that we cannot make available to each other. Van Inwagen (2010) reports being dissatisfied with this response, but the alternative seems to be that we should all be suspending judgment on most of our cherished beliefs about philosophy, politics, religion, and science—a policy which seems even more unpalatable. So, while his view suggests some bullet-biting, it is a plausible way to avoid the skeptical pressures of conciliationism.

Using a similar approach, Fumerton (2010) argues that we can trust our own thinking (evidence processing, inferences, etc.) more than we can trust the thinking of others. If two people have the same evidence and share the same epistemic virtues, one thing they cannot share is their subjective feelings of being confident in their respective positions. Fumerton refers to this as the “egocentric” view and he argues that it is nearly impossible for us to trust the thinking of others as well as we can trust the thinking of ourselves (2010, pp. 92–99). If a disagreement arises over  $p$ , this trust or confidence in one’s own reasoning to a judgment about whether  $p$  is in itself additional evidence that bears on one’s justification that  $p$  (perhaps as higher-order evidence). And, as in the case with van Inwagen’s proposal, this evidence is entirely private in terms of the “feeling” of confidence that comes along with our estimations about how much or little confidence we give to our beliefs regarding  $p$ . According to Fumerton, we may rely on introspective access to our own thought processes as evidence that we have carefully and diligently processed the available (disagreement relevant) evidence. With this additional introspectively gained evidence, we can conclude that the discovery of disagreement need not sway us to change our minds. In Fumerton’s words:

I do know how I reason better than I know how others reason. It is important to keep firmly in mind that in the final analysis there really is no alternative to the egocentric perspective. Even when my discoveries about what others believe defeat the justification I had prior to those discoveries, it is *my* discoveries that are doing the defeating. I can only use the discovery of disagreement to weaken my justification insofar as I trust *my* reasoning. Without such trust, there is no access

even to what others believe. That is not to deny that trust in my reasoning ability can turn on itself—can lead me to doubt the very faculties that I trust. But when that hasn't happened, and when I can't understand exactly what is going on in the minds of others, I'll always turn back to the reasoning I understand best—my own. (2010, p. 106, emphasis in original)

Although Fumerton argues that we cannot escape the egocentric perspective, it is worth noting he does not claim that his self-knowledge or self-trust can *always* tip the evidential balance in one's favor in disagreements. However, when one faces a disagreement with one's potential epistemic peers, Fumerton believes that this self-trust in one's own internal tracking of evidence processing and reasoning is often sufficient to deny that peer disagreement is always or even usually epistemically significant.

Van Inwagen's and Fumerton's respective yet similar views represent the core of the steadfast position; steadfast views usually deny that we ought to conciliate in *all* cases of peer disagreement. If we believe that someone who disagrees with us could be an epistemic peer, we may be permitted to remain steadfast in our belief anyway. To be clear, the steadfast position does not deny that we ought to conciliate in *some* cases of disagreement. However, it does claim that we are permitted to remain steadfast in disagreements when we have reason to believe that we have evidence that our interlocutor does not—or cannot—have.

There are other steadfast replies, but they often follow this same line of thinking: We can demote those whom we reasonably believed to be peers prior to a disagreement if



we have good reason to believe that they lack important disagreement-relevant evidence or virtues that we believe we possess.

#### **2.4: Conciliation and Steadfast Summary**

These two positions—conciliationism and steadfastness—stand at two opposite ends of the spectrum of proposed rational responses to peer disagreements. If we exaggerated, we could say that conciliationists argue that the *only* rational response to peer disagreement is to meet exactly in the middle with our disagreeing interlocutors by setting our credences as the exact average of the credences of the parties involved in the disagreement. Further, the exaggerated steadfast view could be described as saying that we may *always* hold our ground when confronted with disagreement, even against those whom we believed to be epistemic peers prior to the disagreement, and thus we face no normative pressure to conciliate. Of course, these extreme caricatures are not what most philosophers defend in the literature. Christensen and Elga both offer cases where we need not conciliate with those whom we believed to be epistemic peers (say, if we discover that the person disagreeing with us is drunk or that their response is very obviously mistaken), and Fumerton admits that there are times when one should conciliate or change one’s beliefs after carefully considering the evidence (he uses as an example his own intuitive response to the Monty Hall puzzle and his later realization that his initial response was wrong) (2010, p. 94).

Conciliationism and steadfastness are the two main views debated in the literature.<sup>10</sup> If we imagine these two views as occupying opposite ends of a spectrum of

---

<sup>10</sup> Several alternative proposals have been offered to solve the puzzle of the epistemic significance of peer disagreement. For instance, Sosa (2010) questions whether we are actually ever having genuine

responses to the peer disagreement debate, there are also views that fall somewhere in the middle of that spectrum. I call these middle positions the *hybrid positions*, which I discuss in the next section.

## 2.5: Hybrid Positions

After Kelly's initial defense of steadfastness in his 2005 chapter, he developed a new position that he calls the **Total Evidence View** (TEV) (Kelly, 2010). The TEV rejects the claim that the Equal Weight View (EWV) is part of the appropriate rational response to peer disagreements. However, Kelly's TEV is supposed to capture the intuitively appealing aspects of Christensen's (and Elga's) EWV without demanding that peers conciliate in every disagreement. Kelly's attack on the EWV is underwritten by both an implicit attack against the Independence Principle (IP) and his objections to cases of conciliation that lead interlocutors into *less* rational positions than they were before they discovered their disagreement. Let me briefly explain his arguments.

As I discussed above, Kelly distinguishes two broad classes of evidence: first-order evidence (FOE) and higher-order evidence (HOE). FOE is the initial evidence that we possess prior to the disagreement; this is the evidence a person bases their beliefs on. In other words, FOE is the evidence that epistemic peers are supposed to share about some topic. For instance, in RESTAURANT, the FOE is the restaurant bill that the two peers are using to calculate the evenly split the bill. In more complex cases, the FOE is all of the shared evidence, arguments, etc. that one needs to take into account when making a

---

disagreements, and Hawthorne and Srinivasan (2013) argue that arriving at general epistemic norms that could guide our responses to disagreement are difficult to defend, thus, the structure of peer disagreement fails to produce the appropriate norms. However, these alternative proposals are few and far between; a good amount of the growing literature is dedicated to defending either conciliationism or steadfastness.

judgment. Kelly sometimes refers to FOE as “non-psychological” (2010, p. 128) evidence because FOE is generally not about the person holding the evidence, but instead about what has been presented to the person as data, observations, information, inferences, etc. On the other hand, HOE is evidence about our FOE. Paradigmatically, one type of HOE is the discovery of disagreement. If you and I disagree about  $p$  and we are putatively peers on  $p$ , then we have “the same” FOE that bears on  $p$ . Furthermore, once the disagreement comes to our attention, we now also have HOE that a putative peer on  $p$  disagrees with us about  $p$ . The disagreement is not only additional evidence about  $p$ , but it is also HOE about our FOE about  $p$ . So, in this case, our HOE is that someone disagrees with us about  $p$ . Kelly refers to HOE as “psychological” (2010, p.128) evidence because HOE is about the opinions that a person holds rather than the FOE those opinions are based on.

With the FOE / HOE distinction in hand, Kelly argues that there may be cases in which we can rely on the strength of our FOE enough that discovering disagreement, which presents us with HOE about our FOE, gets little to no epistemic weight. Notice that, similar to Kelly’s initial 2005 defense of steadfastness, this position denies that the IP is a good rule for rationally resolving disagreement. Kelly’s 2010 Total Evidence View permits us cite our original FOE alone as justification for remaining steadfast in our beliefs, which is clearly a violation of the IP. Recall that Christensen argues that we should not dismiss the epistemic significance of peer disagreement by ignoring the discovery of disagreement itself as evidence that something has gone wrong. For Christensen, peer disagreement rationally demands that the involved interlocutors must

adjust their beliefs (attitudes or credences) accordingly. Furthermore, the IP tells us that we are not permitted to simply rely on our original disagreement-relevant evidence to remain steadfast—our FOE alone is not sufficient to reject the pressure to conciliate. That is, for conciliationists, *disagreement itself matters as evidence*. However, Kelly (2010) argues that there are cases of putative peer disagreement where returning to our original evidence—i.e., our FOE—to support our position and resist conciliation is rational. His view, then, is based on denying that we ought to follow the IP in all cases of disagreement.

Kelly refutes the IP by illustrating the ways that the EWV can lead us to hold irrational beliefs. For instance, Kelly (2010) revisits Christensen’s RESTAURANT case except he discusses a variant (originally from Christensen 2007) where one of the putative peers (Kelly casts himself in this role in his example) arrives at the amount owed by each diner to be \$450 (recall that in the original case, the peers arrived at \$43 and \$45 respectively). In this case, Kelly argues that it would be absurd to conciliate:

[H]ere we note only how the Total Evidence View offers an extremely straightforward and compelling explanation of why you are entitled effectively to discount my absurd opinion. Quite simply: given the totality of considerations available to you that bear on the question at issue (for example, your knowledge that the total bill is  $n$ , a number that is less than \$450), it would be completely unreasonable for you to give any significant credence to the proposition that a share of the total bill is \$450, despite the fact that this is what I, your peer,

believe. In this case, it is the non-psychological considerations that swamp the psychological considerations into epistemic insignificance. (Kelly, 2010, p. 150)

The point of Kelly's discussion of the modified RESTAURANT case is to demonstrate that there are times when someone we believe to be our peer is so obviously mistaken that we may (and perhaps ought to) dismiss or at least downgrade the epistemic significance of a disagreement with them. Instead, in these cases, it would be more reasonable to assume that the putative peer has made a mistake, so we can remain steadfast in our beliefs. In fact, if we were to conciliate in the modified RESTAURANT case, we could end up in a less rational position about how to split the bill than we were before the disagreement. That's because conciliation could lead us to revise our belief credences about the amount owed in the *wrong* direction (in this case, "towards" the belief that the amount owed by each is \$450, which is clearly mistaken). As Kelly argues, there are instances when ignoring our original FOE after discovering disagreement can be irrational. Hence, the IP can lead to irrational outcomes when applied to (potential) peer disagreements.

Kelly's arguments continue at length, but his Total Evidence View can be succinctly summed up in his own words:

Granted that, on the Total Evidence View, both the first-order evidence and the higher-order evidence count for something, which kind of evidence plays a greater role in fixing facts about what it is reasonable to believe?

It is a mistake, I believe, to think that there is some general answer to this question. In some cases, the first-order evidence might be extremely substantial

compared to the higher-order evidence; in such cases, the former tends to swamp the latter. In other cases, the first-order evidence might be quite insubstantial compared to the higher-order evidence; in such cases the latter tends to swamp the former. (2010, p. 142)

Kelly's view takes into consideration all of our evidence (hence, the "Total" in TEV) and has the advantage of avoiding unreasonable conciliation (as in the modified RESTURANT case) while allowing us to give weight to peers' opinions when, given the rest of our evidence, those opinions seem reasonable. Hence, TEV is a hybrid view that permits us to respond to disagreements on a case-by-case basis while employing all of the evidence that is available to us, both FOE and HOE.

In a somewhat similar vein, Jennifer Lackey (2007) has proposed a **justificationist** response to the peer disagreement question. Her view shares the flexibility of Kelly's TEV view, but for different reasons. On Lackey's justificationist view, the epistemic significance of (peer) disagreement maps onto how justified one's confidence is in one's disputed belief.<sup>11</sup> If one enjoys a high level of justification in one's confidence that  $p$ , then a disagreement with a peer about  $p$  has little to no epistemic significance. However, if one maintains a low level of justification in their confidence that  $p$ , then a disagreement with a peer about  $p$  will have relatively greater epistemic significance. Furthermore, in the former case—since the disagreement is not epistemically significant—one can rationally retain their belief that  $p$  without pressure to revise (i.e., one can remain steadfast about  $p$ ). In the latter case, one is rationally required

---

<sup>11</sup> For clarification, Lackey explains what "justified confidence" refers to: "By a belief enjoying 'a very high degree of justified confidence,' I mean a very confident belief that is highly justified" (2007, fn16).

to revise their level of confidence that  $p$  because they lack justification for high confidence in  $p$ . So, for example, if you believe that you and I are epistemic peers on basic arithmetic and you say that  $1+1=2$  and I disagree with you, you will have an extremely high level of justified confidence for your belief that  $1+1=2$ . So, you will not be required to revise that belief at all. However, if you believe that you and I are epistemic peers on complex matters such as the whether the US was justified to increase tariffs on imported goods and I disagree with your position, you will most likely have a moderate to low level of justified confidence for your belief that the tariffs are a bad policy. In this latter case, according to Lackey, my disagreement with you over tariffs may require some doxastic revision on your part.

Notice that this counts as a hybrid view because the principles for rationally resolving disagreement put forth by Lackey permit the interlocutors to either conciliate or remain steadfast. Which action each interlocutor takes depends on how justified they are in holding their respective confidence(s) about their beliefs. That is, commensurate with one's level of justified confidence in one's belief, one will have a principled way to determine if one ought to rationally conciliate (when justification for confidence is low) or remain steadfast (when justification for confidence is high). So, for this view to be applied, we would need some reliable way to access our levels of justified confidence for our beliefs about disputed propositions. This may not be as straightforward as it sounds, but Lackey's proposal does offer us an alternative to the main views discussed above.

So, on the spectrum from steadfast views to conciliatory views, both Kelly's and Lackey's positions land somewhere in the middle. They offer us ways to rationally

respond to disagreements by revising our beliefs in some cases while remaining steadfast in others, depending on either the strength of our first-order evidence relative to our total evidence (Kelly) or the level of justification we have for our confidence in beliefs in question (Lackey).

These four views—conciliationism, steadfastness, total evidence, and justificationist—represent the vast majority of the positions taken and defended in the literature. Before discussing some of the common objections to these views, I want to pause here to mention that as these positions developed in the literature, defending or denying each became the central focus of contributors. This is a relevant point for my dissertation project because entrenchment in these positions led the authors away from the early goals of applying the literature’s advice to real-world cases of disagreement. That is, while contributors to the literature devised many more idealized toy cases of disagreement to defend their positions, the path from the advice given in the literature (from any position) to actual real-world application became obscured and ultimately lost. Finally, despite all the efforts to defend each of these views, each one suffers from significant defects, some of which I discuss in the next section.

## **2.6: *Prima Facie* Objections to the Main Views**

The views presented above face some *prima facie* objections. Given that I do not intend to defend any of these views here, I leave the reader to ponder the strength of each objection without adding my own commentary. For my purposes, it is enough to list the main objections merely as evidence that there is no cost-free view on offer.



***Conciliationism:***

There are (at least) three reasonable *prima facie* objections to the conciliationist views.

*Epistemic Wrong Way (Bootstrapping)*: One problem faced by conciliationists (noted by Kelly 2010) is that we get an intuitively epistemically bad result in cases like the following: You and I are both deluded into thinking that we have strong justification to believe that *p*. However, we are both mistaken, and *p* is highly unlikely. If my confidence that *p* is 0.75 and yours is 0.85 and we disagree and conciliate, we will both end up with confidence of 0.80 that *p*. However, suppose that our actual level of confidence should be 0.20 that *p*. Here we have a case where, according to conciliationism, we rationally responded to our disagreement and ended up with beliefs that are irrational. In fact, in this case, *my confidence has moved the wrong way* after our disagreement, so it is difficult to see how conciliation in this case was an epistemic benefit; I am less rational than I was before the disagreement and neither of us have come to a rational conclusion by applying the principles from conciliationism. While conciliationism does not require infallibility, we can imagine how revising “away” from the appropriate credence in a belief is a *prima facie* problem for the view.

*Disagreeing about Disagreement*: Let’s say that you and I disagree about the appropriate response to peer disagreement, and we believe that we are peers on the matter. If I am arguing for conciliationism and you are arguing for steadfastness, I will, by my own lights, need to move closer to your position. If we repeat this exercise, I will end up becoming a steadfast. Your view, on the other hand, is not going to be under any such threat because you will have remained steadfast during our multiple disagreements

about peer disagreement. In effect, you have pulled me to your view without having to budge at all. So, in the battle between conciliationism and steadfastness, conciliationism will dwindle away. Given enough time and disagreements about the appropriate response to peer disagreements, there will be no conciliationists left standing.

*Skepticism Rears Its Ugly Head, Again:* The general objection to conciliationism from the beginning of the peer disagreement project—that we will end up agnostic on every complex, meaningful matter that’s controversial because we will almost always find a peer who disagrees with us about our views—remains in full force. This widespread agnosticism is what Feldman cautiously argued for. Enough philosophers found Feldman’s conclusion so distasteful that an entire body of literature has grown up around avoiding pan-agnosticism with regard to complex matters that we disagree about in fields such as science, law, politics, philosophy, religion, etc.

Besides good reasons to maintain steadfast views, there are additional and more nuanced objections to conciliationism that have been discussed in the literature. However, the three *prima facie* objections I discussed above are the most common ones that conciliationists most often face. And there is also the more straightforward worry that when we are obviously right (think of the  $1+1=2$  case in the discussion of Lackey’s view above) in cases of disagreement with putative peers that *always* conciliating, even if only by a little bit, seems absurd.

***Steadfastness:***

The steadfast view faces (at least) one concerning *prima facie* objection:

*Dogmatism:* Steadfasters often argue that there may be times when it is appropriate to be conciliatory. But if one takes van Inwagen's or Fumerton's approach, we may also demote a disagreeing interlocutor to epistemic inferior with ease. We merely need to find evidence that our interlocutor cannot possess to break the epistemic symmetry required for *peer* disagreement. However, this approach invites dogmatism and narrow-mindedness. Most of us have been in disagreements with people who double down on claims that are obviously false or at least highly contentious. If those unreasonable people got ahold of these defenses of steadfastness, they would always have an escape route when faced with disagreement, even when it is obvious that they ought to revise their beliefs. So, steadfasters get their rationality cheaply through mysterious evidence, and, despite claims that they understand when they should conciliate and when they should remain steadfast, this seems dubious given the way many disagreements actually turn out with stubborn people.

***Hybrid Views:***

The hybrid views I've described tend to face fewer *prima facie* objections. After all, the hybrid views can often account for ways to handle many cases of disagreement that ardent conciliationists or steadfasters cannot. Unfortunately, this is both a feature and a bug for these views; they are so permissive that they become less useful as general guides for how we ought to respond to disagreement writ large. Here is my *prima facie* objection to the hybrid views:

*Permissive to the Point of Peril:* The Total Evidence View says that we should determine the epistemic significance of disagreement based on our total evidence (both

first-order and higher-order, with a suggestion that strong FOE can swamp relatively weaker HOE). The justificationist view says that we should determine the epistemic significance of disagreement based on whether we can justify the confidence in our beliefs. The harsh critic will point out that the advice from these hybrid views not only depends too much on individual differences for a general solution, but also that these prescriptions are basically fancy ways of saying: use common sense. That is, these views rely on some general norms of rationality and epistemic principles to help people decide how they ought to respond to disagreements. While this approach will work sometimes, the hybrid views are so permissive that they do not offer as much practical guidance as they purport to. How ought we respond to disagreements? Rationally. (Why didn't I think of that?!) How do we rationally respond to disagreements? By considering all of our available evidence and checking the justifications for our confidence in the beliefs we formed based on that evidence. However, it is not always clear that we are competent at performing the required tasks to apply these views. Recall that for Lackey's view, we need access to our levels of justified confidence, and our estimations of those confidence levels needs to be fairly precise and accurate. But it is unclear whether we *do* have the requisite access to these justified levels of confidence. So, it will be up to us to assign (probably inaccurately) our levels of justified confidence, which is tantamount to advising us to guess how confident we are in a belief. Again, this approach is too permissive because there are too many opportunities to get it wrong without knowing it.

While the above criticisms of the hybrid views borders on hyperbole, they do not overreach. The hybrid views help explain the problems that conciliationism and

steadfastness bring to the debate, but at the cost of not being able to give clear advice on how to recognize when we ought to turn to our FOE and dismiss the Independence Principle, or how we can tell if our justifications for the confidence in our beliefs are sufficient to resist conciliation. Perhaps practicing these epistemic skills will be helpful for resolving disagreements (it can't hurt to check our evidence and justifications), but, as the hybrid theories stand, they lack clear guidance on how to apply their methods to cases of disagreement in a way that definitively resolves the puzzle of peer disagreement.

This brief survey of objections is not intended to be exhaustive; it is enough for my purposes here to highlight the fact that each view faces reasonable *prima facie* objections—none of these views are perfect at first glance. Indeed, the massive growth of the literature can be explained, in part, by extensive efforts to refute these objections along with more nuanced ones.

While each view is subject to the respective objections I listed above, each also employs its own general strategy for resolving the puzzle of peer disagreement. In the next section, I discuss those strategies.

### **2.7: Peer Disagreement Response Strategies and the Puzzle as a Trilemma**

The three views I discussed above employ unique strategies to resolve the peer disagreement puzzle. To make these strategies clear, I present the puzzle as a trilemma and then explain how the two main views—conciliationism and steadfastness—respond to the trilemma. I leave the hybrid views out of the analysis since they each could utilize either strategy, depending on the situation. So, focusing on the strategies of the two main

views in this analysis also provides the reader with the options available to those using the hybrid views (while keeping in mind this flexibility is both an advantage and a flaw).

### **2.7.1: Setting up the Peer Disagreement Trilemma**

Typically, a trilemma is composed of three statements that, when taken together, present inconsistencies or contradictions among the combined statements. In a trilemma, if two of the three statements are true, the remaining third statement appears false or unlikely. One way to understand the peer disagreement problem is as a trilemma.

Setting up the problem as an explicit trilemma is part of my original contribution in this dissertation. The trilemma that follows is of my own design and not currently found in the literature (at least not explicitly in any of the works I have reviewed). There are various ways to approach the problem of how one ought to rationally respond to cases of peer disagreement, so I am not arguing that the problem *must* be framed as a trilemma. However, instantiating the problem of peer disagreement requires three conditions to be true. So, this trilemma is a compact way to express the problem that motivates the literature.

The puzzle of peer disagreement can be made salient by demonstrating the tension between three claims: (1) the Uniqueness Thesis is true, (2) genuine disagreement obtains, and (3) epistemic peerhood obtains. For a peer disagreement to instantiate, all three of these claims need to be true. Let me explain each claim in order.

First, according to the literature, the Uniqueness Thesis (UT) must hold. To review, UT can be stated as the following principle: When considering the veracity of  $p$  and one defined body of evidence  $E$  that bears on the veracity of  $p$ ,  $E$  can support at most

one truth value for  $p$ . To say it another way, here's Stewart Cohen's take: "Uniqueness: Given a proposition  $h$ , and a body of evidence  $e$ , there is a unique attitude toward  $h$  that is rational on  $e$ " (2013, p. 101). Or, on Bryan Frances's interpretation:

If two people have the very same body of evidence, and if they adopt different attitudes towards  $B$  (the attitudes: believe it, disbelieve it, or suspend judgment on it), then only one of those attitudes is reasonable. In other words, a pot of evidence can't make two opposing attitudes reasonable: if the evidence shows that belief is reasonable, for instance, then disbelief and suspension can't also be reasonable given the exact same evidence." (2014, p. 177)

**Importantly, UT will be stipulated to hold for the remainder of this project.**

However, it is worth noting that should UT fail to be true (which seems at least plausible), the trilemma may also be resolved by that route.

Second, and perhaps the most overlooked requirement of peer disagreement (an exception here would be Sosa 2010, cf. Ballantyne 2016), is that the disagreement has to be genuine in the following ordinary sense: If two people disagree about  $p$ , they must *really* be disagreeing about  $p$ . While this seems obvious, it is not difficult to find examples from everyday conversations where supposedly disagreeing interlocutors are talking past each other or are unclear about how each other are using the words employed in the conversation. Or, oftentimes, they are just misunderstanding each other's respective claims (see: social media). In other words, a genuine disagreement is one in which the parties are not having a mere verbal dispute or talking past each other due to a miscommunication or misunderstanding about what is at stake in the conversation.

Third, for a *peer* disagreement to occur, the interlocutors must be epistemic peers.

Recall the evidence condition and virtue condition from Kelly:

Let us say that two individuals are **epistemic peers** with respect to some question if and only if they satisfy the following two conditions:

(i) they are equals with respect to their familiarity with the evidence and arguments which bear on that question, and

(ii) they are equals with respect to general epistemic virtues such as intelligence, thoughtfulness, and freedom from bias. (2005, pp. 174-175, **bolding in original**)

As Kelly discusses in a footnote, in certain contexts, the standards for what counts as “equal” may be so demanding that it could be impossible for epistemic equality to obtain between two or more disagreeing interlocutors. He points out that when the demands of equality become very strict, it may be the case that no two people can be epistemic peers because there will always be some slight variation in the available evidence or in their intelligence, etc. (2005, p. 175, fn 11). Later in this dissertation, I will discuss the issue of epistemic equality at length (see Chapter 3). For now, we can be charitable and assume that a slightly permissive standard for what counts as epistemic equality is sufficient for construing the problem of peer disagreement as a trilemma. Additionally, most authors who have contributed to the peer disagreement literature agree that if one person is an epistemic inferior to another (who is, naturally, an epistemic superior), the epistemically inferior person should defer to the epistemically superior person.<sup>12</sup>

---

<sup>12</sup> See Priest 2016 for more on evaluating inferior / superior relationships.



Taking all three of these conditions together, the problem of peer disagreement becomes salient in the following trilemma for any token case of disagreement with two interlocutors on the question of whether  $p$  is true or false, likely or unlikely, etc.:

(T1) The Uniqueness Thesis obtains (stipulated)

(T2) A genuine disagreement occurs where one interlocutor claims  $p$  and the other claims not- $p$

(T3) Epistemic peerhood obtains for the interlocutors on the question of  $p$  (using Kelly's standard definition above)

When conditions T1, T2, and T3 hold simultaneously, something has gone wrong. In simple terms, if two people mutually satisfy the evidence condition and the virtue condition of epistemic peerhood, and if a given body of evidence can support *at most one* conclusion about a particular proposition, and the two people involved are really having a genuine disagreement about the veracity of that proposition, at least one of these two people is wrong about whatever they are disagreeing about. Put differently, if T1 and T3 hold, then T2 should not occur—if two people have the same evidence and virtues and that evidence supports at most one conclusion, then there should be no disagreement. If T1 and T2 hold, then T3 should not obtain—if two people genuinely disagree and the evidence supports at most one conclusion, then they are not epistemic peers. Finally, if T2 and T3 hold, then T1 cannot be true (which may be the case, but I have stipulated here that UT is true)—if two people genuinely disagree and have the same evidence and epistemic virtues, then UT seems to be false. Hence, cases of peer disagreement present us with a trilemma, which motivates the problem of peer disagreement in general.

### **2.7.2: Peer Disagreement Trilemma**

Allow me to state the trilemma in its simplest form:

- (1) If uniqueness is true and there is a genuine disagreement, then the disagreeing parties are not epistemic peers.
- (2) If there is a genuine disagreement and the parties are epistemic peers, then uniqueness is false.
- (3) If uniqueness is true and the parties are epistemic peers, then there is not a genuine disagreement.

### **2.8: General Strategies of the Leading Responses**

#### ***Steadfastness:***

The steadfast strategy is permissive in the sense that one *may* conciliate in certain instances of disagreement, but one is not always required to conciliate in cases of (putative) peer disagreement. Since the steadfast argument usually involves falling back on the claim that we may have private yet disagreement-relevant evidence or self-knowledge that our peers may not or cannot have, this leaves open the question about whether steadfasters believe it is reasonable to believe that we have epistemic peers at all. On one reading of the steadfast response (call it the “demotion” interpretation), we may believe that our interlocutor is our epistemic peer, and, upon discovery of disagreement, we also discover that our private evidence is relevant to the first-order evidence in the dispute. In this case, according to the steadfast, we may demote our previous peer to the status of an epistemic inferior since they lack evidence that bears on the proposition being disagreed upon. On a second reading (call it the “denial” interpretation), if, according to

the steadfasters, we always (or almost always) have an evidential advantage (all the public evidence plus our unexportable private evidence, etc.) in most cases of disagreement, then it may be the case that the evidence condition of peerhood cannot ever practically be met. If this second reading is the appropriate way to apply the steadfast response, then the steadfast may often regard most interlocutors as non-peers. Perhaps the steadfast could assume that those who are peer candidates also have *their own* respective private evidence and this could cancel out in the overall evidence equality calculation, but that is not the line taken by van Inwagen or other similar steadfasters. Their idea is that there is something special about our own private evidence and trust in our own reasoning such that we can't be sure if another person's private evidence or reasoning has the same weight as our own private evidence and our own introspective access to our reasoning. So, without some convoluted algebra to get two people to balance their disagreement-relevant evidence in just the right way, on the denial reading, steadfasters may just deny that they have any epistemic peers at all.

In any case, the important takeaway here is to notice that the steadfast approach—on either the demotion or denial interpretation—solves the problem of peer disagreement by eliminating epistemic peerhood. That is, the steadfast solution to the trilemma is **peer-denying** and **disagreement-preserving**. Relegating an interlocutor to epistemic inferior status removes the threat of T3 (peerhood) from obtaining. Without peers there is no peer disagreement. Problem solved. Since I am holding the Uniqueness Thesis (T1) constant, we only have T2 remaining as the statement to affirm. That is, in the steadfast solution, either the genuine disagreement remains unresolved or the person who has been demoted

to epistemic inferior status must see the error of their ways and defer to their newly-anointed epistemic superior (i.e., they must conciliate or remain in a disagreement).

***Conciliationist:***

On the other hand, the conciliationist response to the problem leaves epistemic peerhood intact while instead breaking the trilemma at the genuine disagreement statement (T2). Recall that conciliationists advise us to “split the difference” as the appropriate rational response to peer disagreement. When considering all-or-nothing belief models, this means (in a simple case) that prior to the disagreement, one interlocutor will believe that  $p$  and the other interlocutor will believe that not- $p$ . For both parties to meet the conciliationist prescription, the only doxastic middle ground available is to suspend judgment about  $p$ . However, when using belief models that employ credences rather than all-or-nothing attitudes to describe the conciliationist response to disagreement, the conciliationist position seems to permit more than just agnosticism as the result of many peer disagreements. Consider a generic case where one party to a disagreement about  $p$  believes that  $p$  with a credence of 0.60 and another party to the disagreement believes that  $p$  with a credence of 0.80. On the conciliationist view, when the interlocutors are considered to be peers engaged in a genuine disagreement (and when UT holds, of course), they are having a peer disagreement—though, they could share the same all-or-nothing attitudes because their credences are both above 0.50. The trilemma, now in effect, still needs to be resolved. The conciliationist approach tells us that the rational response here would be for both parties to “split the difference.” They accomplish this by taking the fact that they disagree and that they believe that they are

epistemic peers as evidence that they should adjust their new credences to, roughly speaking, the average of their earlier individual credences. That is, the rational response in this case will have each party arriving at a credence of 0.70. If both parties make the conciliationism-prescribed adjustments to their credences, thereby each arriving at 0.70 credence, then they are no longer engaged in a genuine disagreement. Returning to the trilemma, we can see that conciliationists aim to solve the problem of peer disagreement by eliminating the genuine disagreement condition (T2). So, contra the steadfast strategy, the conciliationist approach is **peer-preserving** and **disagreement-denying**. Peers who split the difference will no longer disagree, or at least they will disagree “less,” however that works.<sup>13</sup>

With the strategies of the two main views in the literature explained, we can see that the trilemma succinctly captures their responses to the peer disagreement puzzle well. If we deny peerhood on the grounds that there is an evidential asymmetry between us and our interlocutor, we are permitted to remain steadfast in our belief. If we try to preserve peerhood on the grounds that our peer’s opinion should be given equal weight to our own, we ought to conciliate and adjust our attitudes accordingly. In later chapters, I

---

<sup>13</sup> There are several distinct norms for belief revision in play in the literature: (1) all-or-nothing belief models that prescribed suspending judgment when conciliating, (2) credence adjustment models that prescribe moving closer to a peer’s credences when conciliating, and (3) conciliating in such a way that the peers arrive at an exact average of their respective pre-disagreement credences. It is sufficient for my project to point out that whichever norms for belief revision / credence adjustments are employed in a prescription to conciliate will dictate how, precisely, the conciliationist responds to the trilemma. For norms that don’t require arriving at exact averages of credences, we might say that they are partially peer-preserving, but this is exceptionally complicated to sort out. If I move from 0.2 to 0.45 and you move from 0.8 to 0.55, are we still disagreeing? Are there degrees of disagreement? These questions require further research. For now, it is worth noting that when I say that conciliationists aim to preserve peerhood and eliminate the genuine disagreement condition, I mean that the goal of the conciliationist is to settle the disagreement by way of some method of compromise for all parties involved.

use these strategies to explain why getting clear about one's definition of peerhood is key to understanding how to apply the views in the literature.

## **2.9: Summary and Conclusion of Chapter 2**

In this chapter, I reviewed the literature and briefly explained the arguments for each of the core responses to the problem of peer disagreement: conciliationism, steadfastness, and hybrid views. I also briefly presented some of the *prima facie* objections to each of the responses, which demonstrates there is still an open question regarding the epistemic significance of peer disagreement and how we ought to respond to disagreement in general. Finally, I posed the peer disagreement problem as a trilemma and discussed the strategies underlying each of the core responses to the problem of peer disagreement.

While this review is intended to bring the reader into this project with a better understanding of how the peer disagreement literature developed and has attempted to address the epistemic significance of disagreement, I also demonstrated that the literature—especially in its early development with Feldman and Christensen—is concerned with targeting real-world disagreements and is, in part, intended to be applied to actual cases of disagreement. So, consider the following premise as the relevant upshot of this chapter:

**(P1) The normative advice offered in the peer disagreement literature is intended to be applied to real-world disagreements about our cherished beliefs.**

In the next chapter, I discuss what is required—according to the literature—to identify our epistemic peers in more detail.

### **CHAPTER 3: ANALYSIS OF EPISTEMIC PEERHOOD**

In this chapter, I provide an analysis of epistemic peerhood. Recall that the central problem addressed in the peer disagreement literature is summed up by asking the following question: “What is the appropriate rational response when one discovers that one’s epistemic peer disagrees with them?” Naturally, then, it would be prudent if, before we respond to that central question, we ask “How do we determine who counts as an epistemic peer?” If we want to apply the advice given in the literature to real-world cases of disagreement, we’ll need to understand how to accurately identify our peers.

To help us understand how to identify our epistemic peers, I begin with a discussion of what a good definition of peerhood captures in terms of epistemic symmetry and disagreement-relevant epistemic asymmetries. Then, I survey some of the definitions of epistemic peers from the literature. This survey is not intended to be exhaustive, even though it is relatively detailed. Most of the definitions I present cluster around Kelly’s 2005 evidence / virtue model discussed in Chapter 2, though I will also note some variations and additional requirements discussed by other authors that help clarify (or complicate) what, precisely, we need to do to identify our epistemic peers. Next, I review some extant criticisms of epistemic peers found in the literature. These criticisms cast doubt on the likelihood of finding peers in real-world situations. I continue my discussion by examining the suggestion that epistemic peerhood may come in degrees. I then comment on the burdens real-world interlocutors face when managing

track records in peer evaluations. Finally, I conclude with a summary of the main points presented in this chapter and discuss what is required for us to justifiably believe that our interlocutors are our peers.

### **3.1: Epistemic Symmetry and Disagreement-Relevant Epistemic Asymmetries**

Recall that when Feldman (2006) proposed the problem of peer disagreement, he did not use the term “epistemic peer”. Instead, he discussed cases in which interlocutors found that they could reasonably believe that there was some sort of epistemic symmetry between them. While Feldman was skeptical<sup>14</sup> that actual people could often reach the stage of “full disclosure” (see Chapter 2), the puzzle he presented is predicated on the *absence of obvious epistemic asymmetries* among disagreeing interlocutors. Recall from the trilemma I presented at the end of the previous chapter that one of the main strategies (used by steadfasters) to resolve the puzzle of peer disagreement is to deny that one’s interlocutor is an epistemic peer. Here is another way to describe that strategy: if one wants to avoid the skeptical pressure of peer disagreement, then it is one’s job to find disagreement-relevant epistemic asymmetries among the parties to the disagreement. If, upon discovering an epistemic asymmetry and, for instance, one’s interlocutor turns out to actually be an epistemic *inferior* on the matter being disagreed upon, then one need not consider conciliating on that matter (even according to conciliationists). Relatedly,

---

<sup>14</sup> According to Feldman: “The other stage I will refer to as ‘full disclosure’. In this stage, Pro and Con have thoroughly discussed the issues. They know each other’s reasons and arguments, and that the other person has come to a competing conclusion after examining the same information. There are, of course, intermediate situations in which the various pieces of evidence and the arguments are partially shared. Indeed, almost any realistic disagreement is somewhere between isolation and full disclosure. Nevertheless, it will be useful to think about the extreme situations.” (2006, p. 220). Here Feldman is acknowledging that full disclosure is going to be rare in real-world cases, yet he notes that using perfect or near perfect epistemic symmetry is useful as a thought experiment to get the (later named) peer disagreement project off the ground.



understanding epistemic peerhood better involves understanding how the criteria used to describe peerhood also gives us a method and the means to discover and rule out disagreement-relevant epistemic asymmetries among disagreeing parties.

Christensen (2007) presents a case to help us understand what counts as a disagreement-relevant epistemic asymmetry and what does not. I'll call it WATCHES:

To focus in on the symmetry question, let me begin with an admittedly crude analogy: I look at my watch, a one-year-old Acme that has worked fine so far, and see that it says 4:10. Simultaneously, however, my friend consults her watch—also a one-year-old Acme with a fine track record—and it reads 4:20. When she tells me this, it clearly gives me new evidence that her watch is fast: I should not trust her watch as much as I would have before finding out that it disagreed with mine. But just as clearly, I've just gotten new evidence that my watch is slow, and this should diminish my trust in it. In this case, it's obvious that the fact that one of the watches is on my wrist does not introduce an epistemically relevant asymmetry. (Christensen, 2007, p. 196)

Christensen's example here is, in part, a response to Richard Foley's (2001) view that having a first-person perspective of our reasoning processes gives us some license to trust our own beliefs more than we can trust the stated beliefs of others. According to Foley, we are entitled to some intellectual self-trust in epistemic matters simply in virtue of the fact that we understand why we hold our own beliefs better than we understand why others hold their respective beliefs. We can see how this self-trust may permit us to

remain steadfast when presented with disagreements as well.<sup>15</sup> Recall that Fumerton's (2010) defense of steadfastness, noted in the previous chapter, is closely aligned with Foley's views on self-trust in terms of guiding how we might respond to peer disagreements. So, according to both Foley (2001) and Fumerton (2010), self-trust in one's own reasoning is usually sufficient to maintain one's views in the face of controversy. However, using WATCHES as analogy, Christensen's response to the self-trust line of thinking is this: just because we happen to be the person holding a belief (or "watch"), the mere fact that *we* possess that belief does not give that belief any more weight relative to the beliefs ("watches") of others. In other words, according to Christensen and the conciliationist perspective, only epistemically relevant asymmetries give us good reason to resist the pressure to conciliate during a disagreement, and merely being the possessor of a belief does not count as an epistemically relevant asymmetry by Christensen's lights. At this point, a general definition of a disagreement-relevant epistemic asymmetry is in order:

**Disagreement-relevant epistemic asymmetry:** A relation among interlocutors whereby (at least) one party to a genuine disagreement about the veracity of *p* has an epistemic advantage, *a*, over the other party (or parties) to that disagreement such that *a* can potentially explain why the parties disagree about the veracity of *p*.

Let's look closer at this definition of a disagreement-relevant epistemic asymmetry. First, the basics: (1) here interlocutors or parties to a disagreement are just the people involved

---

<sup>15</sup> Foley's view is more complex than I have presented it here, so my quick-and-dirty summary does some violence to it. However, with apologies, I must move on with my analysis here.

in the dispute and (2) these disagreements are genuine disagreements as described in Chapter 2. More importantly—in the context of peer disagreement—what it means for someone to have an epistemic advantage over another person can be derived from the definitions of epistemic peers presented below. These advantages can be evidential, cognitive, environmental, etc. in nature. So, when one evaluates themselves and others to determine if they are in an epistemically symmetrical relationship (i.e., are epistemic peers), one needs to check each factor listed in whatever definition of epistemic peers they are using in order to determine if they are advantaged (or disadvantaged) on any of those factors. For example, suppose you and I disagree about  $p$  and we want to apply some advice from the literature to rationally respond to our disagreement. If you have, say, one piece of disagreement-relevant evidence regarding  $p$  that I do not have, you will have an epistemic advantage over me. Your advantage over me is, paradigmatically in the literature, sufficient to count as a disagreement-relevant epistemic asymmetry among us regarding  $p$ . Hence, we would not be peers due (at least) to your specific (evidential) epistemic advantage over me.

Furthermore, understanding what it means for an epistemic advantage to potentially explain why we disagree helps us understand the peer disagreement puzzle clearly. In the previous example, the fact that we disagree about  $p$  is, at least in part, *explained by* a disagreement-relevant epistemic asymmetry—namely, the epistemic advantage you have by possessing the piece of evidence that bears on  $p$  that I do not have. Another way to put it: If I had that piece of evidence and we were otherwise

epistemic equals, we would not disagree about  $p$ .<sup>16</sup> Epistemic asymmetries can also be present in the balance of epistemic virtues under consideration in peer evaluations. Christensen (2007) discusses a hypothetical case in which two meteorologists who are putatively epistemic peers disagree about a specific weather forecast. In his example, Christensen discusses how a lapse in otherwise reliable epistemic virtues can also introduce a disagreement-relevant epistemic asymmetry that explains why those putative peers disagree: “Given that my friend and I are generally reliable thinkers who have studied the same evidence, the fact that we disagree will be explained by the fact that at least one of us has made a mistake in this case” (2007, p. 198). So, disagreement-relevant epistemic asymmetries defuse the puzzle of peer disagreement by offering a reasonable explanation for why two people might come to disagree; different inputs (evidence) or errors in calculation (virtues) result in different outputs (beliefs).

What we need now is a stable, precise, and detailed definition of epistemic peers. This definition should capture the epistemic qualities and capabilities that we would need to evaluate to discover disagreement-relevant epistemic asymmetries among real-world interlocutors. In the next section, I survey several definitions of epistemic peers to build that definition.

---

<sup>16</sup> Christensen (2007) discusses how epistemic asymmetries explain why parties disagree and how, when no relevant epistemic asymmetries are present, we are left without a good reason to explain why parties disagree. So, according to Christensen, when we cannot explain why two people disagree through pointing out disagreement-relevant epistemic asymmetries, we are only left with conciliation as the appropriate way to respond to the disagreement. See his 2007, section 3: “Explaining Disagreements and Adjusting Beliefs”, pp. 194 – 199 for more.

### 3.2: Definitions of Epistemic Peers in the Literature

One definition of peers that stands out—one we’ve already seen in Chapter 2—comes from Kelly (2005). Kelly gives credit for coining the term “epistemic peer” to Gary Gutting (1982). While Gutting was primarily concerned with discussions about disagreements between religious skeptics and believers, there does not seem to be any good reason to think that his definition cannot be expanded to cover interlocutors who might disagree about other topics. Here is Gutting:

Ordinarily, disagreement about an important matter among inquirers of apparent equal acuity and goodwill is taken to indicate that those who would take a position on the matter need to provide reasons for doing so. What no one has questioned I may, perhaps, take for granted; but where there is significant disagreement, it seems foolish simply to prefer my intuitions to those of others who seem to be my epistemic peers. (Gutting, 1982, p. 11-12)

Gutting continued:

Suppose I find out that a friend, whom I have every reason to believe is as intelligent and reasonable as I, has a belief that I see as totally unwarranted... Rather than question my friend’s rationality or intelligence, I am inclined to think that I do not properly understand what he believes. When I do not understand someone’s belief, I neither share it nor contradict it. (Gutting, 1982, pp. 15-16)

Gutting’s definition of an epistemic peer suggests that epistemic peerhood obtains when someone has “equal acuity” and is “as intelligent and reasonable” as their interlocutor. In

Gutting's definition, we can immediately recognize Kelly's virtue condition for epistemic peerhood I briefly discussed in Chapter 2.

Before we continue examining Kelly's definition of peers, here is an interesting aside: Gutting suggests that if you discover that you disagree with an epistemic peer about some belief, your first reaction will be to think that *you do not understand what they believe*, and that the appropriate rational response would be *to neither share nor contradict that belief*. Recall that the two main strategies from the peer disagreement literature are to do precisely one of those two things: (1) steadfasters permit remaining *contradictory* (disagreement-preserving) in many cases, and (2) conciliationists advise us in most cases to "split the difference", which is just to say that we should *share* the belief of our disagreeing interlocutors (peer-preserving).<sup>17</sup> Gutting's intuition was that we should do neither, which seems to present a third option not discussed in the literature, especially when we realize that simply not understanding the beliefs of others is a common occurrence in many real-world disagreements.

Returning to our investigation of peerhood, Kelly develops Gutting's definition of epistemic peers by adding an evidential requirement to Gutting's approximation of the virtue condition. Let me repeat the relevant lines from Kelly 2005:

---

<sup>17</sup> In all-or-nothing belief models, one could make the case that suspending belief is akin to "neither sharing nor contradicting" a belief. However, most of the peer disagreement literature eventually moves to credence level models of belief in terms of meeting in the middle. According to credence models, interlocutors do not always arrive at a suspension of judgment after conciliating, so my claim about the conciliationist response here is aimed at the common conciliationist prescriptions for doxastic credence revision. Additionally, I think there is a nuanced difference between not properly understanding someone else's belief and suspending judgment on one's own related and perhaps opposing belief due to a disagreement. Pausing to consider why someone holds a belief that seems unwarranted without immediately comparing our own attitude towards that belief is an option that is not often discussed. However, I leave this topic for another investigation.

Let us say that two individuals are **epistemic peers** with respect to some question if and only if they satisfy the following two conditions:

- (i) they are equals with respect to their familiarity with the evidence and arguments which bear on that question, and
- (ii) they are equals with respect to general epistemic virtues such as intelligence, thoughtfulness, and freedom from bias. (Kelly, 2005, p. 12)

As I discussed in Chapter 2, Kelly's two-part evidence / virtue version of the necessary conditions for epistemic peerhood appeared early in the literature. However, Kelly's definition could use some further examination to arrive at a stable, precise, and detailed definition of what, exactly, it means for interlocutors to be epistemic peers. Below I survey the various requirements and conditions that several authors have proposed in terms of what information one needs to have about themselves and other parties to a disagreement to determine if they are epistemic peers.

Bryan Frances (2014) proposes a list of what he calls *disagreement factors* that need to be considered when performing an epistemic peer evaluation:

- Data
- Evidence
- Time [spent thinking about the question]
- Ability [cognitive]
- Background Knowledge [general factual information]
- Circumstances of Investigation [environment] (Frances, 2014, p. 26)

I'll address each of these disagreement factors in turn. First, we have data, which is (perhaps) just empirical evidence. Second, we have evidence of all sorts (which would include data). So, these first two factors are just the evidence condition found in Kelly's definition. Third, we have time spent thinking about the evidence that bears on some proposition. My take on the time factor is that duration of study alone isn't all that valuable when searching for disagreement-relevant epistemic asymmetries.<sup>18</sup> Instead, we should perhaps think of time spent thinking about the evidence as something like care and effort spent thinking about the evidence, which is broadly, to my mind, an epistemic virtue in itself. Fourth, we have cognitive ability, which is partially just the epistemic virtue condition ("intelligence", etc.) restated. So, in general, time and cognitive ability reflect the virtue condition requirements. Fifth, we have background knowledge. This is where Frances's list of disagreement factors begins to require additional targets of evaluation for epistemic peerhood beyond the evidence and virtue conditions.

According to Frances, background knowledge relates to the general knowledge that one has about a subject matter. In his example, if two people disagree about the results of a chemistry experiment through which they both have the same evidence (the experiment) and equivalent epistemic virtues, but one person has a substantial background in studying chemistry while the other does not, this difference in background knowledge is sufficient to deny epistemic peerhood in this case (Frances, 2014, p. 19-20). Frances's background knowledge factor, then, either expands the evidence condition to

---

<sup>18</sup> I am, however, aware of the effects of letting an idea "marinate" or "percolate" in the background of one's consciousness. Sometimes time spent *not thinking* about an idea seemingly transforms and clarifies our opinions about that idea. So, to use a simple metaphor to quickly make my point, the duration from when the seed is sown until the plant bears fruit is necessary but not always sufficient for obtaining fruit.



include any pertinent peripheral or basic evidence that might bear on the veracity of some proposition, or it adds an additional requirement for peerhood altogether. In either case, checking to see if my potential peer has the same background knowledge increases the scope (and burden) of what needs to be considered to evaluate myself and others for peerhood. Sixth (and finally), we have circumstances of the investigation, which, according to Frances's example, is just the environment one is subjected to while considering the disagreement-relevant evidence. According to Frances, if I process my evidence in a noisy, distracting environment and you process that same evidence in a quiet, peaceful environment, we may not be epistemic peers—even when we otherwise satisfy the standard evidence and virtue equality conditions. I think including the environmental factor as a criterion for peerhood in this way depends on how well each person studies and learns in various environments. That is, it could be the case that you learn well and think clearly in noisy environments (perhaps through habituation) while I do not, which would be a feature we would need to know about each other to accurately evaluate each other for peerhood. So, according to Frances, we should add equality of background knowledge and evidence-processing environments (while acknowledging individual differences when considering environment) to the evidence and virtue conditions.

Frances also proposes that epistemic peerhood can obtain when we are not equals on *each* disagreement factor alone; instead, so long as the sum of all of those factors combined are equal among us, we may consider each other to be epistemic peers. For example, if you and I disagree about  $p$ , and I have a good deal of background knowledge

concerning  $p$  while you have stronger cognitive abilities than me, according to Frances, we can still be epistemic peers in that situation. I say more on this below, but first let's look at how Frances defines peers using his disagreement factors.

By way of a brief example, Frances explains how we can identify others as epistemic peers using his list of disagreement factors:

Suppose that, with regard to a certain question, "Is claim B true?," Ned significantly surpasses Jed in several disagreement factors and Jed doesn't surpass Ned in any Disagreement Factors. Under those circumstances, Ned is the *epistemic superior* of Jed with respect to that question and Jed is the *epistemic inferior* of Ned on that question. If they are roughly equal on all Disagreement Factors, then they are *epistemic peers* on that question. (Frances, 2104, p. 43, emphasis in original)

Here, being "roughly equal on all Disagreement Factors" indicates equivalence of the *sum of the disagreement factors* that I discussed above. This method of epistemic peer evaluation—where two people can be *on balance* epistemic peers—is interesting since it has the potential to either expand who counts as a peer (since 1:1 equality for *each* unique factor is not required) or restrict who counts as a peer (since there are so many factors to consider). Recall that the primary reason we would evaluate ourselves and others along the lines of these various disagreement factors is to detect disagreement-relevant epistemic asymmetries. So, another matter to consider here is which factors count as disagreement-relevant in each token case of disagreement. Some cases of disagreement will require intense scrutiny of empirical data (e.g., disagreements about complex

scientific experiments (LHC experiments come to mind)) while others will require vast cognitive resources (e.g., disagreements about the 475th number in the 5-minute number challenge at one of the World Memory Championships). In short, disagreement factors will be weighted differently depending on what the token disagreement pertains to.

However, for real-world cases, requiring that parties to a disagreement sort out how much weight each disagreement factor should have in the overall peerhood evaluation could be quite burdensome. Furthermore, consulting an extensive list of disagreement factors with varying weights introduces a many-disagreement-factors problem: the more factors that need to be evaluated for peerhood, the more opportunities there are to find disagreement-relevant epistemic asymmetries. If this is the case, then there will hardly be any situations where two interlocutors find themselves in an epistemic peer relationship—a worry I discuss below in section 3.3.

Some elements of Frances's list of disagreement factors are implicitly reiterated by other authors, so his proposal is not idiosyncratic. In what follows below, I offer a brief survey of the various ways that epistemic peers are defined in the literature. Some of these definitions echo Kelly's definition and some seem to suggest that we should include additional disagreement factors, like those proposed by Frances, in our peerhood evaluations. Here I begin my brief survey of the definitions of epistemic peerhood proposed by various authors in the literature:

- Hilary Kornblith (2010): Upon discovery of a disagreement:

I find that I have an opinion, but there are others who disagree with me who are, to adopt a useful term, my epistemic peers: they are just as smart as I am, just as

well informed, and have thought about the issue just as long as I have and just as carefully. (p. 31)

Most of Frances's disagreement factors are present in Kornblith's account: (1) "just as smart" is part of the cognitive factor (i.e., virtue condition), (2) "well informed" is part of the data and evidence factors (i.e., evidence condition), and (3) "have thought about the issue just as long as I have" is the time factor (i.e., possibly a virtue condition).

- Catherine Elgin (2010): When epistemic status is not clearly an inferior / superior relationship:

More problematic are the cases in which opponents are, and consider themselves to be, epistemic peers. Then they have the same evidence, reasoning abilities, training, and background assumptions. (p. 53)

Elgin's treatment of peerhood also includes some of Frances's disagreement factors: (1) "same evidence" is the evidence factor / condition, (2) "reasoning abilities" is the cognitive / virtue condition, and (3) "training and background assumptions" are, probably, Frances's background knowledge factor.

- David Christensen (2007): Revisiting the case for conciliationism given in Chapter 2, now with a focus on how Christensen defines epistemic peers in cases of disagreement:

In particular, there are cases where one does not have any special reason to think that the person with whom one disagrees has more (or less) evidence, or is more (or less) likely to react to that evidence in the right way. Suppose I find out that my friend disagrees with me about P: she has moderately high confidence that it's

true, and I have moderately high confidence that it's false. But to the best of my knowledge, my friend is just as well informed as I am - in fact, we may suppose that my friend and I have had long discussions in which we share every bit of evidence we can think of that's relevant to P. And suppose further that I have good reason to believe that my friend and I are equally intelligent and rational, and that I know of no general reason (like the fact that people tend to be biased toward their children) to think either of us is especially likely to be particularly good, or bad, at reacting to evidence on this particular topic—no reason, that is, aside from the fact that my friend disagrees with me about P. In other words, my friend seems to be what some have called an “epistemic peer.” In this sort of case, should I revise my belief? (2007, pp. 188-189)

Christensen's definition also includes “just as well informed” and “equally intelligent and rational”, which are just the evidence and virtue conditions, respectively. However, Christensen also includes a discussion of how “likely” the interlocutors are to correctly react to the evidence under consideration in the disagreement. Later, I discuss one conception of epistemic peerhood that is measured less by comparing evidence and virtues and more by making assumptions about how likely each interlocutor is to correctly (rationally speaking) react to the evidence. For now, I'll resume the brief analysis of Christensen's definition of epistemic peers.

Notice that Christensen also stipulates that the interlocutors have “had long discussions in which we share every bit of evidence relevant to P.” With this additional condition, it appears that only satisfying the evidence condition and the virtue condition

is not sufficient to establish peerhood. Interlocutors wishing to evaluate each other for epistemic peerhood will also need further familiarity with each other in the form of a **track record** to affirm peerhood. I discuss how we might evaluate those track records below.

So, to summarize Christensen's definition of peerhood, peers must satisfy the evidence and virtue conditions, which may be assessed by way of a likelihood-of-being-right comparison. Furthermore, they will need to have track records of their past attempts to respond to questions similar to the question being disagreed upon. I'll return to likelihood definitions and the track record requirement below. For now, let's return to the survey of definitions of peerhood.

- Ralph Wedgwood (2010): On evaluating epistemic peerhood upon discovery of disagreement and a defense of steadfastness:

Even if initially - before you find out that I disagree with you - you rationally thought that I was just as likely as you to be right, the information that I believe  $p$  may *by itself* give you sufficient reason to think that I am probably less reliable than you are.

This... case shows how important it is to be clear about the definition of what it is for someone to count as one of your "epistemic peers." On the one hand, suppose that we say that for you to regard another thinker as your "epistemic peer" (with respect to a given question) is for you to attach an equally high *unconditional* probability to the hypothesis that that thinker will be right about that question as to the hypothesis that you will be right about that question. Then,

even if you start out by rationally regarding me as your epistemic peer with respect to a given question, it may be quite rational for you to respond to the information that I believe  $p$  (which you believe clearly false) by ceasing to regard me as your epistemic peer with respect to that question.

On the other hand, suppose that we say that for you to regard me as your “epistemic peer” with respect to a given question is for you to assign exactly the same *conditional* probability, on the supposition that you and I disagree about the question, to the proposition that I am right as to the proposition that you are right. Suppose that we also assume that the rational way to respond to new information is by some form of conditionalization. Then, if you rationally regard me as your epistemic peer in this sense, it could *not* be rational for you to respond to the information that you and I disagree by concluding that you are more likely to be correct than I am. This point is not in any way a qualification of this way of understanding the epistemic significance of information about other thinker’s beliefs. This interpretation of what it is for you to regard someone as your “epistemic peer” makes it very unlikely that you will regard many people as your epistemic peers. On any less demanding interpretation of what it is for me to be your epistemic peer, it may be quite rational in certain cases for you to downgrade your assessment of my epistemic standing in relation to your own precisely in response to the information that I disagree with you. (2010, p. 236-237)

This passage from Wedgewood requires some unpacking. Three claims are worth discussing further here: (1) Wedgewood is employing a likelihood measure for epistemic

peerhood, a method of peer evaluation mentioned in Christensen above that I discuss later, (2) Wedgwood is comparing unconditional probabilities to conditional probabilities in the peer evaluation, and (3) Wedgwood wonders if it is likely that any two interlocutors can consider themselves epistemic peers (relative to each other) at all on the conditional probability that two or more thinkers consider each other epistemic peers.

The first point is just definitional; one way that the literature recommends evaluating others for epistemic peer status is through evaluating the relative likelihood that a potential peer will be correct with regard to the veracity of, say,  $p$ , as you are with regard to  $p$ . Again, the likelihood view is discussed below.

The second point advances the operationalization of the first point; one way in which we can evaluate others as epistemic peers is to reflect on whether we unconditionally or conditionally assign probabilities to our peer evaluations. According to Wedgwood, if we unconditionally assign probabilities to our peer evaluations, we can enjoy higher confidence that we are correct even if our potential peers disagree with us. On the unconditional view, the discovery of disagreement is evidence that we were *wrong about our peer evaluation*, and we may therefore demote our disagreeing interlocutors to non-peer status. However, if we permit conditionalization (or Bayesian updating, in this case), we may regard our potential epistemic peers as peers even when we discover that we disagree with them because we can adjust our evaluation of peerhood by conditionalizing. So, the formula presented by Wedgwood suggests that, upon encountering a potential peer's disagreement, we can either change our evaluation of our



interlocutor as an epistemic peer because we are *not open to conditionalization* on the matter being disagreed upon (non-conditionalization), or we can leave room to budge on our evaluation of peerhood through conditionalization upon the discovery of that disagreement. This is a notable approach to epistemic peer evaluations because, as discussed in the trilemma in Chapter 2, the steadfast strategy of denying peerhood is still an option in the unconditional probability assignment. That is, on Wedgewood's view, we may demote our potential peers when we are not required to conditionalize on the belief that they are our peers. While we may have to admit that we misjudged whether an interlocutor was our peer regarding a disputed proposition, we do not have to revisit our beliefs about whether they were peers or not once we become aware of how they stand on the matter being disagreed upon. Conversely, if we are willing to take the conditional view that one is an epistemic peer and we discover that they disagree with us, we may still consider them peers and begin the process of conciliation through conditionalizing our peerhood probability assignment upon the discovery of disagreement. That is, we may update our beliefs about the matter being disagreed upon if we find that we disagree with someone whom we believed to be our peer and the disagreement does not give us evidence that the other person is not our peer.

The third and final point concerns Wedgewood's claim that if we must (or are at least inclined to) only conditionally assign a probability of peerhood (or, to the hypothesis that one is a peer), we will often find ourselves without peers. It is unclear why this would be the case. Initially, this line of thinking seems to allude to a conjunction fallacy since it is more difficult to assign an unconditional probability to a hypothesis, H,

than it is to conditionally assign a probability to H (an unconditional assignment is, roughly, just the sum—or conjunction—of the relevant set of conditional probabilities for a given expected outcome). But we can set this worry aside for now because it is not germane to my project. However, what is germane here regarding Wedgewood’s view is that there are sophisticated ways to deploy the peer-denying strategy of steadfasters. One such tactic in that strategy is, according to Wedgewood, to unconditionally assign probabilities of peerhood to epistemic peer evaluations. Given the epistemic demands of assigning unconditional probabilities to any non-trivial belief, this seems like a workaround that requires more bullet-biting than humility. However, the important lesson from Wedgewood’s view is that one can build their peer disagreement position (steadfastness here) into the way in which one evaluates their interlocutors for peerhood, which short-circuits further investigation of the epistemic significance of disagreement. If one desires to demote an interlocutor, one merely needs to declare (to themselves if not to others) that they originally assigned a high *unconditional* probability to the hypothesis that their interlocutor was an epistemic peer—that is, until one discovered that the potential / former peer disagreed with them.

Wedgewood’s approach here demonstrates that the methods we use for determining who our epistemic peers are can be front-loaded to push us towards (or away from) one of the positions on the appropriate rational response to disagreement in the literature. On the one hand, this is not surprising; perhaps how one defines their epistemic peers determines how much significance (if any) one gives to peer disagreement. However, on the other hand, one could also reverse-engineer a way to determine who

one's peers are by starting with their preferred position (steadfast or conciliationist) and work backwards to a conception of peerhood that is congenial to that position. Here is my point: a general problem for the literature is that, in the absence of a standardized method for performing epistemic peer evaluations, we are permitted to massage our definition of peerhood to accommodate the position that we find most intuitively plausible. So, the flexibility and perhaps ambiguity present in epistemic peer definitions might be affecting the intuitive response to idealized peer disagreement cases—how clearly and precisely one defines the conditions for epistemic peerhood is crucial to using the advice in the literature.

- Richard Fumerton (2010): On encountering disagreement with another and evaluating their epistemic standing:

Furthermore, (and crucially) I have no more reason to think that their evidence is any worse than the evidence upon which I relied in believing my initial conclusion, *nor is their ability to process the relevant evidence*. I also realize, in effect, that there is a perfect *symmetry* in our epistemic situations with respect to one another. (p. 97, emphasis in original)

In contrast to Wedgwood's carefully considered methods for evaluating epistemic peerhood, Fumerton's conception of epistemic peerhood is exceptionally straightforward and fits with the standard evidence condition / virtue condition version that Kelly developed. However, the reason I've included this quote from Fumerton is to demonstrate that there is not a consensus about how *detailed* a definition of epistemic peers ought to be. Recall that Fumerton supports a steadfast view. Also, consider that Fumerton's

definition of epistemic peers not only closely aligns with Kelly's view, but also with Feldman's view of epistemic symmetry. The point here is, in some ways, the opposite of the point about Wedgewood's more complex definition of epistemic peerhood; a steadfast such as Fumerton (see Chapter 2) can hold a definition of peerhood that is nearly identical to that of the conciliationist and reach a completely different conclusion about the rational response to disagreements. In short, one does not *need* a 'bespoke' definition of epistemic peerhood to arrive at either (or any) of the main positions developed and defended in the literature.

- Stewart Cohen (2013): Hedging the definition:

More generally, non-experts should defer to experts about matters within their area of expertise. This is straightforward.

Matters are considerably less clear when parties to the dispute have the same evidence. Of course no two people ever share exactly the same evidence. But in many cases, there is enough shared evidence that there is no reason to suppose that either party to the dispute is in an evidentially superior position... A special case of this problem arises when the parties to the dispute are in general equal in their reasoning abilities, or at least, close enough so there is no basis for supposing either party is in general the superior reasoner. When parties to a disagreement have the same evidence and are equal in their reasoning abilities, they are epistemic peers. (2013, p. 98)

Cohen's take on epistemic peerhood here expands, in a sense, the definition of peers.

When he says that "[N]o two people ever share **exactly** the same evidence" and that the

two interlocutor's respective reasoning abilities are "close enough" rather than equal, this indicates that the equality requirements for epistemic peers could be a bit looser than some of the stricter versions presented above that require tight parity or equality of evidence and virtues. This is an important development in the literature because the boundaries of peerhood have slowly expanded to include people who are not only positively identified as peers but also those who are negatively defined as not obviously superiors (or inferiors). However, as I discuss later (below and again in more detail in Chapter 6), loosening the requirements of peerhood brings problems of its own.

This sampling of the literature's definitions of epistemic peers gives us variations on a theme. However, there are some differences in the way each writer conceives of epistemic peers. Importantly, since determining peerhood is a requirement for applying the prescriptions given in the literature to actual disagreements, concerns about which version or definition of epistemic peerhood one should use to evaluate the epistemic standing of their interlocutors arises.<sup>19</sup> Do we need to consider all of the disagreement factors listed by Frances each time we want to evaluate the relative epistemic standing of ourselves compared to others? If so, and of special interest here: Should we be confident that we are competent to evaluate all of these disagreement factors for ourselves and others in real world cases of disagreement? These are questions that are rarely discussed

---

<sup>19</sup> Of course, one *could* apply the prescriptions from the peer disagreement literature without any sort of peer evaluation, but this willy-nilly approach to using the peer disagreement advice defies any principled defense of why one is choosing to remain steadfast or conciliate. That is, the advice from both camps is probably suitable in many cases of disagreement, but if one doesn't perform the peer evaluation before one decides how to respond to the disagreement, one is not using the methods produced by the conceptual framework of the *peer* disagreement literature; in these cases, one is merely deciding if it is reasonable to change their mind or not in the face of any old disagreement, which, without an epistemic peer evaluation, is a completely different method for rational disagreement resolution that falls outside of the peer disagreement project.

further in the literature, even though the answers to these questions are crucial for the literature to be useful in real-world settings.

Furthermore, the approaches that employ “likely” or “close enough” definitions of peers struggle to be useful in real-world situations because they often don’t help us discover pertinent disagreement-relevant epistemic asymmetries. I discuss these approaches to defining peerhood in the following section.

### **3.2.1: Likelihood and Dispositional Definitions of Epistemic Peers**

There are two notable approaches to defining epistemic peers. Each approach attempts to allay the practical burdens of asking actual people to evaluate several disagreement factors—including the standard evidence and virtue conditions—to determine peerhood in real-world cases of disagreement. One approach to defining epistemic peers that has been briefly discussed above regarding both Christensen’s and Wedgwood’s *likelihood approach*. This approach defines peers as those who are just as likely to get the answer to a question right (or wrong) as you are, or some other version of “just as likely” evaluations. Here is Elga’s justification for using this model:

My use of the term “epistemic peer” is nonstandard. On my usage, you count your friend as an epistemic peer with respect to an about-to-be-judged claim if and only if you think that, conditional [on] the two of you disagreeing about the claim, the two of you are equally likely to be mistaken. On more standard usages, an epistemic peer is defined to be an equal with respect to such factors as “intelligence, perspicacity, honesty, thoroughness, and other relevant epistemic

virtues” (Gutting 1982, 83), “familiarity with the evidence and arguments which bear on [the relevant] question”, and “general epistemic virtues such as intelligence, thoughtfulness, and freedom from bias” (Kelly 2005). In defense of my use, suppose that you think that conditional on the two of you disagreeing about a claim, your friend is more likely than you to be mistaken.

Then however intelligent, perspicacious, honest, thorough, well-informed, and unbiased you may think your friend is, it would seem odd to count her as an epistemic peer with respect to that claim, at least on that occasion. You think that on the supposition that there is disagreement, she is more likely to get things wrong. (Elga, 2007, p. 499 en 21)

On the face of it, this simplifies matters. We no longer need to evaluate a long and complicated list of disagreement factors or have intimate knowledge about the other person’s access to and processing of the available evidence to determine if they are our peers. However, these likelihood-style definitions of peers don’t elaborate on *how we are to arrive at these likelihoods*. When Elga says “conditional on the two of you disagreeing about a, your friend is more likely than you to be mistaken”, it isn’t clear why—other than the mere fact that you disagree with your friend—how you would assess the likelihood that your friend is more likely than you to be mistaken. My guess is we would need to have some way to justify our likelihood estimates by evaluating things like how intelligent and reasonable the other person is, or what kind of evidence they have had access to, or how often they’ve been right about similar matters in the past, etc. So, when one believes that their interlocutor is just as likely to be right about  $p$  as they are (for

peerhood to obtain), one ought to have specific *reasons* why one believes that their interlocutor is just as likely to be right about  $p$  as they are. The idea that one can sense or intuit who their peers are based on a superficial feeling of likelihood is clearly not what is intended by this approach to describing peerhood. It is only reasonable to think of the likelihood definition of peerhood as a shorthand way of taking into account some (or perhaps all) of the criteria for peerhood listed above to inform one (or, to update one's priors, if that is more acceptable for the Bayesians) who one's peers are and are not. In other words, it will be exceptionally difficult to determine if there are any particular disagreement-relevant epistemic asymmetries that explain why two people disagree if we are only basing peerhood on how likely a potential peer is to be right or wrong about some question. It isn't clear how the mismatch in likelihoods of being right (or wrong) alone is evidence that two people are (or are not) peers. For instance, if it turns out that we think our interlocutor is less likely to be right about the disputed question than we are, how would we explain this likelihood without referring to some disagreement-relevant epistemic asymmetry such as a differential in evidential positions or virtues? Imagine telling an epistemic inferior on  $p$  that they just seem less likely to be correct about the veracity of  $p$  compared to you, full stop. Not only would this superficial approach be unreasonable and unhelpful, it invites dogmatism. Thus, the likelihood approach doesn't help us simplify the process of epistemic peer evaluations when the reasons or justifications for those likelihood assessments are not spelled out.



Another alternative approach to defining peers comes from Juan Comesaña (2012). He developed a dispositional definition of epistemic peers that could also resolve the many-disagreement-factors problem. In Comesaña's words:

The dispositionalist view has it that whether we take somebody else as a peer regarding a certain proposition is revealed in our dispositions to change opinions about that proposition, rather than in our beliefs about the other subject... it is the claim that what it is to consider somebody else a peer should not be understood in terms of our explicit beliefs about her, but rather in terms of our dispositions to change our opinions when a disagreement with her arises. The belief and the dispositions are not, of course, unrelated, but no direct reduction is being advocated here. (2012, p. 239)

According to Comesaña, instead of searching through our beliefs about the disagreement factors possessed by us and our interlocutors, we, counterfactually it seems, may consider an interlocutor to be an epistemic peer if we would be disposed to change our opinions about  $p$  if we were to disagree with them about  $p$  in a specific set of circumstances. This approach may seem promising at first since we are, as in the likelihood approach, relieved of the burden of evaluating and judging a long list of disagreement factors to determine how we ought to rationally respond to a disagreement. That is, under the dispositionalist view, we no longer need to check our (non-dispositional?) beliefs about our potential epistemic peers to determine if they are peers; we need only to understand how we would be disposed to respond should our potential peers disagree with us on a certain question in a specific set of circumstances. Comesaña explains the dispositionalist

approach through the following case: Suppose you and I are watching a horse race and there are three horses to watch: horses A, B, and C. Say that horse A and B finish the race at nearly the same time. However, also suppose that horse C was several lengths behind horses A and B when A and B crossed the finish line. On this view, if I believe that horse A won the race and if you were to claim that horse B won the race, I might be disposed to think of you as my peer on the matter of the winner of the horse race, i.e., I might be disposed to change my opinion or suspend judgment in that specific case. Alternatively, if I believe that horse A won the race and if you believe that horse C won the race, I would not be disposed to think of you as a peer on the matter of which horse won the race since it is apparent that either A or B won the race while C definitely did not.

On this view, how I am disposed to respond to a disagreement with you about specific cases informs me about whether you are my epistemic peer. However, the question still remains: *Why* are we in a disposition to treat someone as an epistemic peer if we discover that we disagree with them? As with the likely to be right / wrong view, it seems that we would need further reasons to explain our dispositions rather than just observing our behavioral response to disagreement without consideration of the beliefs that justified that response.<sup>20</sup> Comesaña's definition of peers makes sense only when we

---

<sup>20</sup> It is worth noting that one of the features of Comesaña's view is that we can remain steadfast without demoting our epistemic peer to epistemic inferior. On the dispositional peer view, if we would not be disposed to change our opinion if we disagreed with S about *p* in a specific set of circumstances, and S disagrees with us about *p* in those circumstances, we are not disposed to responding to them as though they were a peer on *p* in those circumstances in the first place. That is, the nature of the disagreement and the strength of our confidence in our beliefs prior to disagreement determine if one is a peer or not. This is similar to both Kelly's Total Evidence View and Lackey's Justificationist View but focuses on our dispositions to concur or not with our interlocutors rather than focusing on our evidence (Kelly) or level of justified confidence (Lackey). Whether this is a successful shift away from the main steadfast response to the trilemma, which seeks to find a way to demote one's prior or potential peers to resolve the puzzle of

are justified in believing that our *interlocutors*' beliefs are way off. But what about those cases when *we* are very wrong, and our interlocutor isn't? It could be us who thinks either Horse A or Horse C won the race in the above example. Our disposition in that case would have us reject the other party's opinion (that Horse A or B won) *because* we were wrong, which seems completely backwards. In other words, using the dispositional view as a guide for identifying peers fails to meet the sole criterion for peer evaluations: to sift out disagreement-relevant epistemic asymmetries. In the case where we are way off—which is due to some sort of epistemic asymmetry—our disposition to *not* treat our interlocutor as a peer (or, better yet, as a superior) is a failure of rationality that seems permitted on the dispositionalist view. So, the dispositionalist view, like the likelihood view, is not without significant problems.

Additionally, and more broadly, how we define peerhood frames which or what kind of disagreement-relevant epistemic asymmetries we count as sufficient to deny peerhood. However, it is unclear if we, as real-world interlocutors, *can* perform the necessary evaluations demanded by the criteria set by whichever definition we choose. That is, regardless of whether we use the standard (Kelly), the expanded disagreement factor list (Frances and others), the likelihood (Elga and others), or the dispositional (Comesaña) definition of epistemic peerhood, there is still a looming question: Should we be confident that we have any real-world epistemic peers in the first place? In the next section, I discuss some of the criticisms of epistemic peers noted in the literature that question whether we will ever encounter epistemic peers in the real world.

---

peer disagreement, would require further analysis. Still, Comesaña's approach is at least creative even if it does appear circular.

### 3.3: Extant Criticisms of Epistemic Peerhood

The various definitions of epistemic peers given above repeat the requirement for epistemic symmetry of some sort between two disagreeing parties. Philosophers are often wary about claims that any two real objects or phenomena are *equal*. “Equal” can refer, in one sense, to a strict identity relation, or, in a less restrictive sense, to things that are merely the same. The relational referent of “same”, in turn, has been widely discussed in various pockets of philosophical literature, but here we are only concerned with what the authors mean by “same” in the context of the targets of peer evaluations in disagreement situations.

#### 3.3.1: Concerns about Epistemic Peerhood

Kelly (2005) includes a telling footnote for his original definition of epistemic peerhood, which is the standard definition we began with in Chapter 2:

It is a familiar fact that, outside of a purely mathematical context, the standards which must be met in order for two things to count as *equal* along some dimension are highly context-sensitive. Thus, inasmuch as classes of epistemic peers with respect to a given issue consist of individuals who are ‘epistemic equals’ with respect to that issue, whether two individuals count as epistemic peers will depend on how liberal the standards for epistemic peerhood are within a given context. That is, whether two individuals count as epistemic peers will depend on *how much* of a difference something must be in order to count as a *genuine* difference, according to the operative standards. In the same way, whether two individuals count as ‘the same height’ will depend on the precision

of the standards of measurement that are in play... Of course, given sufficiently demanding standards for epistemic peerhood, it might be that no two individuals *ever* qualify as epistemic peers with respect to any question. (Perhaps there is always at least some *slight* difference in intelligence, or thoughtfulness, or familiarity with a relevant argument.) Similarly, it might be that no two individuals count as the same height given sufficiently demanding standards that we employ in assessing intelligence and thoughtfulness (like the standards that we employ when measuring height) are sufficiently liberal to allow individuals to qualify as equal along the relevant dimensions. (Kelly, 2005, p. 175, fn 11, emphasis in original)

This footnote is telling because Kelly questions whether any two interlocutors can ever satisfy the equality or sameness requirements for evidence and virtues. That is, even the *standard definition* of peerhood is problematic if one wants to identify peers in real-world settings because, if that definition is too strict, “it might be that no two individuals *ever* qualify as epistemic peers with respect to any question” (Kelly, 2005, p. 175, fn 11, emphasis in original).

In addition to Kelly’s initial concerns about the equality of the necessary conditions (evidence and virtue) for peerhood, and his admission that it may be the case that there are no true epistemic peers in the actual world, other authors have expressed concern about the nature of epistemic peers. For instance, consider Hawthorne and Srinivasan’s (2013) footnote on the matter:

Typically what is meant by ‘peer’ in the disagreement literature is opaque, in part because the relevant notion of evidence used in glosses on peerhood is not fully spelled out. For example, if knowledge is evidence and peers by definition possess the same (relevant) evidence, then disagreements in which one person knows and the other does not are *ipso facto* not peer disagreements (Hawthorne and Srinivasan, 2013, p. 14, fn15).

This point by Hawthorne and Srinivasan is important to consider because on a Williamsonian account of knowledge, the evidence condition for peerhood also becomes a comparative knowledge condition. This *knowledge* parity definition of peerhood requires that one subscribe to the view that one’s knowledge just is one’s evidence, but the general problem here is not as idiosyncratic as it appears to be. One could have alternative views about knowledge (standard JTB, contextual, etc.) or specific views about how evidence functions in our epistemic judgments (externalist, evidentialist, etc.), but a problem still remains: How can one have the same evidence as another person about *p* and disagree with that person about *p*?

Notice that this is just a restatement of the puzzle of peer disagreement, but the stakes are raised to include equal knowledge. If knowledge is factive, the idea that two people can be peers and disagree in the ways described in the literature falls apart because there can be only one fact of the matter about the truth value of a given disagreed-upon proposition. However, it is important to note that the peer disagreement project does not require that peers possess the same knowledge regarding the disagreed upon proposition. Such a requirement would immediately halt the process of peer

disagreement resolution because, analytically, it is impossible for two people to have the same exact knowledge that  $p$  and genuinely disagree about whether  $p$ . So, while the instantiation of a peer disagreement doesn't require that interlocutors disagreeing about  $p$  have the same knowledge that  $p$ , it is unclear how likely epistemic peerhood is in terms of epistemic symmetry even when we are considering (mere) justified beliefs in the place of knowledge.

Furthermore, there are additional criticisms of epistemic peerhood in the literature that aren't hidden in footnotes. Feldman and Warfield (2010) open their co-edited book on disagreement with the following concern about whether we can actually have or reasonably identify our epistemic peers:

In the stipulative sense of “peer” introduced, peers literally share all evidence and are equal with respect to their abilities and dispositions relevant to interpreting that evidence. Of course, in actual cases there will rarely, if ever, be exact equality of evidence and abilities. This leaves open questions about exactly how conclusions drawn about the idealized examples will extend to real-world cases of disagreement. (Feldman and Warfield, 2010, p.2)

Again, as in the case with Kelly's footnote above, we have prominent authors in the literature doubting whether any real-world interlocutors can meet the epistemic symmetry or equality definitions of peerhood.

Picking up on Feldman and Warfield, Tim Kenyon (2020) writes:

Epistemic peerhood is an idea critical for contemporary EoD [epistemology of disagreement] because it defines a puzzle or family of puzzles that has animated a

considerable proportion of that literature. The puzzle is that of what to believe in a case where agents standing in an important sort of epistemic symmetry relative to some area of judgement nevertheless disagree about an issue falling within that area. Thinking in terms of epistemic peers could be a heuristic that helps us reason creatively about what, if anything, distinguishes our epistemic position from that of our interlocutors – it could help us to think about symmetry-breakers (Kenyon 2018). But that’s rarely how it’s been used. In fact EoD has tended not only to idealize epistemic peerhood in the ways noted by Feldman and Warfield, but to treat the satisfaction of known (sometimes “revealed” or “acknowledged”) epistemic peerhood as a background condition for the core phenomenon of interest. Jointly this raises the prospect that the core phenomenon of interest to EoD is virtually non-existent.

There is a lot to say about epistemic peerhood, and I am not going to attempt to say it here. It is enough to note that the idealized nature of epistemic peerhood in the EoD literature has often been flagged as a concern, even if only to be set aside in many instances. (Kenyon, 2020, p. 1-2)

Here, Kenyon reiterates the doubts expressed by Feldman and Warfield concerning the application of the peer disagreement literature’s advice for real-world cases of disagreement. Furthermore, he points out—ten years after Feldman and Warfield—that the disagreement project has yet to correct for the dearth of interest in epistemic peerhood as the motivation for the core phenomenon (peer disagreement and the responses to it).



The problem of finding a working definition of peerhood that is amenable to real-world application still stands.

In response to the worries that idealized definitions of epistemic peerhood are too strict in terms of evidential and virtue parity for any two real-world interlocutors to satisfy, some authors have attempted to loosen the requirements for peerhood. However, as I discuss below, these looser definitions of epistemic peerhood—definitions that might allow for more people to be peers—are problematic in terms of generating the puzzle of peer disagreement in the first place.

### **3.3.2: Criticisms of Epistemic Peerhood: Loose vs. Strict and Degrees of Peerhood**

As discussed above, some authors have been careful to hedge the definition of epistemic peerhood by adding “roughly” (Christensen) and “close enough” (Cohen) when discussing the requirements for epistemic equality or symmetry. So, it is worth exploring whether these hedged definitions of peerhood are sufficient to help us detect disagreement-relevant epistemic asymmetries in real-world cases of disagreement.

Generally, a definition of epistemic peers that is too strict will suffer from the concerns about equality or sameness that Kelly discussed in his footnote to his definition of epistemic peers. However, a definition of epistemic peers that is too loose or permissive will suffer from not capturing differences in evidential positions or epistemic virtues among the parties under consideration for peerhood. That is, definitions that are too permissive will not be conducive to highlighting the disagreement-relevant epistemic asymmetries that explain why two people disagree. Hence, in real-world epistemic peer evaluations, a loose definition of peerhood may end up fooling us into thinking that we

are having a peer disagreement when the real reason(s) for the disagreement can be found in the fact that there are undiscovered disagreement-relevant epistemic asymmetries that are sufficient to explain the disagreement.<sup>21</sup>

Christensen (2007) distinguishes between full peers, near-peers, and not-so-near peers (p. 204, fn. 20) as a way to relax the equality requirements for peerhood. His proposal for these cases—ones that he claims better resemble real-world cases of potential peerhood—is that we might treat disagreements with those who are close to being peers as though they are peers. One of Christensen’s examples of this “near-peer” relationship: You and your friend conduct the same poll on different populations, but your sample size is somewhat larger than your friend’s sample (2007, p. 211-213). If you and your friend disagree about  $p$ , and the main evidence for  $p$  is the polling data, you may be tempted to declare that your friend’s smaller sample size is sufficient to generate a disagreement-relevant epistemic asymmetry—one that explains why you and your friend disagree about  $p$ . However, Christensen argues that merely having a smaller sample is perhaps *not enough* of an evidential difference to generate a disagreement-relevant epistemic asymmetry in the evidence condition comparison. In these cases, according to Christensen, you would still be wise to conciliate even though your friend, due to her smaller sample size, may not be fully your peer in terms of evidential equality.

---

<sup>21</sup> In his own analysis of epistemic peerhood, Kirk Loughheed (2020) discusses at length the loose vs. strict definitional problem for conceptualizing epistemic peerhood. Loughheed, following King (2012), argues that definitions of peerhood that are too strict limit its use in everyday applications because neither of the conditions (evidence or virtue) for peerhood will ever be met. Additionally, he argues that definitions of peerhood that are too loose are too permissive will not capture relevant epistemic asymmetries that would help us decide who is in a position to have better justified beliefs about whatever is being disagreed upon. This argument is the conjunction of Kelly’s footnote-worthy worries about epistemic equality and Christensen’s explanation of disagreement-relevant epistemic asymmetries tied together in a clear and accessible detailed description of the problem.

Additionally, there will be times when one's interlocutor is not obviously an epistemic superior, but also may be a "not-so-near-peer", as in cases where you are fairly sure that you have spent long hours carefully studying the arguments and evidence that bear on the veracity of some proposition and you believe that your interlocutor—while familiar with the same arguments and evidence—has not attended as carefully to the disagreement-relevant evidence as you have.

Christensen goes on to say that cases where we slightly relax the disagreement factor equality requirements are more similar to real-world cases of disagreement. That is, and quite importantly, Christensen believes that whatever lessons we learn from our intuitions about the relaxed cases are how we move from the idealized cases of evidential and virtue equality to cases with slight epistemic asymmetries as cases that better resemble the conditions of real-world disagreements:

[O]ften one won't have much of a specific idea how well informed other people are. Still, this does not eliminate the evidential value of their beliefs. *Pace* Foley, I should take the beliefs of others into account in many such cases, for in many cases, where I'm not at the well informed end of the general spectrum of people, it is only reasonable for me to expect that a fairly random person has a level of evidence which makes her opinion a valuable epistemic resource. It seems that the lessons of the pure evidential equality cases generalize quite widely...

In sum, then, it turns out that the lessons of the artificially pure cases apply to a great many ordinary situations. In general, unless one has reason to consider oneself to be in a highly privileged epistemic position—both with respect to the

evidence and with respect to one's ability to respond correctly to that evidence—the disagreement of others will provide good reason to revise one's beliefs.

(Christensen, 2007, pp. 212-213)

Recall that Christensen proposed that disagreement-relevant epistemic asymmetries explain why two people disagree, and when that happens, there is no puzzle of peer disagreement. Here, in defense of his conciliationist view, he seems to backpedal to a position that permits some disagreement-relevant epistemic asymmetries (beyond strict equality at the very least) to be overlooked when considering whether one ought to conciliate or not. The method for applying the literature's advice to real-world cases of disagreement now looks like this: start with a case of disagreement with idealized epistemic peers and note your intuitive response to those cases. Then, loosen the requirements for peerhood such that those cases of peerhood resemble real-world cases of disagreement. Next, check your intuitions again and they should be nearly the same as they were when you responded to the idealized case—after all, we've only permitted slight disagreement-relevant epistemic asymmetries to obtain in these near-peer closer-to-real-world-cases. Those latter intuitions should then guide you in how you ought to respond to real-world disagreements. However, there are two problems with this approach: (1) steadfasters and conciliationists have different intuitions to many of the same idealized cases, so those differing intuitions will carry over to the relaxed cases quite easily (so, in terms of deciding which position in the literature one should take advice from, the relaxed cases are just as much of a wash as the idealized cases), and (2) now interlocutors must not only evaluate each of the (perhaps many) disagreement

factors in a peerhood evaluation, they must also decide if the disagreement-relevant epistemic asymmetries are *severe enough* to explain away the disagreement.

One might argue that by loosening the epistemic equality requirements, the burden on the interlocutors is actually allayed since the looser definitions of epistemic peers do not require precise awareness of exactly what pieces of evidence a person has encountered and do not require detailed information about a person's intelligence or acumen in certain subjects. This is a reasonable response since it solves two problems at once: the problem of strict epistemic equality noted by Kelly and others as discussed above, and the problem of overburdening real-world interlocutors with too many necessary targets of evaluation for peerhood that I've been hinting at above. I return to the topic of strict vs loose definitions of peerhood in Chapter 6. For now, it is worth noting that the only way we can reasonably apply the advice from the peer disagreement literature is to loosen any strict equality requirements for peerhood.

One proposal to loosen the equality requirements for peerhood comes from Axel Gelfert (2011). First, Gelfert noticed the dearth of interest in the literature in determining who our peers are and are not:

What the preceding discussion suggests is that determining whether two disputants are indeed epistemic peers—and, if they are not, identifying in what ways their relationship falls short of peerhood—has primacy over the question to what extent the mere fact of disagreement offers a *pro tanto* reason for each party to adjust their credences. I wish to take a first step towards developing a richer notion of epistemic peerhood, before arguing... for a plausible connection

between peerhood and questions of epistemic value, which so far appear to have been ignored in the philosophical debate about disagreement. For now, I wish to argue that our understanding of peerhood can be enriched by recognizing that peerhood comes in degrees. (2011, p. 512)

Gelfert then borrows the idea that epistemic peerhood can come in degrees from Wald (2009), and presents the descriptions of the degrees as follows:

(1) **Close Peers:** Those who (i) assess the evidence pertaining to a particular topic in similar ways to you, and (ii) have a good reason to believe that both parties have similarly good track records in forming true beliefs on the basis of evidence.

(2) **Distant Peers:** Those who fail to meet one of the conditions (i or ii) of Close Peers.

(3) **Remote Peers:** Those who fail to meet both of the conditions (i and ii) of Close Peers.

Gelfert added a further comment:

The significance of being aware of one's own epistemic predicament—not least with respect to one's larger epistemic environment—is not adequately reflected by traditional definitions of epistemic peerhood in terms of either (individual) epistemic virtues or mere reliability on a given occasion. In most cases of persistent disagreement, the relative epistemic standing of the disagreeing parties is far from self-evident. Rather than taking epistemic peerhood for granted and battling over the correct normative response to prima facie instances of disagreement, epistemologists would be well-advised to pay greater attention to

the causes of disagreement and its persistence, and to the many ways in which peerhood can be undermined by tacit commitments or failure of reflective awareness of one's own epistemic predicament. (p. 514)

Gelfert's proposal that peerhood might be better framed as coming in degrees is motivated by two points given in the quotes above: (1) the literature generally lacks any substantive realistic discussions about how people might come to justifiably believe that they can find any real-world epistemic peers in the first place, and (2) the literature focuses on defending various normative positions without consideration of the descriptive limitations inherent in real-world peerhood evaluations. If Gelfert's proposal can help us address those two concerns about the applicability of the peer disagreement literature's advice, then his model of epistemic peerhood as coming in degrees merits further investigation here.

For Close Peers, there are two requirements: (i) interlocutors must assess the disagreement-relevant evidence in similar (but not equal) ways, and (ii) interlocutors must believe that they each have comparably good track records of reliably forming true beliefs from the given evidence. The first requirement is slightly different from Kelly's evidence condition since it does not require evidential equality. Instead, the first requirement only demands that potential peers assess the evidence in similar ways. Taken at face value, this seems to be an oversight; without evidential parity, we would be justified in being suspicious that a disagreement-relevant epistemic asymmetry could be present among our interlocutors. For instance, if you believe that  $p$  and I believe that not- $p$ , and we come to disagree, it could easily be the case that the reason we hold different

beliefs about  $p$  merely comes down to one of us possessing a crucial piece of  $p$ -relevant evidence while the other does not. If we're both trying to figure out why a car engine won't start, the person who realizes that the battery is missing will be the epistemic superior to the person who hasn't looked under the hood. So, to be charitable, let's modify Gelfert's proposal to include the following in his (i) under (i\*): Those who assess *the same and only the same evidence* pertaining to a particular topic in similar ways... This modification in (i\*) gives us, roughly, the evidence condition, which adequately maintains the appropriate level of vigilance for disagreement-relevant epistemic asymmetries in our epistemic peerhood definition. Both (i) and (i\*) also give us part of the virtue condition since assessing or processing evidence in "similar ways" can be understood as shorthand for meeting the parity requirements for the cognitive, time, and circumstances of investigation disagreement factors. Or, put differently, how we assess the evidence we are given is generally guided and constrained by how well our epistemic virtues are functioning at the time. I discuss the connection between evidence and virtues in Chapter 6, but for now, it is sufficient to point out that on the (i\*) reading of Gelfert's (i) condition for Close Peers, both the evidence condition and the virtue condition are present as targets of evaluation for epistemic peerhood.

For the second requirement for Close Peers, Gelfert proposes that the interlocutors have "similarly good track records in forming true beliefs on the basis of evidence" (see (ii) above). Recall that in Christensen's paradigmatic case of epistemic peers above, he sets up his cases so that "[T]o the best of my knowledge, my friend is just as well informed as I am - in fact, we may suppose that my friend and I have had long



discussions in which we share every bit of evidence we can think of that's relevant to P.” (Christensen, 2007, p. 188). This intimate familiarity between interlocutors is repeated in his RESTAURANT case (from Chapter 2): “Let us suppose that my friend and I have a long history of eating out together and dividing the check in our heads, and that we've been equally successful in our arithmetic efforts: the vast majority of times, we agree; but when we disagree, she's right as often as I am.” (Christensen, 2007, p. 193). It isn't immediately clear if Christensen's track record is the same sort of track record that Gelfert describes for his (ii) above. One could read Gelfert's proposal as requiring that the interlocutors have independent yet comparable records of assessing evidence and drawing reasonable conclusions from that evidence. On this reading, you and I might be able to meet requirement (ii) if I have demonstrated that I have taken evidence pools A, B, C, and D and arrived at a set of justified beliefs about propositions relevant to those evidence pools while you have taken a different set of evidence pools E, F, G, and H and similarly arrived at a set of justified beliefs about propositions relevant to those distinct evidence pools. However, in this case, our respective evidence pools and ensuing beliefs may not be related at all here and we could meet Gelfert's (ii). That is, we could both have good track records of being reasons responsive on disparate topics and meet part of the requirements for being Close Peers. On another reading—one that is closer to what Christensen has in mind for track records in his ideal definition(s) of epistemic peers—for Gelfert's (ii) to be met, our track records would not just be a longitudinal evaluation of our proper belief forming habits in general but instead directly tied to how well we assess the *same* pools of evidence and arrive at justified beliefs over time. In this case, we

must have both assessed, say, evidence pool A and both arrived at reasonable and similar attitudes towards propositions pertaining to evidence pool A. Then we must have both assessed evidence pool B and both arrived at reasonable and similar attitudes towards propositions pertaining to evidence pool B, and so on. I'll call the first reading the *independent track record* proposal and the second reading the *mutual track record* proposal.

The test for whether a definition of epistemic peerhood (even for Close Peers) is useful for the peer disagreement project is, once again, to assess whether the necessary requirements for peerhood given in that definition help us ferret out disagreement-relevant epistemic asymmetries. On the independent track record reading of Gelfert's (ii), there are more opportunities for unchecked disagreement-relevant epistemic asymmetries to arise than in the mutual track record proposal since disagreements, in the peer disagreement framework, are about specific propositions and one shared pool of evidence that bears on those respective propositions. That is, in the case of the independent track record proposal, you and I are only generally peers because we both have good track records of taking some evidence and arriving at reasonable conclusions (beliefs) based on that evidence. So, all we get with the independent track record proposal is something more akin to a general virtue condition being met; we are both reasonable people. When it comes to a disagreement about the truth value of a particular proposition,  $p$ , being generally reasonable may count as a relevant disagreement factor, but it also leaves too many openings for disagreement-relevant epistemic asymmetries to slip through regarding  $p$ . So, since the mutual track record proposal—where our track records are on

past encounters with each other *on the currently disagreed upon topic*—both aligns closer to Christensen’s definition of epistemic peers and helps us pick out areas where disagreement-relevant epistemic asymmetries may be present, we should modify (ii) to (ii\*) for Close Peers to read something like having “similarly good track records in forming true beliefs on the basis of evidence *on the topic that is being disagreed upon.*” With that change in mind, it will be helpful to have ways to define the quality and suitability of track records to fit with (ii\*) and, more loosely, with the kinds of track records Christensen had in mind in his original proposals for defining when two people are epistemic peers.

In his 2013 dissertation on peer disagreement, John Bundy discusses track records as a requirement for evaluating potential epistemic peers. He distinguishes several levels of track records: extraordinary, ordinary, less-than-ordinary, and (mere) general background information. An extraordinary track record is one in which the following conditions must be met: (1) the record is long (in terms of quantity of samples, I assume), (2) the track record tracks the history of true beliefs your potential peer has had about relevant propositions related to the key proposition that is under consideration in the current disagreement, (3) knowing when your interlocutor is joking, drunk, etc. compared to past experiences, (4) tallying up how many times you have been wrong compared to your potential peer in ways relevant to the proposition under consideration, and (5) exercising due diligence in fact-checking to be sure that your evaluation of your potential peer’s epistemic history have been verified (i.e., not just relying on your own judgments of when the peer was right or wrong, etc.) (Bundy, 2013, pp. 150-160). Bundy admits

that having an extraordinary track record with anyone is going to be rare (hence the term “extraordinary”) and that these five requirements may not even be exhaustive. The requirements for the remaining track record types (ordinary, less-than-ordinary, and general background) become less detailed and less demanding down the scale to virtually no track record at all, as we might have (or lack) when we encounter a complete stranger. The point here is not that we need to find a precise definition for different types of track records; the point here is that, following Christensen and Gelfert, track records are: (1) relevant for potential peer evaluations and (2) on a spectrum from highly detailed and multifaceted to sparse in terms of what is required to generate and maintain various levels of track records with our interlocutors.

Since my project is aimed at assessing whether the peer disagreement framework is amenable to real-world applications, it is worth pausing here to think about the kinds of relevant track records that we have with the people in our daily lives. As several authors mentioned in the criticisms of epistemic peerhood above, one of the deficiencies of the literature in general is that its authors are often too eager to gloss over the details of what it means for any two real-world people to believe that they are peers. So, here I aim to correct that oversight first by simply asking my reader: how many people do you know that you could count as having the sort of established track records proposed by Christensen and the modified (ii\*) from Gelfert? Pause to think it over.

Speaking for myself, I can think of only *two* people in my life who could meet the rather onerous track record condition as proposed in the literature. Furthermore, of those two people, the topics that we could actually claim to have peer-evaluation-worthy track

records on are exceptionally limited. The main problem here is not just that I haven't had sufficient time to discuss various topics at length with others, though. The problem is that nearly all the topics I discuss at length with these people do not have definite outcomes or objective answers. Kirk Loughheed (2020) makes a distinction between alethic expertise and non-alethic expertise, which relates to the truth-tracking methods we would use to evaluate the kinds of track records that Christensen, Gelfert, and Bundy propose for peer evaluations. Alethic expertise, put succinctly, is expertise in areas where one can verify if the expert's assertions or predictions are objectively correct. If a math teacher claims that  $1 + 1 = 2$ , we have a way to check that claim to determine its truth value. If a meteorologist claims that it will be sunny tomorrow, we have a way to check that claim tomorrow as well. However, non-alethic experts are those with expertise in subject areas that tend to lack objective truth-tracking methods (at least, within reason of human epistemic limitations). Loughheed's examples of non-alethic experts are those who are experts in subject areas such as religion, morality, and philosophy (Loughheed, 2020, pp. 54 - 63). Many of the questions studied in those fields do not have objectively verifiable answers (now or maybe ever). According to Bundy's proposed types of track records above, in order to have an extraordinary or even ordinary track record, the interlocutors must meet the following condition: (2) the track record tracks the history of true beliefs your potential peer has previously had about relevant propositions related to the key proposition that is under consideration in the current disagreement. That is, for any two people—experts or not—their track records can only be verified if their track record evaluation method tracks the history of true beliefs. Hence, even ordinary track records

are only applicable in situations where the interlocutors have a history of disagreeing about alethic subject matters rather than non-alethic matters. This limits the situations in which a peer evaluation can include a track record as a target of that evaluation since many real-world disagreements are on non-alethic matters. However, there are some matters of disagreement that are alethic or alethic-leaning that one could use to develop the relevant sort of track records under analysis here. I offer the following quaint but true anecdote to illustrate one instance of having a track record that would count for meeting that requirement for peerhood.

My partner likes to watch a TV show called *The Great British Bake Off* (GBBO), which is a baking competition where contestants bake specific foodstuffs that are judged by professional bakers. At the end of each episode, one of the contestant-bakers is disqualified (“sent home”) and another is named Star Baker (winner) for the week (until the next episode ends). I like to spend time with my partner, so I watch this show with her. Over the last decade, we’ve watched ten seasons of that show and each season is usually ten episodes. So, we’ve watched 100 episodes of the GBBO together. Somewhere around season 5, we started playing a game where after the judges tasted the contestant’s various baked goods but had not yet announced which contestant was disqualified and which contestant won the Star Baker award, we would each guess which contestant would be disqualified, and which one would win Star Baker. Over the past five years of playing this game with each other around 50 times, we’ve developed a track record where, in aggregate, both of us have been right with much better than chance odds (so, we’re good at this game) and we have both been right about the same number of times.

So, ostensibly, my partner and I meet the track record requirement for disagreements involving who will be disqualified and who will win the Star Baker award on the GBBO. This might sound quite pedestrian, but it is worth noting that this is one of the only few solid examples of a relevant epistemic peer track record that I could come up with when reflecting on my life. Here is the crucial fact: the set of propositions on which I could have a potential epistemic peer who meets the track record condition is infinitesimal compared to the set of propositions that I could find myself in a disagreement about. Furthermore, the number of people I've known and interacted with long enough to have the sort of relevant track record data required for peerhood is certainly less than five. My partner is the paradigmatic example of someone whom I could develop track records with, and yet I can only come up with a handful of examples where those track records are present. Perhaps my partner and I don't disagree about much, so there are limited opportunities to repeatedly test who is right and wrong about certain subjects. Perhaps when we do disagree, we don't keep a clear tally of who was right and who was wrong. It is entirely plausible that my lack of relevant alethic track records is an outlier situation, and maybe my reader has come up with many examples of track records that would fit the alethic model of truth tracking discussed above. However, just based on the sheer durational volume of time and meticulous record keeping that it would take to establish an ordinary or better track record with one person, it is reasonable to doubt that there are very many people who would fit the bill for the relevant sort of track records required here. Of all of the conditions required for epistemic peerhood, track records seem to be such exceptional byproducts of hyperidealizing cases that the track record requirement

should be ignored for any attempts to demonstrate how the literature's idealized cases could transfer to real-world epistemic peer evaluations. I won't argue this point further, but later I suggest reasons why even when we think we have good track records, we are probably mistaken.

Returning to Gelfert's proposal that peerhood may come in degrees, and with the modified versions of (i\*) and (ii\*) in place (while maintaining an incredulous stare toward the track record requirements of ii\*), Close Peers appear to be just regular epistemic peers. If this is the case, then the (modified) definition of Close Peers helps us pick out disagreement-relevant asymmetries as much as the standard or even the enhanced versions of the definitions of peers discussed above. Let's turn now to the proposal for degrees of peerhood as that Gelfert labeled Distant Peers and Remote Peers.

Distant Peers are those who satisfy *exclusively* either (i\*) or (ii\*). However, due to the burdens that ordinary or better track records impose on us, I'm setting (ii\*) aside. Now it is up to us to decide if (i\*) is a sufficient definition of peerhood that can help us pick out disagreement-relevant epistemic asymmetries that could explain why two people disagree. With the modified version, (i\*), we concluded that meeting (i\*) is just another way to say that two (or more) people meet the evidence condition and some major part of the virtue condition (i.e., evidence processing). So, if Distant Peers only meet (i\*), they are still epistemic peers in the relevant sense. However, if Distant Peers only meet (ii\*), they are not peers in the relevant sense because (ii\*) lacks both the evidence and virtue conditions. This is fine since satisfying (ii\*) alone will only apply to an exceptionally small number of real-world cases anyway. So, satisfying (i\*) is, perhaps roughly, just



satisfying Kelly's definition of peers. Satisfying (ii\*) alone is not sufficient as a way to pick out any peers since (ii\*) alone does not help us detect disagreement-relevant epistemic asymmetries *and* it will only apply to a severely limited number of cases.

Finally, Gelfert proposes that we might have Remote Peers who do not meet (i\*) or (ii\*) (or even i or ii). It is unclear how two people could fail to meet (i\*, or even i) and consider themselves peers. As I discussed above, if two people disagree and do not meet (i\*), the reason for their disagreement could easily be explained away by acknowledging that they just don't have the same evidence. With that explanation in hand, there is no need to turn to the peer disagreement literature for specific advice on how they ought to respond to the disagreement; the differential in their evidence suggests that the parties to the disagreement first need to get on the same page by examining each other's evidence. So, using the charitably modified definitions of (i\*) and (ii\*) above: Close Peers are just standard epistemic peers, Distant Peers are also just standard epistemic peers once we remove track records from the requirements, and Remote Peers cannot be epistemic peers in a way that can utilize the framework of the peer disagreement literature.

I've spent some time discussing Christensen's near-peer idea and Gelfert's proposal that peerhood can come in degrees to further demonstrate that reasonable attempts to loosely define peerhood simply do not fit in the framework of the peer disagreement literature. Any pertinent disagreement-relevant epistemic asymmetry can defuse the problem of peer disagreement before it begins. So, the current peer framework in the literature cannot handle loose definitions of peerhood or accommodate proposals

for degrees of peerhood—at least not without bending the rules that generate the puzzle in the first place.

Despite the dearth of analysis of epistemic peerhood in the literature, a few authors have tried to solve the loose vs strict problems with peerhood definitions. Realizing that overly strict definitions of peerhood rule out real-world application, several authors have merely hand waved at the problem by loosening the requirements for peerhood so that people may be “roughly” peers or “close enough” in terms of meeting the evidence and virtue conditions. However, while strict definitions of peerhood render the application of the peer disagreement literature beyond difficult to realize, loose definitions of peerhood include too many opportunities for disagreement-relevant epistemic asymmetries to go undetected. So, a usable definition of epistemic peers not only needs to be stable, precise, and detailed enough to capture disagreement-relevant epistemic asymmetries in real-world cases, the equality requirements proposed in that definition should also not be too strict or too loose. I discuss this further in Chapter 6.

### **3.4: Summary and Conclusion of Chapter 3**

In this chapter, I provided an analysis of epistemic peerhood as it is presented in the peer disagreement literature. Here are the take-home messages from this chapter:

- The conceptual framework and motivating puzzle of the peer disagreement literature depends on how we define epistemic inferiors, epistemic superiors, and—most importantly—epistemic peers.

- For the advice given in the peer disagreement literature to be applied to real-world cases of disagreement, we need to understand how to accurately and competently perform epistemic peer evaluations.
- Understanding how to accurately and competently perform epistemic peer evaluations requires using definitions of epistemic peers that captures disagreement-relevant epistemic asymmetries among potential peers.
- For a definition of epistemic peers to meet the requirement of capturing disagreement-relevant epistemic asymmetries, there are, at minimum, two general targets of evaluation that should be included: those that pertain to the evidence condition and those that pertain to the virtue condition.
- Some authors have proposed additional targets of evaluation such as background knowledge, circumstances of investigation, and the presence of accurate and lengthy track records.
- Requiring an accurate and lengthy track record that tracks the history of true beliefs formed by the parties to a disagreement severely limits who might count as a peer.
- The targets of evaluation in epistemic peer evaluations (evidence and virtues, at least) need to be equal among the interlocutors in such a way that they are not too strict to rule out any possibility of real-world people attaining peerhood but also not so loose that disagreement-relevant epistemic asymmetries might be present but remain undetected by the parties to the disagreement.

- Some authors have attempted to solve the strict / loose problem by suggesting that peerhood can come in degrees or in near / not-so-near categories. However, these attempts have been unsuccessful because they permit too many opportunities for disagreement-relevant epistemic asymmetries to remain undetected.

Allow me to condense these take-away points into a claim that will serve as a premise in my central argument presented later in Chapter 6:

**(P2) Accurately and competently performing real-world epistemic peer evaluations requires the following: (1) using a definition of epistemic peerhood that appropriately captures disagreement-relevant epistemic asymmetries among potential peers, and (2) using a definition of epistemic peerhood that is not overly strict or excessively liberal in terms of equality or parity of the targets of evaluation given in that definition.**

Keeping this premise in mind, in the next chapter I discuss Nathan King’s 2012 article that brings the concerns about equality or parity of the evidence and virtue conditions in applied settings to the forefront. To date, King’s criticism stands out as one of the clearest and most substantial criticisms of the problems with attempting to use the standard evidence / virtue definition of epistemic peerhood in real-world disagreement situations—so much so that King’s article deserves an entire chapter in my project.

#### **CHAPTER 4: KING’S CRITICISM OF EPISTEMIC PEERHOOD**

Nathan King’s 2012 article “Disagreement: What’s the Problem? *or* A Good Peer is Hard to Find” is one of the few serious critical examinations of peerhood in the literature (Kenyon 2020 and Lougheed 2020 also stand out here). As its title suggests,

King's article focuses on how the problem of peer disagreement is contingent on there being epistemic peers at all. Recall that the puzzle of peer disagreement challenges us to think about the epistemic significance of disagreement. Feldman (2006) argued that peer disagreement is epistemically significant because these disagreements should induce skepticism and lead to suspension of judgment (see Chapter 2). King argues that if we don't often find ourselves engaged in peer disagreements, the skeptical pressure—and therefore epistemic significance—of *peer* disagreement is markedly diminished. For King, if the advice from the literature is going to be useful in real-world disagreements, then we should discuss whether or not epistemic peerhood is, or at least is thought to be, a common occurrence.

To set up his argument, King explains his motivation for investigating peerhood:  
[Peer] disagreements give rise to several interesting philosophical questions.

Among them:

- Can it be rational to retain our beliefs in the face of disagreement with an equally intelligent, equally well-informed subject (a so-called *epistemic peer*)? If so, *how*?
- How often are our dissenters our epistemic peers? And how often do we have reason to think that our dissenters are peers?

Much recent work in epistemology has addressed the first pair of questions. The second pair has gone relatively under-explored. This is a curious situation. For a prominent theme in the literature is that widespread peer disagreement mandates

widespread doxastic attitude revision. This theme is perhaps most vividly expressed as an argument for a certain kind of skepticism. The argument requires answers to *both* pairs of questions above. It goes roughly like this: (i) If one finds oneself party to a disagreement with an acknowledged epistemic peer, it is irrational to retain one's belief—one is rationally required to suspend judgment under such conditions. (ii) We often find ourselves party to disagreements with acknowledged peers (indeed, such disagreements often concern our most cherished beliefs). Thus, (iii) it is quite often irrational for us to retain our cherished, controversial beliefs; we are often rationally required to suspend judgment about such matters—at least until further evidence comes in... In recent discussion of the skeptical argument from peer disagreement, premise (ii) is often left untouched. In effect, the discussion has assumed that we are often party to acknowledged peer disagreement.

But *are* we? ... I argue that we are not. I show that peer disagreement is rare, and that we rarely have reason to think it obtains in a given case. If this is right, the skeptical argument from peer disagreement is a failure, *irrespective* of how the discussion of claim (i) turns out. Moreover, if I am correct, participants in the current discussion about disagreement have been laboring under a false assumption. Real-world disagreements concerning issues we care about are *not* peer disagreements. Or at any rate, they seldom are. As a result, it is not clear to what extent the contemporary discussion of peer disagreement is relevant to the

rational status of our most cherished beliefs. (King, 2012, pp. 249-250, emphasis in original).

This passage from King embodies the spirit of my entire project. However, before discussing King’s argument further, I offer a point of clarification. It is important to recognize the distinction between the *ontological question* of epistemic peers and the *epistemological question* of epistemic peers.<sup>22</sup>

The ontological question, “Are there epistemic peers?”, is perhaps interesting in its own right, but not the central question that King or I desire to answer. Recall that in Chapter 3, I explained that the cases of peerhood in the literature are primarily idealized. If we were only concerned with examining ideal cases of peer disagreement, the ontological question would perhaps be a suitable way to investigate the phenomenon. We would (merely) need to believe that epistemic peers are metaphysically possible—which they are—and then the investigation could turn to how idealized interlocutors should rationally respond to ideal cases of disagreement.<sup>23</sup> However, along with King, I am

---

<sup>22</sup> Elga 2007 makes a similar distinction in a footnote: “Note that in setting up the problem [of peer disagreement], the initial assumption is that you *count* your friend as your epistemic peer. That contrasts with some presentations, in which the initial assumption is that your friend *is* your epistemic peer. The former assumption is appropriate, however. For example, one sometimes is reasonable in thinking wise advisors to be foolish. Evidence, after all, can be misleading. In such cases, one is reasonable in being guided by one’s *assessments* of the advisor’s ability, even if those assessments are in fact incorrect.” (2007, fn14, p. 499). Furthermore, Comesaña 2012 elaborates on these distinctions even further by splitting the question about peer assessments into metaphysical (we have epistemic peers), psychological (we consider others epistemic peers), and epistemic (we justifiably believe others may be peers) questions. I do not need to be this fine-grained here; the distinction between ontological and epistemological conceptions of peers will be sufficient for my project.

<sup>23</sup> Robert Aumann’s 1976 “Agreeing to Disagree” is a good example of a formal model of purely idealized rational agents with epistemic symmetry that disagree, which Aumann argues is untenable. Most proposals in the peer disagreement literature do not idealize peers to this extent, but the literature does employ a good deal of idealization at any rate. Kelly 2005 discusses Aumann’s work in relation to peer disagreement and concludes that the ‘no agreeing to disagree’ position offered by Aumann is only tangentially related to the puzzle of peer disagreement. (p. 11-13) Finally, Matheson 2015 and Loughheed 2020 discuss some relevant issues with relying on idealized cases to defend the various positions found in the literature.

concerned with how the literature can play a role in helping us respond to actual disagreements. So, for actual cases, the ontological question is only of value if it turned out that epistemic peers were *not* metaphysically possible, which is not the case. The epistemological (doxastic) question, “Are we justified to believe that we have any epistemic peers?”, is much more important to answer. Furthermore, for my project, I am interested in what I will call the *applied social epistemological question of peerhood*: is it likely that we can accurately identify others as epistemic peers in real-world cases? King does not believe so, and, as I argue in Chapter 6, neither do I, but for different reasons. To probe the question further, I present King’s case here.

King summarizes the requirements for a peer disagreement to occur as a conjunction of four necessary conditions:

- (a) *The disagreement condition*: S believes P, while T believes that  $\sim$ P.
- (b) *The same evidence condition*: S and T have the same P-relevant evidence, E.
- (c) *The dispositional condition*: S and T are equally disposed to respond to E in an epistemically appropriate way.
- (d) *The acknowledgement condition*: S and T have good reason to think conditions (a) – (c) are satisfied. (King, 2012, p. 252-253)

Let’s pause to consider whether King’s necessary conditions for peer disagreement and epistemic peerhood align with the conditions defended broadly in the literature (see Chapter 3). For the sake of precision, condition (a) needs a bit of unpacking. King adds the caveat that condition (a) must be a *genuine* disagreement since it is possible for S to believe that *p* while T believes that *not-p* yet S and T are not genuinely disagreeing. For



example, if I believe that my computer is fast and you believe that it is not the case that my computer is fast, we might not actually be genuinely disagreeing if we are using “fast” in two different senses. Indeed, issues of vagueness, ambiguity, context sensitivity, and equivocation arise often in casual, real-world disagreements. Furthermore, it is sometimes the case that two people disagree indirectly rather than asserting direct contradiction. For example, say you and I are discussing human free will and you discover that I am a libertarian and I discover that you are a determinist. I claim  $p$ : “Humans have free will.” You claim  $q$ : “The universe is fundamentally deterministic.” Are we disagreeing? The answer to that question will depend on whether your version of determinism (and perhaps your version of compatibilism) entails that humans don’t have free will. That is, if  $q$  entails that not- $p$ , but not- $p$  is not clearly stated during the discussion, it could still be the case that you and I are genuinely disagreeing here. With this caveat in hand, King’s condition (a) is sufficient to generate the genuine disagreement horn of the trilemma I presented in Chapter 2.

King’s same evidence condition, (b), is roughly a reiteration of Kelly’s evidence condition (see Chapters 2 and 3), so (b) seems to align with what I have been calling the standard definition of epistemic peerhood. However, King’s test cases employ what he calls dialectical evidence, which is just the arguments and their constituent parts that bear on whatever is being disagreed upon. I explain why King gravitates towards this category of evidence below.

The dispositional condition, (c), is similar to the virtue condition from Kelly’s standard definition of peerhood as well. However, King focuses on how reliable each

interlocutor is in terms of how they respond to the evidence that bears on the disputed proposition (this is similar to Gelfert's (i) from Chapter 3). If you are more reliable in responding to evidence than I am, then, due to that virtue-based disagreement-relevant epistemic asymmetry, we do not meet (c). If we disagree but do not satisfy (c), we're not having a peer disagreement.

Interestingly, King's condition (d), the acknowledgement condition, is not often discussed in the literature. What does it mean to acknowledge that one is an epistemic peer? According to King, the acknowledgement condition is a "[H]igher-level claim about subjects' reasons for thinking that they are engaged in a genuine disagreement with an epistemic peer" (King, 2012, p. 261). In other words, for a peer disagreement to occur, the disputants must believe that (a) – (c) obtain simultaneously to satisfy condition (d). Without this higher-level belief—one that acknowledges that (a) - (c) obtain—we cannot have a peer disagreement.

Recall that, according to the literature, a peer disagreement can be generated when King's (a) – (c) alone are met. This means that the genuine disagreement condition, the evidence condition, and the virtue condition are all met. So, why does King think we need a fourth condition for a real-world peer disagreement to obtain? Often in the idealized cases from the literature, the imagined interlocutors are merely aware of (a) – (c) and, presto, a peer disagreement is instantiated. So, it is reasonable to wonder if King's (d), the acknowledgement condition, is redundant here. It may be. However, it is also a crucial condition for real-world interlocutors to affirm (or deny) (a) – (c) before moving to their preferred prescription from the literature (conciliate, remain steadfast,

etc.). If I want to use the advice given in the literature, it is perhaps not enough for (a) - (c) to be met; I need to also be aware (i.e., acknowledge) that those conditions have been met. Being able to believe with justification that one's interlocutor is an epistemic peer on whatever matter is being disagreed upon is key to *applying* the normative advice offered in the literature.

So, we've affirmed that King's conditions (a) – (c) (with caveats) align well with the requirements from the literature for peer disagreements. Furthermore, King's additional condition (d) may be redundant in some respects, but it is a relevant condition when considering the usefulness and applicability of the prescriptions offered in the literature. In the next section I discuss King's argument against the likelihood of condition (d) being met.

#### **4.1: King's Argument**

King argues that neither the same evidence condition (b) nor the dispositional condition (c) are likely to be met in real world cases of disagreement, which leads to doubts that condition (d) can be *justifiably* met. Concerning the failure of (b) in epistemic peer evaluations, King offers the following example: Philosophers Mike and Keith have a disagreement about metaphysics—specifically about the real-ness of universals (King, 2012, p. 254-255). In King's example, Mike and Keith have been exposed to the same *dialectical* evidence. According to King, for interlocutors to satisfy the dialectical evidence condition, they only need to be familiar with the same arguments that bear on *p* rather than being mutually familiar with *all* of the evidence that could bear on *p*.

I'll paraphrase King's case of Mike and Keith to explain further: Let's say that Mike has studied arguments A, B, and C regarding the topic of universals. Let's further say that Keith has studied arguments B, C, and D regarding the same topic. The question here is simple: given that Mike and Keith both studied arguments B and C, do they meet condition (b), that is, the same evidence condition? King thinks not since Mike also has dialectical evidence from argument A that Keith does not, and Keith has dialectical evidence from argument D that Mike does not. So, both Mike and Keith have carefully studied arguments B and C and have that in common, but each one also has studied additional arguments that bear on the question under consideration. However, according to King, in this case Mike and Keith—even on this relatively permissive dialectical version of the “same evidence” condition—do *not* satisfy condition (b) for epistemic peerhood. On King's view of dialectical evidence, the requirements to meet condition (b) require that the interlocutors must have the *same and only the same* evidence, which aligns with a fairly strict yet popular version of the evidence condition from the main literature.

According to King's dialectical view of evidence, the evidential equality demanded by (b) is rarely met in real world cases—even among philosophers who might otherwise generally consider each other to be epistemic peers—because it is rarely the case that one interlocutor has been exposed to *exactly* the same arguments as their disagreeing interlocutor and no others. In the parlance I used in Chapter 3, one party exclusively possessing argument A (or D) in this case would count as a disagreement-relevant epistemic asymmetry. Therefore, on King's view, Mike and Keith cannot be

epistemic peers even on a permissive account of evidential equality. Furthermore, hopes for evidential equality actually diminish after we consider that Mike and Keith might have coextensive dialectical evidence regarding  $p$  (so, they have been exposed to the same arguments regarding the veracity of  $p$  and no others) but could still fail to meet a more general version of evidence condition (b). How so? King has us suppose that Mike and Keith have been exposed to the same arguments regarding  $p$  and no others, yet Mike still holds that  $p$  and Keith still holds that  $\text{not-}p$ . According to King, even when Mike and Keith satisfy the dialectical evidence equality condition in this example, it could be the case that Mike or Keith are affected by environmental conditions, intuitions or philosophical “seemings”, or background factors (recall Frances’s list of disagreement factors in Chapter 3) that somehow sway their respective beliefs about  $p$ . If this is the case, the situation is now quite messy; Mike and Keith might be very familiar with exactly the same arguments regarding  $p$ —or, put differently, might meet King’s permissive dialectical version of evidential equality—but they could still have very different evidential positions given their respective background factors, intuitions, sensory inputs, etc. If this is the case—as it usually is in real-world disagreements—it means that Mike and Keith *do not share* the same  $p$ -relevant evidence in such a way that the parity or symmetry required by condition (b) can be realized. When we take a sober look at the evidential positions of our real-world interlocutors, it is apparent that they do not share even the same dialectical evidence with us, let alone *all* of the evidence that bears on the disagreement. Therefore, King believes that condition (b) will very rarely if ever obtain in real-world scenarios among disagreeing interlocutors.

King also addresses the problems that might arise with the dispositional condition (c), and his argument against interlocutors satisfying condition (c) is worth exploring further here. Let's stipulate that two interlocutors have exact equality of evidence such that they meet condition (b). Despite how unlikely the perfect fulfillment of condition (b) may be, does this stipulation help us confidently determine who our epistemic peers might be? It turns out, according to King, that we are no better off in this situation. Recall that condition (c) requires that "S and T are equally disposed to respond to E in an epistemically appropriate way." (King, 2012, p. 252-253). King proposes that *justifiably believing that his dispositional condition (c) has been met* will be at least as difficult as *justifiably believing that condition (b) has been met in real world cases of epistemic peer evaluations*. Let's see why.

Recall that King uses the epistemic virtue condition from the standard definition of peers (Chapters 2 and 3) to inform his requirements for meeting the dispositional condition (c). In his words:

This condition concerns the relative merits of the dissenting subjects as evidence assessors. The most important feature of such equality, I suggest, is equal reliability with respect to the relevant field of inquiry. For plausibly, the reason such items as intelligence and intellectual virtue are often included in definitions of "epistemic peers" is that these characteristics typically render subjects reliable as evidence assessors. (King 2012, p. 258)

Here, epistemic reliability is the crucial property under examination when evaluating ourselves and others for the dispositional condition. King provides an example of

interlocutors who do not meet the dispositional condition, which, again, I'll paraphrase here. Let's say that S and T both study two news articles. One is a sensational tabloid and the other is a "good" news source. Now let's say that S gives more credit to the sensational tabloid for evidence regarding  $p$  while T gives more credit to the good news source for evidence regarding  $p$ . The question of reliability here is interesting because S and T could both be equally reliable in the relevant sense *if* they shared the appropriate background beliefs about the credibility of their respective news sources. However, because S gives more credit to the less credible source than T does, there is something less reliable about S's assessment of the  $p$ -relevant evidence compared to T's assessment of  $p$ -relevant evidence. Despite S and T having the same evidence (the two articles), because of S's background belief that the tabloid is more credible, S has failed to properly assess the credibility of all the  $p$ -relevant evidence in this case. So, S is less reliable than T in assessing the credibility of the evidence that bears on  $p$ . According to King, this disparity in reliability shatters the epistemic symmetry required by condition (c). We do not need to look far to find people disagreeing about which news sources are more credible, so, anecdotally, it seems like King's example generalizes to real-world interlocutors well.

The reliability question (so, the dispositional question) also bears on academic debates in a general sense:

In many cases of disagreement in academic fields, equal reliability may require equality with respect to a range of epistemic virtues such as honesty, carefulness, and freedom from bias. Suppose that one of two dissenting subjects is more

careful than the other, or less biased, or more honest. All else being equal, we would expect the more careful, less biased, more honest person to be more reliable than her dissenter. (King, 2012, p. 259)

Finally, King leaves some room to doubt that condition (c) can *never* be satisfied among parties to a disagreement. Regarding his claims about the likelihood of condition (c) being met in real-world situations:

None of the above should be taken to imply that it is *impossible* for subjects to satisfy the dispositional condition on peerhood. However, it should be clear that satisfaction of the condition is in many cases a very complicated matter. It is plausible to think that its satisfaction is fairly rare. If this is right, then we should not accept uncritically the claim that two subjects satisfy it in a given case. (King, 2012, p. 261, emphasis in original)

Finally, according to King, failing to satisfy the evidence condition, (b), or the dispositional condition, (c), entails failing to satisfy the acknowledgement condition, (d). Finding ourselves in an acknowledged peer disagreement is not only unlikely because we rarely find ourselves disagreeing with people who have the “same” evidence or the “same” dispositions, but also because we will not be able to identify our epistemic peers due to the uncertainty of our higher-level evidence as it bears on the requirements for (b) and (c). If you cannot be reasonably sure that your interlocutor has the same *p*-relevant evidence (evidence condition) or the same disposition to be as reliable as you are in evaluating *p*-relevant evidence (roughly, the virtue condition), you cannot reasonably believe that your interlocutor is a peer. So, if we cannot justifiably believe that (b) and (c)



have been satisfied, we also cannot believe that (d) has been satisfied (i.e., it would be unreasonable to believe that your interlocutor is your peer). Without (d), according to King, we cannot have a peer disagreement.

King anticipates a critical objection to his view, namely, that his definition of epistemic peerhood is too strict. He counters this objection by pointing out that his evidence and disposition conditions are charitable in that they do not require perfect epistemic symmetry. It is easier to meet the mere dialectical version of evidential parity than to meet evidential parity for *all* of the disagreement-relevant evidence, including difficult-to-share evidence such as intuitions, etc. Additionally, satisfying the dispositional condition is less demanding than satisfying a comparison of a long list of epistemic virtues. From one of King's examples, a golf caddy can give expert advice on golf without having the same exact IQ as the golfer he is assisting. So, due to King's liberal versions of evidential and virtue parity, it is not the case that his definitions of peerhood are too strict. King also notes that when equality requirements are too lenient, it is unclear how much significance disagreement has. King writes:

If, for instance we loosen the requirements so that peerhood is consistent with small differences in evidence, there may be cases in which a small evidential difference between subjects makes a *large* difference in what it is rational for the subjects to believe. A single piece of evidence may in some cases be the *key* piece. (2012, p. 266)

Recall that I began a discussion about loose vs. strict conceptions of peerhood in Chapter 3, and, following King's lead, I discuss that problem further in Chapter 6. It is worth

noting here, however, that King realizes that lenient versions of peerhood permit too many opportunities for disagreement-relevant epistemic asymmetries to go unnoticed. Again, a definition of peerhood that does not capture most of the disagreement-relevant asymmetries is not a good definition of peerhood.

King further argues that even if one of the conditions, (b), (c), or (d) can be met, satisfying all three will be difficult:

Taken jointly, the conditions are quite difficult to satisfy—so difficult that it is probably rare for two subjects to satisfy them in cases of at least moderate complexity. Perhaps there are often acknowledged peer disagreements concerning simple subjects like the rules of tic-tac-toe. And maybe acknowledged peer disagreement is common among novices whose evidence is meager and whose track records are short. But these aren't cases of the sort that drive the current interest in peer disagreement. When it comes to issues we tend to care about, it is rare for subjects to find themselves involved in a genuine disagreement with someone who is, and who they have good reason to believe is, their epistemic peer. (King, 2012, p. 263)

Later in King's article, he argues that while the epistemic significance of disagreement may be salient and forceful, this significance is not due to the peerhood aspect of *peer* disagreement. That is, after casting serious doubt on the idea that the conditions for peerhood are likely to obtain for all but the most trivial disagreements in real-world cases, King argues that *disagreement itself* (regardless of the relative epistemic standings among the interlocutors) is epistemically significant. On his view, dissenting opinions

from others should give us pause in many cases even if we are not certain about whether those opinions are held by epistemic peers. I agree with King's point in general, but I also think there is more to be said about the reasons why we should not rely solely on epistemic peerhood as the criterion for deciding how we ought to rationally respond to disagreements. When our disagreeing interlocutor is obviously our epistemic inferior or superior, it is at least *prima facie* reasonable to remain steadfast or to defer, respectively. However, it is less clear how we ought to respond to disagreements when we are unsure about our epistemic standing relative to others. I discuss this issue further in Chapter 6.

#### **4.2: Summary and Conclusion of Chapter 4**

Once we combine what we learned in Chapter 3 with our discussion of King here in Chapter 4, we're in a position to gain new insight into how and whether we can identify and justifiably believe that others are our epistemic peers. Ultimately, there are reasons to doubt that either the evidence or virtue conditions can ever be met in real cases. These specific reasons to doubt that peerhood can be recognized support the following premise:

**(P3) Criticisms of epistemic peerhood generally rely on doubts that the required evidential parity or epistemic virtue parity conditions are likely to be met in real-world situations.**

In the following chapters, I present evidence that supports King's claim that it is unlikely for any two interlocutors to justifiably acknowledge that they are epistemic peers. In the next chapter, I discuss some of the empirical findings that suggest that we are not cognitively well-suited to accurately perform relative epistemic peer evaluations

on ourselves or our disagreeing interlocutors. If this is the case, even if it is likely that we commonly satisfy the evidence, virtue, and perhaps acknowledgement conditions, we should still not believe that we are competent to judge when those conditions have actually been met.

## **CHAPTER 5: BOUNDED RATIONALITY AND THE HEURISTICS AND BIASES PROGRAM**

In earlier chapters, I reviewed the peer disagreement literature, analyzed various definitions of epistemic peerhood, and discussed the extant criticisms of those definitions. In the previous chapter, I discussed Nathan King's incredulity regarding the likelihood of acknowledged peerhood obtaining in real-world situations. In this chapter, I shift the focus of my project from the disagreement literature to psychological research on cognitive biases. This research illuminates the descriptive, real-world problems we might face when performing epistemic peer evaluations. In the next chapter, I argue that, due to what we have learned from the research I present here, there are good reasons to believe that it is unlikely that we will ever find ourselves in what King called acknowledged peer disagreements.

In this chapter, I begin with an overview of Herbert Simon's theory of bounded rationality. Next, I discuss how that theory developed into the Heuristics and Biases Program through the influential work of Daniel Kahneman and Amos Tversky, among many others. Then, I present several examples of robust cognitive biases that unconsciously affect how we evaluate the epistemic competencies of ourselves and other people. The purpose of this descriptive account of how human minds typically function is

to demonstrate that evaluating the epistemic standing of others relative to ourselves is a process that is fraught with potential errors. Importantly, as I argue in Chapter 6, these errors should not be ignored when attempting to evaluate ourselves and others for epistemic peerhood. I conclude this chapter with a definition of what I call “The Cognitive Bias Problem”, which I deploy in the next chapter to support my thesis.

### **5.1: Overview of Simon’s Bounded Rationality**

In the mid-20<sup>th</sup> century, psychologist Herbert Simon called the prevailing conception of *Homo economicus* into question.<sup>24</sup> *Homo economicus*, as a general model of human behavior, assumed that agents are perfectly rational in market transactions. In opposition to those assumptions, Simon argued that we should change how we model rational choice optimization and decision-making:

Broadly stated, the task is to replace the global rationality of economic man with the kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist. (Simon, 1955, p. 99, quoted from Wheeler, 2018)

Simon’s goal was to constrain models of rationality so that modeled agents could in fact (descriptively defined) behave in the ways that the model represented them. The descriptive constraints that Simon proposed aimed to roughly define the *boundaries* of human capacities and capabilities to act as rational agents in economic markets. Hence,

---

<sup>24</sup> The concept of *Homo economicus* is often credited to J.S. Mill, though Mill never used that exact phrase. According to Persky (1995), the term *Homo economicus* became popular as a way for Mill’s commentators and critics to summarize his theories. For a further description of the history of that label, see Persky 1995.

Simon's approach is called **bounded rationality**: humans can be rational, but not perfectly rational in actual world scenarios because their abilities are bounded by practical cognitive, behavioral, and environmental limitations (Gilovich and Griffin, 2002, pp. 1-2).

One shorthand way we can describe the foundation of Simon's work (especially for philosophers) is to say that the theory of bounded rationality strongly suggests that "**ought implies can**". What I mean by this is that the power of a model to predict states of affairs is improved by limiting the possible elements of the model to reasonable representations of actual states of affairs. For example, imagine a subject faced with the choosing to complete one and only one of the following tasks, A or B:

**Memory Experiment 1 (Actual World):**

- A) Receive \$5 for memorizing 10 words in one minute and perfectly recalling all 10 of those words at a later time
- B) Receive \$1000 for memorizing 1000 words in one minute and perfectly recalling all 1000 of those words at a later time

All but perhaps those very few people with exceptional, savant-like abilities would choose task A simply because task B is too daunting if not nigh impossible. However, consider:

**Memory Experiment 2 (Counterfactual World):**

The same tasks are presented in the same manner as Memory Experiment 1, but our working memory and long-term storage and recall abilities can typically easily handle the requirements of task B.

In Memory Experiment 2, it would be rational for the (counterfactual) person to select task B over task A since the payout is double for B and only requires slightly more effort (in the counterfactual world) than task A. The point of comparing these two experiments is not to (trivially) demonstrate that our memories are functionally limited. We have limits to our cognitive capabilities—this should be obvious. The point here is that if we wanted to design a model that would predict which task most subjects would (and should rationally) choose, we will get different predictions by including the different descriptive states of affairs between Memory Experiment 1 and Memory Experiment 2. A model that does not take the cognitive limitations of typical subjects into account will favor task B because task B maximizes the expected utility / payout when compared to task A. However, if our model insisted that *actual world* subjects ought to favor task B, that model would be useless for offering guidance to actual people with limited memory capacities. While this example is (deliberately) an oversimplification of how we build reasonable normative models of rationality, it illustrates why we should include descriptive limitations (i.e., boundaries) in those models. Simon explains this point further:

[S]ome of the constraints that must be taken as givens in an optimization problem may be physiological and psychological limitations of the organism (biologically defined) itself. For example, the maximum speed at which an organism can move establishes a boundary on the set of its available behavior alternatives. Similarly, limits on computational capacity may be important constraints entering into the

definition of rational choice under particular circumstances. (Simon, 1955, p. 101, quoted from Wheeler, 2018)

Simon's proposal for bounded rationality implores us to consider the descriptive limitations of our cognitive capacities in normative models of our rationality. Of course, bounded rationality is much more complex than I have presented it here (as are the alternative theories of economic modeling methodology) and I do not intend on arguing for bounded rationality over its competing theories. I included this brief overview of bounded rationality for two reasons. (1) The motivation for bounded rationality suggests a general principle: we ought to consider a balance of descriptive and normative factors when offering prescriptions for rational behavior (again, ought implies can). (2) researchers investigating the bounds of bounded rationality laid the foundation for the Heuristics and Biases Program, which, as I demonstrate later, is relevant to the problem of peer disagreement. More specifically, the Heuristics and Biases Program is relevant to the problems inherent in performing relative epistemic status evaluations among interlocutors. So, the relevance of reason (1), if Simon was right, should be apparent; prescriptions aimed at real-world application and compliance should be attenuated by what can actually be the case. The relevance of reason (2) is explained below in my discussion of the development of the Heuristics and Biases Program.

## **5.2: An Overview of the Heuristics and Biases Program**

Simon's proposal for bounded rationality was *a priori* plausible in view of one important fact: it is not contentious to believe that there are significant differences between idealistic and realistic conceptions of human behavior, generally speaking.



However, it is one thing to observe that humans have cognitive limitations, and it is another thing to demonstrate what those limitations are. That is, if *bounded* rationality is going to inform us about ways to build better applied and predictive models of human behavior (in economic market transactions or elsewhere), then determining where to draw the boundaries of our cognitive capabilities is crucial for that project.

Paul Meehl, a contemporary of Simon, conducted research that investigated the cognitive limitations of human judgment (overviews in: Gilovich and Griffin, 2002; Wheeler, 2018). For example, in a landmark study, Meehl (1954) compared the clinical outcomes of patient care with actuarial data on the outcomes of patient care. Meehl's study found that actuarial predictions of patient outcomes were more accurate than the predictions of the clinicians who were treating those patients. That is, in some cases, a statistical analysis of a patient's situation was a better guide for prognosis and treatment than the attending clinician's judgment. Furthermore, when clinicians were asked to rate how confident they were that their prognoses and treatments were accurate and effective, the clinicians reported higher (often much higher) confidence in the accuracy of their judgments than was warranted by actual patient outcomes.

With the search for cognitive limitations in mind, Meehl's seminal work is notable for several reasons. First, it strongly suggested that clinicians could give patients better care by using statistical algorithms instead of, or at least alongside, standard human evaluations. If this is the case, then one probable cognitive limitation worth noting is that our cognitive evidence processing faculties do not use (good) statistical algorithms to make decisions. If we did, there wouldn't be much difference between the outcomes

predicted by the actuarial analyses and the clinicians' evaluations (assuming that clinicians aimed for beneficence in those cases). Second, Meehl's research motivated later studies on overconfidence. As I discuss below, studies on overconfidence present us with another probable limitation: we aren't good judges of our own judgment. Finally, Meehl's work offered a powerful impetus to continue research into defining the boundaries of rationality.

In later studies, Simon, Meehl, and others began addressing the boundary-defining project for bounded rationality. That work eventually evolved into the Heuristics and Biases Program (HBP henceforth). Early pioneers of the HBP—now famous in academic psychology—were Daniel Kahneman and Amos Tversky. From the introduction of the 2002 compendium *Heuristics and Biases: The Psychology of Intuitive Judgment*:

In the late 1960's and early 1970's, a series of papers by Amos Tversky and Daniel Kahneman revolutionized academic research on human judgment. The central idea of the "heuristics and biases" program—that judgment under uncertainty often rests on a limited number of simplifying heuristics rather than extensive [cognitive] algorithmic processing—soon spread beyond academic psychology, affecting theory and research across a range of disciplines including economics, law, medicine, and political science. The message was revolutionary in that it simultaneously questioned the descriptive adequacy of ideal models of judgment and offered a cognitive alternative that explained human error without invoking motivated irrationality. (Gilovich and Griffin, 2002, p. 1)

Kahneman and Tversky's early research focused on how humans assess risk when they are faced with incomplete information or small sample sizes—or, in short, when faced with judgment under uncertainty (Kahneman, Slovic, and Tversky, 1982; Gilovich, Griffin, and Kahneman, 2002). As the HBP developed, it became apparent that humans are often bad at probabilistic reasoning. While Kahneman and Tversky's approach and conclusions has had critics, the surprising experimental results from early research in the HBP at least merited further investigation.<sup>25</sup> This further investigation led to Kahneman and Tversky's **Prospect Theory**.

Prospect Theory proposed that people are generally risk-averse to potential losses, but risk-tolerant of potential gains, all else being equal. Kahneman and Tversky (1979) found that people tend to evaluate financial risk differently depending on whether the risk is posed as a gain or a loss (i.e, whether their “prospects” are favorable or unfavorable, even when the expected utility of the risk is the same).<sup>26, 27</sup> Prospect theory criticizes the

---

<sup>25</sup> One outspoken critic of the “pessimistic” side of the HBP is Gerd Gigerenzer. His work (included in Gilovich et al. 2002, Ch. 31) promotes the idea that our heuristics are largely beneficial and the few errors they might cause do not indicate widespread irrationality. Gigerenzer's positive project argues for a “fast and frugal” model of judgment that shows how our heuristics help us more than they hurt us (see Minimalist and Take the Best approaches in Gilovich et al. 2002, pp. 562-564). While I see no problems with looking for the positive aspects of the HBP, much of Gigerenzer's work is based on theoretical models rather than experimental data (so, we're already comparing apples to oranges here). It is also worth noting that Gigerenzer's positive project questions the applicability of study designs that involve testing for biases in statistical reasoning (e.g., base rate errors, law of “small” numbers (discussed below)). However, his criticisms of those particular elements of the HBP do not always carry over to the HBP's extensive findings about non-statistical reasoning errors found in confidence biases, halo effects, or confirmation bias, all of which I discuss below.

<sup>26</sup> The gains and losses in the initial prospect theory experiments were posed as *gambling* gains and losses. Kusev et al. 2009 argue via their experimental results that when losses and gains are framed in different contexts (e.g., as insurance options), the propensity of risk aversion and risk-seeking behaviors fall on a spectrum depending on the context of gains and losses.

<sup>27</sup> It is important to note here that Prospect Theory has been used in both behavioral economics and social psychology experiments to support the idea that there is an Endowment Effect (for example, see Kahneman, Knetsch, and Thaler, 1991) where subjects desire to receive more money for an item that they

idea that agents interacting in market transactions are ideally rational, as neoclassical theories of economics assumed in the mid-20<sup>th</sup> century (often broadly captured by the label “Chicago school of economics”).

Kahneman and Tversky’s early research not only offered a window on how typical people might irrationally respond to perceived gains and losses, but it also exposed how certain heuristics can lead us astray. Hence, the research by Kahneman and Tversky on Prospect Theory led us to the evidence that Simon needed to establish an empirically grounded practical theory of bounded rationality for agents operating in economic markets. Recall that one of the problems of defining bounded rationality pivoted on this question: How (or “where”) is our rationality bound? Prospect Theory pointed us towards the HBP’s answer to Simon’s original question.

### **5.3: Anchoring, Availability, and Representativeness Heuristics in Early HBP Studies**

While Prospect Theory demonstrated that people are often inappropriately risk-sensitive in some situations (whether risk-averse or risk-tolerant), Kahneman and Tversky were also interested in further studying how people responded to questions about the likelihoods of certain events. Prior to the development of Prospect Theory, Kahneman and Tversky (1973) had already proposed that we use cognitive heuristics (henceforth “heuristics”) to make judgments under uncertainty. Along with many other psychologists,

---

have been gifted than a buyer would offer for the item, even when the roles of receiver / buyer are reversed for in-between subject experiments. However, it is worth noting that Prospect Theory is not merely synonymous with the Endowment Effect since the Endowment Effect has faced some difficulties in replication (see Shogren et al. 1994 for one early criticism).

Kahneman and Tversky continued investigating our use of those heuristics in a series of groundbreaking studies. Below, I briefly present findings from those early studies.

Kahneman and Tversky proposed three heuristics: anchoring, availability, and representativeness. I'll address the research on each briefly below, but first it is important to generally define what a heuristic is. For the purposes of this project, a heuristic can be thought of as an input–output algorithm; a neurological mechanism by which presented or stored information is unconsciously evaluated as evidence for making judgments. That is, heuristics are, in this sense, fixed, evolutionarily mediated information processing channels that “answer” questions (especially under uncertainty) without the agent’s conscious awareness of the processing involved in reaching that “answer”. One might wonder, then, how can we be sure that we have or use heuristics in the first place? The short answer is that we can model our heuristics based on the behaviors of subjects who make judgments with incomplete information, as we’ll see in the descriptions of various heuristics below.

### **5.3.1: Representativeness Heuristic**

Kahneman and Tversky describe the representativeness heuristic succinctly in their introduction to *Judgement Under Uncertainty* (1982):

[P]robabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B. For example, when A is highly representative of B, the probability that A originates from B is judged to be high. On the other hand, if A is not similar to B, the probability that A originates from B is judged to be low. (Kahneman and Tversky, 1982, p. 4)

One way to understand representativeness is to think of it as a heuristic that applies specific stereotyping to judgment-making procedures. While stereotyping sometimes deserves negative connotations, the general application of stereotyping also serves us well in many situations. Take a couple of examples. Food that smells like it's decomposing should generally not be eaten (throwing out leftovers that smell like sewage is beneficial to our health, even if that means throwing out perfectly edible durian fruit). Making judgments about the character of a person solely because they phenotypically appear to belong to a certain race is usually not beneficial and often appalling, epistemically and morally speaking. However, Kahneman and Tversky discovered that we systematically (reliably) make mistakes in probabilistic reasoning when we (unconsciously) rely on the representativeness heuristic to make judgments about likelihoods. Those mistakes fall into various categories: discounting base rates, discounting the power of sample sizes, and misconceptions of chance (the gambler's fallacy), to name a few. I discuss each of these types of mistakes below, starting with discounting base rates.

How can the representativeness heuristic influence us to discount base rates? In a series of experiments (Kahneman and Tversky, 1982; Gilovich, Griffin, and Kahneman, 2002), researchers tasked subjects with judging how likely it was for a person to hold a certain profession after reading a short description of that person's personality traits. For example, in one experiment (Kahneman and Tversky 1982, pp. 4-5, 49-57) subjects were instructed to read a vignette that contained a brief description of a person who fit the stereotype of either a lawyer or an engineer. The subjects in one group were then told that

the base rate for lawyers in the target population was 0.3 and the base rate for engineers was 0.7, and another group was told the opposite base rates. When the subjects were asked to determine which profession (lawyer or engineer) that a person held, their responses ignored the given base rates for those professions and correlated instead to the stereotypes of people holding those professions. In this case, the representativeness heuristic seemed to override the explicitly stated base rates because the traits given in the description of the target person *represented* the stereotypical traits for a person holding that profession. Kahneman and Tversky's conclusion was that stereotypes outweighed explicitly stated probabilities (e.g., base rates here) as judgment-making evidence, which led them to investigate the representativeness heuristic further.

Kahneman and Tversky also hypothesized that the representativeness heuristic would skew judgments about events with given sample sizes. In another study (Kahneman and Tversky, 1982, p. 6), they asked subjects about whether a small-population hospital or a larger-population hospital would record, for instance, more than 60% male births over female births *on any given day*. As predicted, subjects were generally insensitive to sample sizes because they believed that the smaller and larger hospitals' respective daily births would have roughly the same percentage of girls to boys. Of course, the smaller hospital—with less total births, i.e., a smaller sample size—would have more days where 60% or more of births were, say, males. Kahneman and Tversky (sarcastically) named this propensity to ignore sample sizes the “law of small numbers” (a play on words of the “law of large numbers”, a term that describes regression to the mean). Kahneman and Tversky wondered if the subjects' insensitivities

to sample sizes were an artifact of deficient training in statistics and probabilities. So, they asked their colleagues and peers to participate in similar experiments that focused on sample size sensitivity. In a survey of peers from a mathematical psychology meeting and a survey of members of the American Psychological Association, they found that *even their peers* seemed to adhere to the law of small numbers. In their words:

Our questionnaire elicited considerable evidence for the prevalence of the belief in the law of small numbers. Our typical respondent is a believer, regardless of the group to which he belongs. There were practically no differences between median responses of audiences at a mathematical psychology meeting and at a general session of the American Psychological Association convention, although we make no claims for the representativeness of either sample. Apparently, acquaintance with formal logic and with probability theory does not extinguish erroneous intuitions. (Kahneman and Tversky, 1982, pp. 29-30).

This revelation is worth pausing over here. In many of the studies discussed below, it turns out that experts and so-called “sophisticates” (people who have been trained in statistical analysis or probability theory) will still more often than not make the same errors as untrained laypeople. In fact, this is a crucial foundational claim for the HBP: no one is completely immune to the unconscious misapplication of the heuristics that Kahneman and Tversky describe. Academics often read these studies and believe that the prevalence of these errors in reasoning are due to lack of education or carelessness. Not so. While experts in logic and probability theory do *somewhat better* with regard to arriving at the correct response to the problems presented, it is usually the case that the



*majority* of those highly trained subjects still make the same mistakes as the so-called layperson. This does not demean the academic's efforts to improve her reasoning skills; this is just further evidence that our brains are structured to use heuristics—correctly or incorrectly—before deliberately engaging in careful, methodical reasoning when making judgments under uncertainty.<sup>28</sup>

Finally, the representativeness heuristic can be found in use in casinos worldwide. When the history of a roulette wheel's previous outcomes is displayed (as it is in many casinos), most people will think that a run of, for instance, eight reds will ensure that the next outcome will be black. This type of predictive reasoning is often called the gambler's fallacy. While the fallacy is common to observe in gambling environments, the cognitive mechanisms by which they operate is not as well understood. Kahneman and Tversky's proposal for a representativeness heuristic offers a straightforward explanation of why the gambler's fallacy is prevalent in most people's gambling behaviors. While each spin of a roulette wheel has discrete odds, gamblers often place their bets by assuming that the history of previous outcomes somehow restricts the possibilities for future outcomes. This, of course, isn't how probabilities work. These misconceptions of chance fall under the representativeness heuristic because the heuristic "tells us" that a fair coin flip should look fair even within a small sample of flips (see above section on law of small numbers). However, when flipping a coin, getting eight tails in a row does

---

<sup>28</sup> In Kahneman's popular 2011 *Thinking Fast and Slow*, he explains the benefits of deliberative ("slow", a.k.a. "System 2") thinking to overcome some of the more egregious examples of heuristics gone awry. While this topic is tangential (though, not wholly unrelated) to my project, it is worth noting here that some of the errors in judgment caused by heuristics can be mitigated by purposefully engaging in careful, analytical thinking. However, since this is not a stated prerequisite for evaluating others as peers in the disagreement literature, I leave out some of Kahneman's (et al.) interesting findings in and after *Thinking Fast and Slow*.

not mean that the ninth flip will be more likely to be heads than the first or second or one-millionth flip will be heads. Oddly, realizing this does not seem to deter most of us from thinking that previous results indicate future performance (if this sounds familiar, the same advice is often given to stock market investors: previous performance is not a guarantee of future gains, according to standard warnings).

The researchers discussed so far have contributed to Simon's project by setting the boundaries for bounded rationality. If the heuristic that leads us away from rotting food also makes us racists and bad gamblers, this is a significant discovery about the descriptive limitations of human cognition. And, as I explain below (see section 5.4.2), biases that potentially arise from the misapplication of the representativeness heuristic are specifically relevant to the peer disagreement project. But, before I get ahead of myself, there are two other heuristics to discuss.

### **5.3.2: Availability Heuristic**

Kahneman and Tversky (1982) describe the availability heuristic as activating in "situations in which people assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind" (p. 11). In an early study (Kahneman and Tversky, 1982), researchers asked subjects to judge whether a certain letter was more common as the first or third letter in English words. For instance, one of the test questions was whether the letter *r* was more common as the first or third letter in English words. The hypothesis was that subjects would more easily recall words that began with the letter *r* rather than words where *r* was in the third position (when thinking of the spelling of words, we tend to start with the first letter

rather than a latter letter). Indeed, subjects thought of more words that began with *r* rather than words where *r* is the third letter, even though the letter *r* is more common in the third position of English words compared to the first position.

The letter position experiment might seem like a mere trivia question, but the availability heuristic also affects real-world relationship dynamics, judgment about the frequency of lethal events (i.e., how likely something is to kill us), and how jurors make judgments in court cases. Ross and Sicoly (1979) designed an experiment where married couples were asked to estimate how much of the housework they performed relative to their partners. Unsurprisingly, each person in each couple claimed that they contributed more (roughly 60% or higher) to the housework than their partner. While there are various reasons why people believe they contribute more than the average person in a given situation, the availability heuristic explains this tendency well; we can more easily recall the times that we've done the dishes, changed the diapers, taken out the trash, or prepared a meal than we can recall when our partners performed those tasks. Why? Because, in these cases, it is easier to remember our own actions than the actions of others. It is also easier to recall events from sensationalized news stories when estimating likelihoods of various events. In another interesting early experiment by Lichtenstein et al. (1978), researchers asked subjects to estimate which of two causes of death were more likely through a series of pairwise comparisons for various causes of death. The researchers also surveyed recent newspaper articles to determine the relative frequency that each cause of death was mentioned and how much (in printed inches) each cause of death was discussed in the article. Reber (2017) summarizes that study succinctly:

Although stomach cancer is more than five times more frequent than homicide, participants estimated that homicide is about 1.6 times more frequent than stomach cancer [as a cause of death]. Moreover, media coverage was high for homicides, but zero for stomach cancer, and media coverage predicted the frequency estimates of causes of death. The authors (Lichtenstein et al.) concluded that estimates of [the] frequency of lethal events are based on high availability of vivid or sensational events. (Reber, 2017, p. 190)

From this experiment, it is apparent that the availability heuristic helps us make judgments about probabilities by pulling information from our most salient memories of information related to those probabilities. In cases when we need to update our priors rapidly to increase our odds of survival, the availability heuristic might serve us well. We can easily imagine early humans needing to recall which berries are safe to eat or which hunting paths are more dangerous by employing this heuristic, for example.<sup>29</sup> However, as in the case of making judgments about risks of dying from a particular cause of death, the availability heuristic can also lead us astray due to the way we consume information about causes of death (i.e., what is most popular and publicized versus what is most prevalent and objectively likely). Interestingly, the availability heuristic does not merely use information that is most recent or present in our memory; the heuristic also operates by using information that is most “vivid or sensational.”<sup>30</sup> In a study by researchers at

---

<sup>29</sup> This is, admittedly, my quick and dirty just-so story. I find the story plausible enough to move on, but I make no claims about it being true.

<sup>30</sup> The primacy / recency effects on memory recall are well documented (see Kahana 2014 pp. 189-196 for more), but the availability heuristic *also* functions by using the somewhat opaque concept of “vividness” to inform our judgments. So, the information used by the availability heuristic need not be recently acquired, though recency is probably sufficient for explanations of how the availability heuristic functions. However,

Arizona State University, students who were asked to imagine easy-to-imagine symptoms of a fictional illness were more likely to believe that that illness was more likely to affect them than when the symptoms of that same fictional illness were less-easy-to-imagine (Sherman et al. 2002). That is, the vividness with which one can imagine an event affects how “easy” it is to bring that event to mind when one is tasked with estimating the likelihood of the event occurring. Finally, in a study by Reyes et al. (1980), mock jurors were swayed by evidence that was presented more vividly for the prosecution than the defense—or vice versa—for a case centered on a drunk driving incident. In these experiments, one set of mock jurors were presented with evidence *from the defense* that was more vivid (more descriptive) than the prosecution. Those mock jurors were more likely to enter a verdict of *not guilty* for the defendant. However, when another set of mock jurors read about the same case where *the prosecution’s evidence* was presented more vividly, those jurors were more likely to *enter a guilty verdict* for the same defendant.

The availability heuristic may work well for survival in some cases, but it leaves much to be desired. Whether we are trying to develop healthy domestic relationships, deciding which causes of death we ought to focus on preventing, or performing our civic duties as jury members, the availability heuristic sometimes leads us astray.

---

due to the vividness effect, an old but perhaps traumatic or otherwise emotionally charged memory from the subject’s distant past may also influence how the availability heuristic prioritizes the information that it selects to produce “answers”.

### 5.3.3: Anchoring Heuristic

The anchoring heuristic is a well-established heuristic that exemplifies the issues that arise when using heuristics to make judgments under uncertainty. Consider what Bahník, Englich, and Strack (2017) write:

“Anchoring” is defined as the assimilation of a judgment to a previously considered standard. Anchoring is a remarkable influence on human judgment, for at least two reasons. First, anchoring effects are very pervasive and robust. Second, the mechanisms that produce anchoring are after many years of investigation still a matter of lively debate. (Bahník et al., 2017, p. 224)

What does it mean for a judgment to be assimilated to a previously considered standard? Here is a simple example: If I ask you how many jellybeans there are in a large jar, you’ll hazard a reasonable guess based on various visual and mental calculations. However, if I ask you instead if there are more or fewer than 800 jellybeans in the jar, your guess will be pulled towards (i.e., anchored and adjusted towards) the number 800, even if it is obvious that the jar contains far more or far fewer than 800 jellybeans. In terms of studying how heuristics help or hinder our rational capabilities, a plethora of experiments have demonstrated that we often use irrelevant information to make judgments.

Furthermore, anchoring’s “pervasive and robust” presence in human cognition also helps us understand the nature of how heuristics invisibly influence our most basic judgments. Several studies (see Bahník et al. 2017) have found that we are affected by neuromarketing strategies that take advantage of the anchoring heuristic. Additionally,

anchoring also affects sentencing decisions made in real-world criminal justice settings. I discuss examples of each below.

Neuromarketing takes advantage of the anchoring heuristic in the form of the “limit X per customer” signs that are often displayed in grocery and other stores (Chapman and Johnson, 2002). For instance, customers who read a sign that stated a sale item has a “limit of 12 per customer” bought closer to 12 of that item than customers who weren’t exposed to that sign. Additionally, we often witness examples of neuromarketing that use the regular price vs. sale price method. When we are told an item was priced at, say, \$89.99, but is now on sale for \$30 (infomercials and social media ads are commonplace examples of this tactic), we often believe the sale price is a good bargain regardless of the market value of the item.

We find similar effects in courtrooms during sentencing hearings. Judges and prosecutors who were asked how long they would sentence or recommend sentencing, respectively, were influenced by questions from journalists about the sentence duration. For instance, for convictions of similar crimes, some judges were asked if they would sentence the person to one year while other judges were asked if they would sentence the person to three years. On average, the judges who were asked whether they would sentence the person to three years gave sentences that were eight months longer than the judges who were asked if they would give a sentence of one year (Bahník et al., 2017). So, while we might shrug off anchoring-influenced judgments about jellybeans, there are real-world consequences of anchoring effects.

The brief discussion of the studies discussed above are just a small sample of how the anchoring heuristic affects our daily lives (see Gilovich, Griffin, and Kahneman 2002 and Pohl 2017 for more examples). What is interesting and germane to this project is that the studies on the anchoring heuristic demonstrate that we are easily influenced by irrelevant information when making judgments. The fact that these influences are invisible to us is cause for concern, especially when we consider what happens when these cognitive illusions become systematic errors in judgment, otherwise known as cognitive biases.

#### **5.4: Cognitive Biases**

In the overview of the HBP above, I discussed the various ways in which heuristics lead us to incorrectly judge the likelihood that certain states of affairs obtain (or will obtain) and discussed how our heuristics might compel us to use irrelevant information when making judgments. Our undeniably valuable heuristics sometimes lead us astray when we make judgments under uncertainty. When our heuristics **systematically** lead us astray, the HBP designates those patterns of systematic errors as **cognitive biases**. This is not to say that all cognitive biases can easily be charted back to a defined or well-studied heuristic, just that the “heuristics” part of the HBP led to the discovery of robust and pervasive systematic errors in reasoning. These predictable patterns of systematic *mis*judgments—cognitive biases—have been demonstrated extensively in psychological experiments that have withstood the replication crisis. In fact, studies on cognitive biases have been used to reasonably *explain* the causes of the replication crisis in psychology (Bishop, 2020). In any case, we (you, dear reader, and I)



are biased. Perhaps not all the time and certainly not to such an extent that we are hopelessly irrational in every judgment we make. But, more often than not, when we make judgments without complete information, odds are that we are being affected by the unconscious misapplication of various heuristics, or: cognitive biases.

According to the research on cognitive biases, it is highly likely that biases affect many of our everyday judgments. This is important to remember for understanding the rest of this project; if, in light of the results of research on cognitive biases, our judgments are often biased, then this descriptive fact about human cognition ought to be accounted for when discussing how we actually evaluate our so-called epistemic peers.

In the sections that follow, I focus on how the effects of specific cognitive biases pertain to my thesis. As a reminder, the short version of my thesis is: Due to the effects of cognitive biases, we are not cognitively well-suited to perform epistemic peer evaluations. However, the discussion of these specific cognitive biases that follows is not in any way intended to be exhaustive. The complete list of cognitive biases that may be relevant to the descriptive limitations of our abilities to discern who our epistemic peers are is much longer. Additionally, not only are there many other cognitive biases that affect our judgments in ways that are relevant to epistemic peer evaluations, but there are also a host of other general cognitive and perceptual “illusions” that play a role in defining the descriptive boundaries that real-world agents are constrained by when encountering disagreements (e.g., memory illusions, optical illusions, tactile illusions, motivated reasoning, emotional reasoning, etc.). Here, I pick out only four cognitive

biases that are pertinent to defending my thesis. But we should keep in mind that these cognitive limitations—*boundaries*, if you will—are only the proverbial tip of the iceberg.

Below I discuss studies on four cognitive biases: over- and underconfidence bias, halo and horns effects, confirmation bias, and bias blind spot. Note that if the majority of experimental subjects (from appropriately representative samples) exhibit these biases—and they do—it is likely that the majority of humans (you and I and others) also exhibit these biases. This means that more people than not will be affected by these biases, including people who are put in a position to make epistemic peer evaluations during disagreements.

#### **5.4.1: Cognitive Biases: Evaluating Our Own Competencies and Abilities**

Overconfidence and underconfidence biases are well documented and highly prevalent. These biases affect how we evaluate our confidence in our beliefs and abilities. However, while many studies on confidence focus on the ways in which subjects are overconfident in their respective epistemic (or doxastic) positions, there are also conditions in which they are reliably underconfident (skeptical and unsure about their convictions), and this spectrum of overconfidence and underconfidence is important to remember for my discussion in Chapter 6 on epistemic peer evaluations. Here, I'll start with Griffin and Tversky's (2002) summary of the research program on overconfidence:

The weighting of evidence and the formation of belief are basic elements of human thought. The question of how to evaluate evidence and assess confidence has been addressed from a normative perspective from philosophers and statisticians; it has also been investigated experimentally by psychologists and

decision researchers. One of the major findings that has emerged from this research is that people are often more confident in their judgments than is warranted by the facts. (p. 230, Griffin and Tversky (posthumously), 2002)

When we evaluate our own rational competencies and abilities, we are often overconfident (and sometimes underconfident). Researchers use what they call *calibration measures* of confidence (epistemologists may roughly think of these measures as imprecise belief credences) to demonstrate that we are often not good judges of how confident we ought to be (see Hoffrage 2017). A calibration measure compares a subject's judgment of their confidence in their accuracy in response to a survey question relative to the objective accuracy of that response. For instance, on a half-scale model, subjects may be presented with forced-choice questions such as, "What is absinthe: a precious stone or a liqueur?" and, after making their choice, the subjects report their confidence level that their response is correct on a scale from 50% to 100% in 10% increments.<sup>31</sup> In these experiments, the majority of subjects report higher confidence in their responses than they ought to when their responses are compared to the objective facts (Hoffrage 2017, Griffin and Tversky 2002). For general knowledge questions or memory abilities, subjects are generally overconfident relative to the objectively correct responses to the questions. This tendency is called the **overconfidence effect**.

Overconfidence effects spill over into other rational self-evaluation tasks as well. In "better-than-average" studies, most subjects believe that they are above-average when

---

<sup>31</sup> Note the half scale is used here because with a random guess on a forced-choice binary question, it would be rational to be at least 50% confident even on a random guess. See Hoffrage 2017 for differences in half-scale and full-scale (0% to 100%) calibration measures for overconfidence experiments.

evaluating their competencies or abilities relative to others. For instance (gathered from various sources noted in Hoffrage 2017):

- 19% of people think that they belong to the richest 1% of the population.
- 83% of people say they are in the top 30% of safe drivers
- 80% of students think they will finish in the top half of their class
- 68% of lawyers in civil cases believe that their side will prevail
- 86% of [members of a] Harvard Business School [class] say that they are better looking than their classmates (Hoffrage, 2017, p. 294)

While there are some mitigating factors here<sup>32</sup>, we have all encountered people who believe that they are far more skilled, intelligent, attractive, etc. than they ought to.<sup>33</sup> So, I won't belabor the point here.

However, we have also encountered people who believe the opposite: they believe that they are *below average* relative to their peers. This is in part due to the **underconfidence effect**, which is also a well-documented psychological phenomenon (Hoffrage, 2017). Underconfidence effects are most noticeable in studies when the survey questions are *exceptionally easy*. For instance, when subjects report that the general knowledge questions on an assessment were “easy”, researchers noted that many subjects

---

<sup>32</sup> For example, a skewed curve representing actual driver safety could show that 10% of drivers cause 90% of accidents, which would somewhat vindicate the 83% of people who say they are in top 30% of safe drivers, etc.

<sup>33</sup> The well-documented Dunning-Krueger Effect demonstrates this propensity for overconfidence at various levels of education and training. However, the Dunning-Krueger Effect is also misleading; overconfidence is not restricted to people who have some training in a subject but not comprehensive or no training in that same subject. That is, according to general research on confidence, most people – regardless of the amount of training in a subject or skill – are overconfident in their knowledge or beliefs about that subject or skill.

were underconfident in their responses (Hoffrage, 2017). This is often called the “hard-easy effect”. When the target questions are hard (as judged by the study participants), an overconfidence trend appears. However, when the target questions are easy (also as judged by the participants), an underconfidence trend appears. This phenomenon is interesting in that it is counterintuitive: faced with difficult questions, we ought to be less confident in our correctness and vice versa for our confidence in our responses to easy questions. Studies on confidence-level judgments show just the opposite (Hoffrage 2017).

Various models have been proposed to explain why it is that we are sometimes over- and underconfident. The details of these models are not relevant for my purposes, though it is important to note that several theories have been proposed to explain overconfidence as a general trait of human psychology. The prevailing theory is, essentially, that a “fake it till you make it” attitude is advantageous for practical human flourishing. In an example given by Hoffrage:

Convincing others is obviously easier if we are convinced ourselves in the first place—and possibly even more than would be justified (Trivers, 2011). Consider a physician who is overconfident that a particular treatment will benefit her patient. Showing high confidence that it will help may be essential for a placebo effect to occur. If the objective chances that the treatment will help are, a priori, 30%, and if they increase, objectively to 60% a posteriori (i.e., after the physician expressed a very high confidence of, say 80%) who wants to blame her for being overconfident? After all, it helped her to be convincing which, ultimately, helped the patient. (2017, p. 308)

Are there times when we can feel secure in our judgments about our confidence assessments? One study (Murphy & Winkler, 1977) showed that meteorologists were accurate in their estimations of their confidence in forecasts of precipitation relative to actual precipitation outcomes, but the authors also showed that various other highly-trained experts such as physicians, lawyers, clinical psychologists, engineers, and security analysts are were generally overconfident in their respective real-world predictions about the outcomes in their respective fields (see Griffin and Tversky, 2002). That is, there is scant but perhaps promising evidence that some experts have appropriate levels of confidence in their abilities and beliefs. However, in general, even experts making predictions in their own domains of expertise are not immune to overconfidence effects when their predictions are scrutinized under controlled conditions.

As with other biases, over- and underconfidence effects are well-documented, prevalent, and resistant to expertise and sophistication. So, as is the case with many cognitive biases, it is highly likely that over- and underconfidence biases surreptitiously affect all of us more often than we'd like them to.

#### **5.4.2: Cognitive Biases: Evaluating the Abilities of Others**

Everyone reading this sentence will have had some experience evaluating the abilities of others. More specifically, many readers of this dissertation will have a good deal of experience formally assessing the epistemic abilities of their students. As educators, we typically use explicit or implicit rubrics to assign grades to our students, and the grading process is a paradigmatic example of what it means to evaluate others' epistemic abilities. Even if the reader is not an educator, it is likely that you have

evaluated the epistemic abilities of other people several times today. Evaluating others' epistemic abilities is fundamental for navigating the social world.

However, sometimes when we evaluate the epistemic abilities of others, our evaluations and subsequent judgments are affected by biases. One such bias is called the **halo effect**, which “occurs when perceivers make unwarranted inferences about the positive or negative qualities of a person based on information about other, unrelated characteristics” (Forgas and Laham 2017, p. 289). If the description of this effect seems familiar, that is because it's reminiscent of stereotyping and prejudice (recall how the representativeness heuristic affects our reasoning in terms of stereotyping).

“Stereotyping”, according to the way we commonly use the word, often refers to making judgments about the abilities of others based on their race, ethnicity, sex, gender, etc., and falls into a subcategory of halo and horns effect biases. However, halo effects occur in situations beyond our typical definition of stereotyping. For instance, halo effects have been observed when the independent variable represented a relatively homogenous population, and the usual categories of discrimination and prejudice are not found. For example, the halo effect was shown to influence schoolteachers grading a sample essay labeled with either a common white male name (David, Michael, etc.) or an unusual but typically white male name (Elmer, Hubert, etc.). In this experiment, the essays labeled with common white male names earned more favorable grades than the essays with less familiar names (Harai and McDavid, 1973).

The halo effect was also prevalent when subjects evaluated the cognitive abilities of others where physical attractiveness was the independent variable (Forgas and Laham,

2017). That is, subjects judged more physically attractive people as smarter, more intelligent, or more likely to do well on academic assessments than less physically attractive people (Forgas and Laham, 2017). Further studies of halo effects found that height, weight, names (familiar and unfamiliar), and previous but unrelated performance all influenced subjects' judgments of the intelligence and capabilities of others as well (Forgas and Laham, 2017). Additionally, halo effects are resistant to several direct mitigation strategies. In the words of Forgas and Laham (2017):

Halo effects appear extremely reliable and pervasive, and it seems that explicit interventions and instructions are not particularly effective in reducing their occurrence. For example, Wetzel, Wilson, and Kort (1981) tried to eliminate the halo effect using Nisbett and Wilson's (1977) procedure, by (1) asking participants to pay close attention to what they were feeling, (2) giving participants a prior description of the halo effect and telling them to try [to] avoid committing it, or (3) giving a description of the halo effect and instructing participants to commit it. None of these interventions had an influence on the size of the halo effect. (Forgas and Laham, 2017, pp. 288-289)

However, Forgas (2011) did show that the halo effect was reduced or eliminated when participants were in a bad mood. Some studies have found that negative affect tends to transition participants into a more analytical and critical mindset (Forgas and Laham, 2017), and so the proposed explanation for this mitigating strategy is that people who are in a bad mood tend to disregard the social cues that contribute to the halo effect.



We are also prone to exhibit a **horns effect** (a.k.a., negative halo effect or devil effect) (Forgas and Laham, 2017). The horns effect is the shadowy complement to the halo effect: subjects will often evaluate others as less capable if an unrelated negative attribute is present. I'll spare the reader examples here; think of things a bully might say to a victim as a list of the negative attributes studied. However, it is important to remember that, as with over- and underconfidence biases, halo and horns effects work "both ways."

Finally, halo and horns effects tend to be described as biases that affect how we judge others' competencies during a brief interaction such as a "first impression" situation. However, in a six-week longitudinal study, Jacob and Kozlowski (1985) found that halo and horns effects are not mitigated by increasing the time the subjects spent with those they were evaluating. In fact, the results from that study supports the idea that "familiarity breeds contempt" since some subjects rated their targets less and less competent over time based on attributes that were not relevant to their target's competencies. So, being acquainted with someone over time (as one might be if they had ordinary track records with that person) does not seem to eliminate the halo or horns effect.

The halo and horns effect is both pervasive and reliable in empirical studies, and, as is the case for the other biases discussed here, the effect is also resistant to debiasing strategies thus far conceived. While we are acutely aware of common stereotyping biases—such as biases in favor of or against people of certain races, genders, etc.—halo / horns effects are not often openly discussed as causes for those stereotyping biases.

However, again, the halo / horns effect is both robust and prevalent, and its effects are not limited to the typical prejudices that get the most attention; we judge others' general intelligence (a requirement of epistemic peer evaluations) based on height, body fat ratio, attractiveness, familiarity of name, and many other factors that are usually irrelevant for determining the epistemic qualities of a person. One can only imagine how drastically the halo or horns effect might lead us to misjudge the relative epistemic standing of others based on the typical racial and gender-based prejudices alone. But importantly those misjudgments are not the only ones we would need to manage to combat the halo and horns effect.<sup>34</sup>

So far, this discussion has focused on how we evaluate the epistemic standing of ourselves and others. However, biases also affect how we evaluate incoming information as well, both in terms of how we seek evidence and how we process available evidence. In the next section I discuss the findings of studies on the widely mentioned yet frequently misunderstood cognitive bias that descriptively affects how we process information: confirmation bias.

---

<sup>34</sup> Here it is important to clarify that the research on the halo / horns effect is not the same as, though possibly connected to, Greenwald et al.'s implicit association test (IAT) and the so-called implicit biases that have been postulated. IAT has not fared well in replication and has a host of credible critics and experimental design problems. This is not to say that we do not have implicit biases (by definition, the cognitive biases I discuss here are implicit or not accessible to consciousness), but rather that IAT and the "implicit bias program" (to sum up that research) is distinct from the research I cite about the halo and horns effects. Unfortunately, the research I discuss here suggests that most people make decisions about others' epistemic status based not only on perceived race and gender, but also based on multiple other phenotypical (e.g., height) and contingent (e.g., first names) factors. That is, distancing halo / horns effect research from IAT does not lead to more potential for optimistic outcomes. However, the research I discuss here is well-established, replicated, and robust, unlike the IAT program's research.

### 5.4.3: Cognitive Biases: Evaluating Evidence in General

One of the most pernicious and pervasive biases that humans exhibit is confirmation bias. In the 17<sup>th</sup> century, Francis Bacon remarked in his book of aphorisms, *Novum Organum* (1620), that people tend to search out facts that support their own views while ignoring information that would be contrary to their opinions. Four hundred years later, psychologists have confirmed Bacon's observations about how we seek and handle evidence that relates to our most cherished beliefs.

Psychologist Raymond Nickerson (1998) defines confirmation bias as “the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand” (p. 175). Below, I draw on studies by Nickerson and others to explain the various types of confirmation biases that affect many of us in our daily lives.

Consider some types of confirmation bias that have been experimentally demonstrated:

- *Restriction of attention to a favored hypothesis*: Subjects ignore alternative hypotheses even when diagnostic observations are presented. A diagnostic observation is evidence that two or more competing hypotheses are roughly equal in terms of explaining some phenomenon. By ignoring the likelihood that an alternative and disconfirming hypothesis may be equal to or higher than one's favored hypothesis, subjects irrationally maintain that their favored hypothesis is the most likely of all competing hypotheses simply by ignoring alternatives (Nickerson, 1998, p. 177-8).

- *Preferential treatment of evidence supporting existing beliefs*: Closely related to restriction of attention to a favored hypothesis, preferential treatment of evidence occurs when subjects give greater weight to information that supports their existing beliefs while giving less weight to information that contradicts their existing beliefs (Nickerson, 1998, p. 178-9). This type of confirmation bias is often referred to as the *my-side* bias because subjects unconsciously give more weight to evidence that supports their own existing beliefs. For example, Stanovich and West (2008) performed an experiment where subjects were surveyed on their beliefs and behaviors about a range of topics (e.g., belief in God, smoking tobacco, etc.) and were given a series of propositions about each topic (e.g., “Religious people are generally more honest than non-religious people”, “Secondhand smoke is a health hazard for non-smokers”, etc.). Unsurprisingly, subjects (in aggregate) agreed with propositions that supported or aligned with their pre-existing views and rejected propositions that did not support or align with their pre-existing views regardless of whether the propositions made claims that were either unsubstantiated or substantiated.
- *Looking only or primarily for positive cases*: When subjects are asked to test a hypothesis, they tend to only search for ways to confirm the hypothesis rather than to disconfirm it. Peter Wason’s (1960) famous selection task asks subjects to test the following rule: “If there is a vowel on one side, then there is an even number on the other side” and are then shown four cards with the following letters or numbers showing face up: A, 7, D, 4. Subjects are then asked which cards need

to be turned over to determine whether the rule is correct. Most participants select either just the A card or both the A and 4 cards. However, checking the reverse of the 4 card does not help us determine if the rule stands since the rule doesn't say what should be on the other side of an even-numbered card.<sup>35</sup> The D card is similarly useless for rule checking since the rule says nothing about consonants. However, checking the 7 card could disconfirm the hypothesis if there were a vowel on the other side of that card, and checking the A card is useful either way because it can confirm or disconfirm the rule. Other studies (see Nickerson 1998 for several) demonstrate that subjects tend to search for confirmatory (positive) rather than disconfirmatory (negative) evidence when testing hypotheses.

- *Overweighting positive confirmatory evidence*: Related to only looking for positive evidence, subjects also tend to give more weight to evidence that confirms their hypotheses and give less weight to evidence that disconfirms their hypotheses (Nickerson, 1998). A classic example is found in people who believe that astrology can accurately predict the future. "People who wish to believe in astrology or the predictive power of psychics will have no problem finding some predictions that have turned out to be true, and this may suffice to strengthen their belief if they fail to consider either predictions that proved not to be accurate or the possibility that people without the ability to see the future could make

---

<sup>35</sup> Hugo Mercier (2017) - and others before him - argues that the Wason selection task does not demonstrate that subjects have a confirmation bias because the subjects are simply checking the cards that are most relevant to the rule: the vowels and the even-numbered cards. According to Mercier, the Wason selection task is flawed since, by pragmatic inference, checking the mentioned cards is the rational thing to do in most cases. I don't find Mercier's argument convincing, but it is worth noting that this type of confirmation bias has its critics.

predictions with equally high (or low) hit rates” (Nickerson, 1998, p. 180).

Overweighting positive confirmatory evidence is often referred to as “**counting the hits and missing the misses.**”

- *Seeing only what one is looking for:* Part framing effect, part halo effect, and part self-fulfilling prophecy effect, seeing only what one is looking for is essentially our tendency to look for patterns in data where there are none or to assign attributes to people based on association rather than investigation (Nickerson 1998). For example, in one experiment (Darley and Gross, 1983), a random selection of subjects was divided into two groups and tasked with assessing the academic abilities of a child completing an academic test. The first group were told the child’s socioeconomic background was low and the second group were told the child’s socioeconomic background was high. As expected, the first group rated the child’s performance as above grade level while the second group rated the child’s performance as below grade level (note that all subjects were presented with the same videotaped footage of the child taking the test). Seeing what one is looking for has been demonstrated in many different contexts through tasks that involve stereotyping behaviors and forming illusory correlations; at the extreme end of this type of bias is hypochondria (health anxiety) and psychotic paranoia (Nickerson, 1998).

As we can see, confirmation bias comes in many guises and affects the way we seek and reason about our evidence. And, as with the other biases discussed so far, confirmation bias is ubiquitous and resistant to mitigation strategies.

#### **5.4.4: Cognitive Biases: Introspective Bias Detection Failure**

In the preceding sections, I discussed several well-established cognitive biases that affect the way we evaluate our own competencies and abilities, the way we evaluate the competencies and abilities of others, and the way we search for and weigh evidence for our beliefs. A common reaction to the news that cognitive biases skew the way we evaluate the world often goes as follows: “Well, that may be true for others, but not for me. I would know if I was biased.” Immediately, we should recall overconfidence and above-average bias, and we should suspect that line of thinking is flawed. Not only does the empirical evidence strongly suggest that it is a mistake for anyone to believe that they are immune to the effects of these biases, but there is also an additional bias that describes this behavior: the bias blind spot (BBS). Social psychologist Emily Pronin (2002) describes BBS as the tendency for people to recognize bias in others but not in themselves. In a series of experiments by Pronin (2002, 2007), 70% - 80% of subjects ranging from psychology undergraduate students to San Francisco airport patrons attributed more biased reasoning to others than they do to themselves. This was true even in some versions of the experiment when those subjects were explicitly told that: (1) the vast majority of people are affected by cognitive biases, and (2) that we tend to attribute more bias to others than we do ourselves (i.e., they were told that BBS affects our reasoning). Furthermore, studies on BBS have been replicated robustly in the US and in a cross-cultural sample from Hong Kong (Chandreshekar et al. 2021). At first, these results may seem to make sense; it seems reasonable to believe that we understand our own thinking processes better than we can understand others’ thinking processes (recall from

Chapters 2 and 3 that this was the precise line of thinking from Foley 2001 and Fumerton 2010). When we are shown evidence that a large majority of people are susceptible to biases, it is reasonable to attribute biased reasoning to other people (after all, we don't want to commit the fallacy of ignoring the base rate). That is, when making decisions about the competencies of ourselves relative to others, it is more likely that *any* given person is operating under the influence of a bias than not. So, when guessing whether someone (anyone) is affected by a cognitive bias in a given scenario (experimental or otherwise), odds are better than a coin flip that the person's judgment is being skewed by some bias. However, when it comes to evaluating whether we ourselves are biased, we tend to believe that we have a special sort of access to our thinking processes that makes us immune to the tendency to be biased, despite strong and robust evidence to the contrary. One of the leading explanations for this behavior, posited by Pronin (2002), is that we rely on a form of naïve realism where we believe the world just is as it seems to us. However, since we don't have access to the way the world seems to others, we can't be sure if *they* are misinterpreting *their* evidence. In Pronin's words:

This proposal of an asymmetry in perceptions of bias arises from recent accounts of "naïve realism" (Griffin & Ross, 1991; Pronin, Puccio, & Ross, 2001; Ross & Ward, 1996; also see Ichheiser, 1970), which hold that people think, or simply assume without giving the matter any thought at all, that their own take on the world enjoys particular authenticity and will be shared by other openminded perceivers and seekers of truth. As a consequence, evidence that others do not share their views, affective reactions, priorities regarding social ills, and so forth



prompts them to search for some explanation, and the explanation most often arrived at, we argue, is that the other parties' views have been subject to some bias that keeps them from reacting as the situation demands. As a result of explaining such situations in terms of *others'* biases, while failing to recognize the role of similar biases in shaping their own perceptions and reactions, individuals are likely to conclude that they are somehow less subject to biases than the people whom they observe and interact with in their everyday lives. (2002, pp. 369-370)

In the sections above, I discussed how general heuristics and specific biases have been shown empirically to affect the way we reason about our abilities, the abilities of others, and the way we process information. When people are confronted with this information about biases, it is not uncommon for them to think that biases affect others but not themselves. However, according to Pronin et al.'s research, this way of reasoning is in itself the result of a cognitive bias (we might think of bias blind spot as a sort of meta-bias). The takeaway from the research for now is this: Most people are affected by cognitive biases, *and* they are unaware of when they are being affected by those biases. This exacerbates the problems caused by biases because while we cannot stop them from occurring, we often feel justified in attributing biases to others to rationalize defending our beliefs.<sup>36</sup>

---

<sup>36</sup> Nathan Ballantyne (2015) describes this line of thinking as “debunking reasoning” (pp. 142-144). After a careful examination of four different strategies we could use to justify debunking reasoning, Ballantyne argues that none of those strategies are promising, especially in light of the data we have on bias blind spot. His conclusion, which I partially agree with, is that the lack of any good strategies to defend debunking reasoning suggests we ought to exercise more intellectual humility when engaged in disagreements (and in general).

## 5.5: The Cognitive Bias Problem

In the preceding sections, I discussed several cognitive biases and how they affect our judgments. It is worth summing up the general points about cognitive biases here so that I can refer back to the findings presented in this chapter simply as the **Cognitive Bias Problem** (CBP). The following facts comprise the CBP and will be applied during my evaluation of epistemic peer evaluations in the next chapter:

**The Cognitive Bias Problem:** The cognitive bias problem is, in general, the claim that empirical studies on human cognition and judgment-making under uncertainty demonstrate the conjunction of the following claims:

- **Cognitive biases are prevalent in the general population.** In most of the studies cited above and in the later replications of those studies, it is common for 70% or more of subjects to display the biases under examination. The typical person is, then, more likely to be biased than not in situations similar to the experimental conditions.
- **Cognitive biases are undetectable through introspection or by self-monitoring behaviors.** The heuristics that we use to make decisions about likelihoods, our own competencies, the competencies of others, and how we seek and weigh information operate on an inaccessible, subconscious level. We usually aren't aware of when cognitive biases are affecting our judgment.

- **Cognitive biases are resistant to mitigation strategies.** In many experiments that replicated the early findings of the HBP, researchers attempted a variety of common-sense interventions that would presumably reduce or eliminate the biases they were finding in their studies. Unfortunately, these debiasing and mitigation strategies have only been marginally effective, if at all.
- **Cognitive biases affect experts, too.** Being well educated, highly intelligent, or keenly aware of the studies on cognitive biases does little to reduce one's susceptibility to bias. People who are experts in a domain tend to exhibit the biases discussed above. Again, it is reasonable to believe that biases affect everyone.

## **5.6: Summary and Conclusion of Chapter 5**

In this chapter, I discussed Herbert Simon's proposal for a theory of bounded rationality. Through the work of Daniel Kahneman and Amos Tversky (among many others discussed above), psychologists gathered the empirical data that Simon's theory needed to set the boundaries of rationality. The early experimental results from the HBP led to the discovery that we exhibit patterns of systematic errors in reasoning when we make judgments with incomplete information. These errors are the results of cognitive biases, and these biases have continued to be cataloged since the inception of the HBP over some fifty years ago. After I briefly reviewed the evolution of the HBP, I then discussed studies on several specific robust cognitive biases. Those biases affect the way we judge our epistemic capabilities relative to others (overconfidence / underconfidence

and halo / horns effects) and the ways that we select and process information (confirmation or my-side bias). I also discussed findings of a particularly interesting bias that often leads us to believe that we are not as biased as others, the so-called bias blind spot. Finally, I summarized what was presented in this chapter as the Cognitive Bias Problem, which is a succinct way to refer back to the general findings I discussed above. With that in mind, this chapter supports the following premise:

**(P4) Cognitive biases affect our judgments in ways that we are not aware of when we evaluate our own epistemic abilities, the epistemic abilities of others, the evidence we use to justify our beliefs, and whether we are more or less biased than other people.**

In the next chapter, I return to the realm of peer disagreement and discuss how these biases specifically could and very likely do affect epistemic peer evaluations.

## **CHAPTER 6: RELATIVE EPISTEMIC PEER EVALUATIONS AND THE COGNITIVE BIAS PROBLEM**

In this chapter, I explain what is required to evaluate ourselves and others as peers. First, I explain what a relative epistemic status evaluation is, and I discuss the possible epistemic statuses we can assign to ourselves and others. Next, I discuss the connection between the evidence condition and the virtue condition and argue that they do not come apart easily. Then, I continue my discussion from Chapter 3 about problematic definitions of peerhood that are too strict or too loose in order to find a case of peerhood that is neither. With those preliminary considerations in place, I return our attention to a fitting paradigmatic case of peerhood, Christensen's RESTAURANT, as a

test case for my thesis. After slightly modifying RESTAURANT, I give specific examples of how the Cognitive Bias Problem (CBP) could affect the hypothetical interlocutors' peer evaluations in that paradigmatic case. To do this, I examine how over- and underconfidence, halo and horns effects, and confirmation bias could affect those interlocutors if they were treated as though they were actual people.

Ultimately, if the CBP's effects on our interlocutors' peer evaluations in that paradigmatic case are plausible, then those effects will also be plausible both for the other representative cases of peer disagreement raised in the literature and, more importantly, for cases of real-world disagreements. I conclude this chapter with a sketch of my central argument and, finally, I end with a sharpened, valid version of that argument to demonstrate my main thesis.

### **6.1: Relative Epistemic Status Evaluations (RESE)**

It is not lost on me that the peer disagreement literature is primarily about disagreement and not about peers. The spirit of the debate implores readers to assume that there are epistemic peers (ontological claim) or that we can reasonably believe that others are our epistemic peers (epistemological claim). That is, the authors contributing to the literature presume that it is at least plausible that we can perform epistemic peer evaluations accurately. Much like assuming that the Uniqueness Thesis is true to motivate the problem of peer disagreement, we are also expected to assume that epistemic peerhood is commonplace. By skipping over the worries about whether we actually have any epistemic peers, the authors can focus on the epistemic significance of

peer disagreement unhindered by descriptive limitations. This provides them with the freedom to explore the main question in the core debate.

Recall that the debate in the literature is centered on the normative question: How ought we respond to disagreement *if* we have good reason to believe that our interlocutor is our epistemic peer on some proposition? However, if we want to take the lessons learned in the literature and apply them to real-world disagreement situations, we must *first* evaluate our actual interlocutors for epistemic peerhood (on some proposition).

*Recall that the purpose of this evaluation is to detect any disagreement-relevant epistemic asymmetries.* I call this process a *Relative Epistemic Status Evaluation* (RESE). For my purposes here, when I say that we perform a RESE, that is a shorthand way of describing the process of evaluating oneself and others for the requisite conditions for peerhood described in Chapter 3. Here is my brief definition of a RESE:

- **Relative:** A peer evaluation is relative to the disputants and what is being disputed. More fully, peer evaluations are both relative to the proposition under dispute *and* relative to the evidence and virtue conditions of the disputants. So, when one is evaluating someone else as a potential peer, the targets of evaluation are (1) the interlocutors' doxastic attitudes towards the proposition that is being disagreed upon and (2) the relevant evidence and epistemic virtues possessed by the interlocutors that bear on that proposition.
- **Epistemic:** The evaluation focuses on and is contained by the relevant general epistemic factors that pertain to the rationality and doxastic

attitudes of the interlocutors. In other words, the evaluation is only concerned with how well each party forms justified beliefs, responds to defeaters for those beliefs, and how well each party handles the evidence they've been given to form those beliefs. Hence, the evidence and virtue conditions are crucial for the epistemic part of the evaluation.

- **Status (or Standing):** There are four coarsely defined statuses that one can assign to an interlocutor (including oneself) as the target of this evaluation: **inferior**, **peer**, **superior**, or **unknown**.<sup>37</sup> Status is determined by searching for parity of, at least, the relevant elements in the evidence and virtue conditions that pertain to what is being disagreed upon.
- **Evaluation:** The process of making judgments and assigning values (qualitative or quantitative) to a target of evaluation. In this case, the values assigned during the evaluation determine the Status (see above) of the interlocutors involved.

I think much more could be said about what a good (or at least sufficient) epistemic peer evaluation looks like. However, my definition of a RESE is at least a humble start—it will serve its function in this project. Furthermore, I find it odd that there isn't a detailed guide on *how* to evaluate others as peers in the literature. Again, for the advice offered in the literature to be applied to real-world cases, it is important to understand how to

---

<sup>37</sup> I am confident that I dispelled the idea that we might have near-peers or that peerhood may come in degrees in Chapter 3. Hence, these four epistemic statuses are the only ones that make sense *in the framework of the literature*. This is not to say that there aren't other perhaps more complex statuses that we could assign to ourselves and others, but it remains to be seen whether those additional statuses would fit well in the peer disagreement framework.

identify our peers in the first place. So, a RESE is a necessary yet underdiscussed part of understanding and applying the peer disagreement literature's advice to real-world cases.

One important aspect of the RESE as I've defined it here is the explicit addition of an "unknown" category. The unknown category can be (at least) one of two things: (1) We have no idea if our interlocutor is a peer, superior, or inferior, or (2) we are at least somewhat confident that our interlocutor is not obviously our epistemic inferior or superior. Interpretation (1) is interesting because in cases where we have no idea if our interlocutor is an epistemic peer, superior, or inferior— and we disagree with them—the question of the significance of disagreement (*sans* peers) is laid bare. That is, we are left to consider the significance of disagreement, full stop. This version of unknown status—call it unknown-1—is what we might encounter if we disagree with strangers at a party, neighbors at a town hall meeting, or commenters on social media. So, while unknown-1 cases of disagreement are not substantially discussed in the literature, that interpretation of the unknown status represents a significant and relevant set of people we often find ourselves disagreeing with. Interpretation (2) gravitates towards the literature a bit more than unknown-1 since this version of the unknown status positions us in an area somewhere between our epistemic inferiors and superiors. So, for example, interpretation (2)—call it unknown-2—describes situations we might encounter if we disagree with someone at a professional conference or with colleagues that we have known only briefly. Interpretation (2), then, represents another important set of people whom we often disagree with. However, it is unclear what it means to be confident that someone is not an epistemic inferior or superior while also being at least somewhat confident that



they might not be peers in a near-strict sense. What *epistemically relevant and useful* status could there be between inferior status and peer status, or between peer status and superior status? Frances and Matheson (2018) discuss how unknown-2 might be useful for the study of disagreement:

A related question is whether there is any important difference between cases where you are justified in believing your interlocutor is your peer and cases where you may be justified in believing that your interlocutor is not your peer but lack any reason to think that you, or your interlocutor, are in the better epistemic position. Peerhood is rare, if not entirely a fictional idealization, yet in many real-world cases of disagreement we are not justified in making a judgment regarding which party is better positioned to answer the question at hand... An analogy may help. It is quite rare for two people to have the very same [body] weight. So for any two people it is quite unlikely that they are ‘weight peers’. That said, in many cases it may be entirely unclear which party weighs more than the other party, even if they agree that it is unreasonable to believe they weigh the exact same amount. Rational decisions about what to do where the weight of the party matters do not seem to differ in cases where there are ‘weight peers’ and cases where the parties simply lack a good reason to believe either party weighs more. (2018, section 5.5.4 “Irrelevance of Peer Disagreement”).

In this quote from Frances and Matheson (2018, section 5.5.4), the authors explicitly say: “Peerhood is rare, if not entirely a fictional idealization...”, which indicates a lack of faith that peers exist at all. If this is the case, then of what use is the concept of peerhood

in applied situations? Frances and Matheson promote the idea that, in real-world applications of the disagreement literature, it could be the case that we ought to treat disagreements where we aren't justified in thinking that we are epistemically advantaged or disadvantaged as though our interlocutors are or at least could be epistemic peers. According to Frances and Matheson, this is true even in "Cases where you may be justified in believing that your interlocutor is not your peer" (2018, section 5.5.4). However, dropping the requirement for at least some semblance of peerhood while still applying the lessons from the literature is problematic for two reasons: (A) recall that the looser the definition of peerhood, the more opportunities there are for disagreement-relevant asymmetries to arise that explain why two parties disagree, which defuses the puzzle of peer disagreement, and (B) using the advice given in the literature—even without the focus on peerhood—still requires us to evaluate others for inferior, superior, or unknown status.

I've already discussed (A) in Chapter 3, and I expand on the problems with loose definitions below in this chapter. With regard to (B), it seems that unknown-2 is either an exceptionally loose definition of peerhood, which renders that status useless for applying the *peer* disagreement literature's prescriptions, or it isn't related to peer disagreement at all, which means it falls outside of the scope of this dissertation project. So, an open question remains regarding how we ought to respond to disagreements with those whom we believe are not our epistemic superiors or inferiors, but we also lack good reasons to assign them peer status. However, this open question is perhaps one that we ought to focus on in the disagreement literature since unknown-1 and unknown-2 statuses should

be assigned to many of the people whom we end up disagreeing with. Hence, despite the fact that many people we disagree with will fall into the broader unknown status category, evaluating how we ought to respond to unknown status cases is currently outside the scope of the *peer* disagreement literature, so it is also outside the scope of this dissertation.

## **6.2: The Relationship Between the Evidence and Virtue Conditions in RESEs**

The relationship between the evidence and virtue conditions is often not extensively discussed in the literature. If we use Kelly's definition of epistemic peers, both of these conditions need to be met for peerhood to obtain, or we must at least *believe* these conditions have been met. The evidence and virtue conditions are often analyzed as independent components of peerhood (see Chapter 3). However, depending on the situation, those conditions are difficult to evaluate independently. To see why, let's start with the evidence condition. It is worth restating that there are times when a single key piece of evidence makes or breaks an evaluation for disagreement-relevant epistemic asymmetries. According to the advice in the literature, when one person has a crucial piece of evidence and the other does not, the person possessing that crucial evidence is clearly an epistemic superior on that matter. This is why it is difficult to understand why Gelfert believes that we can be Distant or Remote Peers; it seems that either two people are Close Peers with the added evidence condition (see Chapter 3) or a sufficient amount of disagreement-relevant epistemic asymmetries are present and settle who ought to defer to who. So, we understand why we need the evidence condition for peerhood. But do we need the virtue condition?

Determining epistemic virtue parity among interlocutors is a challenge because epistemic virtues are difficult to precisely measure. In the context of evaluating the epistemic virtues of ourselves and others, we could use a general catch-all epistemic virtue such as being reasons-responsive. However, one could be adept at cultivating and exercising these general virtues and still fail to possess some disagreement-relevant virtue. This is because there are different types of reasoning: logical, spatial, sequential (cause-and-effect), and so on. Some of these types of reasoning have sub-types, which is why, for instance, philosophers teach their students about the differences between deductive, inductive, and abductive reasoning in logic courses; each is sufficiently different and deserves special training for its proper application. Additionally, spatial reasoning gives us our sense of direction and our capacity to untie knots, but it isn't appropriate to use when deciding if it is likely that all swans are white. Sequential reasoning helps us make predictions about the expected outcomes and consequences of, for instance, our dietary behaviors or our long-term financial investments, but it doesn't help us as much with deciding if our new couch will fit in our friend's mid-size SUV. In short, these types of reasoning are distinct: being competent with one doesn't always mean being competent with another. Yet, these different types of reasoning could all be used when making judgments about disputed propositions. In one case of disagreement, mathematical reasoning may be the most important virtue to evaluate. In another case, careful attention to semantic connections may be most important. To complicate matters further, we often encounter disagreements on propositions that we judge as true or false according to *several* different beliefs that we arrive at by using different skill sets. For

instance, perhaps being skilled at meticulous data analysis is key to justifying one belief, *b*, while following the logic of a complex argument is key to justifying another belief, *c*. If two people disagree about a proposition that involves both *b* and *c*, each person would need to evaluate their data analysis skills as well as their interlocutor's data analysis skills. Additionally, they would need to evaluate their own argument-following skills and their interlocutor's argument-following skills. This is all to say that evaluating epistemic virtues can be exceptionally complex, especially when we consider that our epistemic virtues are often employed when we *process our evidence*.

Evidence processing is discussed in parts of the literature, though usually in the context of how we handle evidence (i.e., evidence processing is often tied to the evidence condition rather than the virtue condition). However, it is important to understand that evidence processing is a cognitive ability, so this is broadly an epistemic virtue and not only part of the evidence exposure or access condition. What one does with their evidence can be just as important as what evidence one has collected. For example, if I read several papers on some topic and decide that those papers were only somewhat interesting, I probably won't spend much time carefully reviewing and reflecting on their details. If I found myself in a disagreement with someone where those papers are targets of a RESE, it would not be enough to simply tick the evidence condition boxes for those articles without considering how carefully I read them. That is, some of my relevant epistemic *virtues* in this case are tightly connected to determining what is in my set of evidence. Recall from Chapter 3 that several authors, notably Frances (2014), discussed the connection between evidence gathering and evidence processing. These authors tend

to agree that for two people to be epistemic peers, they should have both reviewed and considered the dispute-relevant evidence with *the same care and attention*. However, my point here is that satisfying the evidence condition for peerhood also requires meeting certain evidence-processing related epistemic virtue conditions as well. How we manage received evidence *through our virtues* changes the weight or import of that evidence in our own subjective evidence pools. Therefore, it could be the case that not all evidence is managed equally, even when exposure or access to that evidence is “equal”.

Some peer evaluations might focus more on evidence and less on virtues. For example, if you and I disagree about the final score from last night’s game, the most important piece of evidence is going to be our respective source of information for the score. That is, assuming we have similar basic memory and perceptual abilities, our RESEs will primarily search for parity in our evidence about the score. In this case, our epistemic virtues don’t seem significant. However, if it turns out that you didn’t really care about the outcome of that game and I did, that is an important virtue-based target of our RESE that will need to be included in the evaluation. Without considering our virtues in our RESEs, we may overlook a disagreement-relevant epistemic asymmetry—one that could easily explain why we disagree. So, our RESEs—even in cases where our evidential position seems more important—need to include an evaluation of virtues as well.

Conversely, our shared evidence may be less important than the parity of our virtues when performing a RESE. For example, let’s say you and I played a game where a third party shouted out random numbers and our task was to continuously add each

number to the previous number, summing all the numbers as they were shouted out. Let's further say that at some point in our game, we discover that we disagree about the current sum of those numbers. Our shared evidence (the numbers shouted out by the third party) will be important as evidential targets of a RESE, but not as important as the primary target of our RESE: our respective levels of skill at summing a random sequence of numbers. In this case, determining if our summing skills are "equal" is key to detecting disagreement-relevant epistemic asymmetries. So, again, our RESEs will need to target our virtues.

Thus, while there may be cases when it is more important that we've been exposed to the same evidence, most real-world RESEs will also require an accurate evaluation of the virtues that are in play as well. That is, in most cases, if our evidence is relevant to the disagreement, then our virtues are relevant to that disagreement as well. This will be important in later sections when I discuss how cognitive biases affect the accuracy of our RESEs. Before I begin that discussion, I offer a further explanation (carried over from Chapter 3) of why our definitions of peerhood cannot be too loose or too strict when looking for real-world peers.

### **6.3: Epistemic Peer Calibration Cases**

In this section, I roughly define a 'Goldilocks' zone for peerhood equality requirements that aligns with the literature's varying definitions of peerhood. To set the boundaries of the Goldilocks zone, I present three cases of disagreement: one that is too tight, one that is too loose, and one that is just right. My goal for finding the just right case of (idealized) peerhood is to use that case in my later discussion of how the CBP can

affect our RESEs. If it is plausible that the CBP affects our RESEs in a “just right” case of idealized peerhood, it should also be plausible that the CBP will affect our REESs in more complex real-world cases of disagreement as well. Let’s look at our first case below.

### **6.3.1: Peerhood Case 1: COMPUTERS**

Kelly (2010) asks readers to consider a case of disagreement where two people each have a thermometer and those thermometers display different temperature readings (p. 114). Assuming that there isn’t some obvious reason that one thermometer is displaying a higher temperature than the other, Kelly asks how those two people should respond to the “disagreement” between the thermometers (recall Christensen’s WATCHES case from Chapter 3 as a similar example). Since neither person has any reason to believe that their thermometer is more accurate or reliable than the other person’s, it seems that both parties ought to lower their confidence about which thermometer is correct. This is, of course, an example that intuitively supports conciliationism since both parties should agree that they don’t have good reason to believe one thermometer’s readings over the other’s. It would be odd, and perhaps irrational, for one person to trust their thermometer more without good reason. Furthermore, this is a case where even a steadfast should concede that, given the available shared evidence, neither party is better positioned to judge which thermometer they ought to believe. Therefore, they ought to suspend judgment about the actual temperature in this case.



I have a similar case using instrument outputs in mind, but I'm going to alter it to demonstrate that exceptionally tight or strict definitions of peerhood are not helpful for guiding our response to real disagreement. Here is the case:

#### COMPUTERS:

The world's leading computer scientists and engineers have designed and built—through painstakingly redundant quality checks—two all but numerically identical computers. They use the same hardware, firmware, software, power supplies, and so on. By “same” here, I mean that they each use custom designed and repeatedly debugged systems that are indistinguishable from one another upon close comparison (we're nearing Leibniz's Identity of Indiscernibles here). They also operate in carefully controlled and monitored environments with many failsafes built in. Both computers are designed and instructed to calculate Pi to the “last” digit using the same exact algorithms. These computers also check each other's new result after every digit is added to the irrational number as well as the monitoring reports generated by an external system. If the entire string of resulting digits doesn't match during these checks, an alarm goes off. After operating smoothly for several decades, the alarm goes off.

COMPUTERS is analogous to the thermometer case (and WATCHES) except for a few details. First, this case stipulates that each computer possesses the same evidence (initial inputs), the same “virtues” (hardware, software, and algorithms, etc., if computers can have virtues), and have extremely extensive and precise track records— they would surely meet Bundy's (2013) definition of an extraordinary track record. The

“interlocutors” in the case also seem to have nearly identical disagreement factors (recall from Frances’s list in Chapter 3: data, background “knowledge”, environment, etc.). So, when the alarm sounds, these computers are in a sort of hyper-idealized peer disagreement. But, in this case—one that paradigmatically represents the structure of an ideal peer disagreement where very tight definitions of peerhood are met—what *could* go wrong? I’ve all but stipulated that it is impossible for these computers to arrive at different results. The problem with this case is that it should be practically impossible. Furthermore, this case illuminates the fact that the puzzle of peer disagreement is practically incomprehensible when “peers” actually do have the exact same evidence and virtues. So, if we have very strict requirements for evidential and virtue parity for peerhood, it is not only unlikely to ever obtain, but it also amounts to a nonsensical result if those peers disagree. Clearly, *something* needs to be different about the computers in this case to arrive at the result where the alarm goes off, and whatever is different—no matter how small that difference is—it would count as a disagreement-relevant epistemic asymmetry because it would explain why these “interlocutors” disagree.

Of course, I realize that human brains aren’t computers and that the literature’s authors do not claim or imply that we are infallible. However, COMPUTERS serves to illustrate that, while stipulating that Uniqueness is true, peer disagreements must contain *enough* disagreement-relevant epistemic asymmetries among the interlocutors to explain why the disagreement occurs at all. Otherwise, the puzzle is just nonsense. *The epistemic significance of peer disagreement is just the epistemic significance of disagreements where disagreeing interlocutors with at least slightly unbalanced disagreement-relevant*

*epistemic elements (evidence, virtues, etc.) disagree.*<sup>38</sup> The significance of peer disagreement cases is primarily found in whether one can give good reason to think that the imbalance of epistemic elements is sufficient to reject the opinions of others (steadfasters) or insufficient to reject the opinions of others (conciliationists). I won't argue this point further here since my thesis does not depend on this claim being true, but it should be apparent that we cannot intelligibly use an overly tight version of epistemic peerhood when setting up our cases. This realization should be sufficient to change the framework of the disagreement literature in a way that excludes any notion of a strict peerhood requirement for real-world cases, but it is unclear how loose the requirements for peerhood can be before the problem fails to instantiate altogether. I will discuss that issue in the next case.

### **6.3.2: Peerhood Case 2: POLICY**

Here's a case that demonstrates what happens when the definition of peerhood is too loose:

**POLICY:**

A few months ago, two people connected on social media through friends of friends. They have had some casual discussions about a variety of topics and have gotten to know each other. They both have had a few college classes on economics and political science at different universities. Additionally, they both have found each other's points on the topics of economics and politics interesting

---

<sup>38</sup> Recall the quote from Kelly 2005 in Chapter 3 here where he makes a similar point in a parenthetical note in a footnote: "(Perhaps there is always at least some *slight* difference in intelligence, or thoughtfulness, or familiarity with a relevant argument.)" (Kelly, 2005, p. 175, fn 11)

and thoughtful, and they consider each other to be reliable sources of testimony regarding many of the topics they often discuss, including policy decisions. In response to an article posted about a new policy in a bill before the legislature, one person asserts that the policy cannot achieve its stated goals while the other person thinks the policy can achieve its stated goals. They disagree about the expected efficacy of the policy under consideration.

On the assumption that this is a genuine disagreement, is it a peer disagreement?

According to a loose definition of peerhood, it could be. The interlocutors seem to have a similar educational background relevant to the disputed matter. Also, they seem to appreciate each other's epistemic virtues to the extent that neither person believes the other is clearly an epistemic inferior or superior. Furthermore, it seems that they have the same evidence in that they are both responding to information contained in an article they both read about the new policy (assume they haven't been flooded with varying opinions about the new policy yet). Finally, they've developed a sort of track record for each other over a few months in their casual discussions about politics and many other subjects. They seem to be epistemic peers. Maybe they even believe, implicitly perhaps, that they are each other's peers on the proposition they disagree on—they might even be *acknowledged* peers.

When evaluating this case, we must determine if these two people meet the requirements for peerhood by searching for potential disagreement-relevant epistemic asymmetries. To defuse the peer disagreement puzzle, all we need to find is one plausible epistemic asymmetry to explain how it is that these two potential peers could come to

different conclusions with ostensibly the same evidence and virtues. I can reasonably imagine a few such asymmetries, starting with the (stipulated) fact that these two people had different professors since they went to different universities. Perhaps one person's economics professor taught that trickle-down economics is viable while the other's taught that trickle-down is empirically false. Or perhaps one professor taught a balanced approach to evaluating socialism vs. capitalism while the other presented strong arguments against only one of those systems. Relatively small differences such as these are both plausible in the case (and real-world disagreements) and, from our third person perspective of POLICY, those differences can explain why these two acquaintances disagree: they have sufficiently different backgrounds or evidence. So, the peer disagreement puzzle is no puzzle at all here.<sup>39</sup>

Another way to understand why a loose definition of peerhood is not sufficient to generate an interesting puzzle is that the steadfast response "wins" by default in cases like this. How so? Each person can rationally maintain their views (assuming they arrived at those views with *good* reasons and reasoning) precisely because they have different evidence or backgrounds. Recall that the main steadfast strategy to resolve the peer disagreement trilemma is to deny peerhood due to some sort of disagreement-relevant epistemic asymmetry (private evidence, self-trust, or self-knowledge, to name a few). In cases with loose definitions of peerhood where many disagreement-relevant epistemic

---

<sup>39</sup> Recall King's quote from Chapter 4 here: "If, for instance we loosen the requirements so that peerhood is consistent with small differences in evidence, there may be cases in which a small evidential difference between subjects makes a *large* difference in what it is rational for the subjects to believe. A single piece of evidence may in some cases be the *key* piece." (2012, p. 266)

asymmetries seem plausible, the steadfast doesn't even have to invent special evidence or unique virtues to justify their position on the significance (or lack of significance) of peer disagreement. So, if we are going to use the peer disagreement framework for understanding the significance of disagreement in real-world cases, we cannot use loose definitions of peerhood (even though cases like POLICY represent a major set of commonplace disagreements).

### **6.3.3: Peerhood Case 3: RESTAURANT (RETOLD)**

What, then, is the Goldilocks zone for peerhood parity requirements in idealized cases of disagreement if we want those cases to be plausible for real-world comparison? I contend that Christensen's original RESTAURANT case has most of the features that a case of idealized peer disagreement needs if it is going to approach resembling actual disagreement situations. So, we've circled back around to one of the most iconic cases from the literature. Here is the case from Christensen in a condensed 2011 version:

After a nice restaurant meal, my friend and I decide to tip 20% and split the check, rounding up to the nearest dollar. As we have done many times, we do the math in our heads. We have long and equally good track records at this (in the cases where we've disagreed, checking with a calculator has shown us right equally frequently); and I have no reason (such as those involving alertness or tiredness or differential consumption of coffee or wine) for suspecting one of us to be especially good, or bad, at the current reasoning task. I come up with \$43; but then my friend announces that she got \$45. (2011, p.2)

The features I have in mind that make this the basis for a good paradigmatic case of epistemic peerhood are:

- Determining the targets for evaluating the evidence condition is easy since the shared body of evidence is limited to the total of the dinner check— a number
- Determining the targets for evaluating the virtue condition is about as easy as a real-world case can be since visual attention and basic arithmetic skills are the only salient and relevant virtues that each party needs to perform the task at hand
- The actual answer to the question of the divided check is alethic— there is an objective answer to the question that both parties could easily discover

These points are a good start, but I'm going to slightly modify the case to clear up any potential confusion. Here is my modified case:

RESTAURANT RETOLD:

After a nice restaurant meal, Anna and Bob decide to tip 20% and split the check, rounding up to the nearest dollar. The check arrives with a clearly printed total at the bottom, as usual, and both Anna and Bob can clearly see and read the total of the check. As they have done many times, they use that total to do the math for the divided check in their heads. Anna and Bob have long and equally good track records at this (in the cases where they've disagreed, checking with a calculator has shown them right equally frequently); and Anna is not aware of any reason (such as those involving alertness or tiredness or differential consumption of coffee or wine) for suspecting that Bob is especially good, or bad, at the current reasoning task. Additionally, Bob is not aware of any reason for suspecting that

Anna is especially good, or bad, at the reasoning task. Anna arrives at \$43 as the split total. However, Bob disagrees and states that he believes the total is not \$43. In RESTAURANT RETOLD, I've made some adjustments to Christensen's original RESTAURANT case. First, I've given us the "God's eye view" of the case so it is clear that Anna and Bob mutually believe each other to be peers. Viewing the case from this position will be helpful when I discuss how the Cognitive Bias Problem can affect Anna and Bob differently. Second, I made sure that it is clear that Anna and Bob have a single shared piece of evidence before them and that neither is in a better position to evaluate that evidence (this description was present in Christensen's original 2007 presentation of the case as noted in Chapter 2). Finally, I changed the disagreement to a  $p$  / not- $p$  format ("The total divided by two is \$43" / "The total divided by two is not \$43") so that I can refer back to the disagreed upon proposition with one atomic variable.

Recall that the crucial targets of a RESE are the evidence and virtue conditions. So, to be clear, here is the solitary disagreement-relevant piece of evidence:

- The numerical total printed on the dinner check

and here are the disagreement-relevant epistemic virtues in play:

- Mathematical reasoning abilities (mental arithmetic)
- Attention to detail (reading the check carefully)
- Available and functional working memory capacity to perform the calculations (not drunk or tired or...)

and, finally, there is an established track record for this task present among the interlocutors. For now, I'm setting aside the track record issue while I examine what can



go wrong in a RESE for the virtue conditions alone. However, recall that the virtue condition should nearly always be considered in a RESE (see section 6.2 above). Also, we'll just assume that both Bob and Anna are equally matched in their attention to detail and working memory capacities. Below, I will focus on the two diners' respective evaluations of their own mathematical reasoning abilities as well as their evaluations of each other's mathematical reasoning abilities, which is the primary epistemic virtue here.

The next step is to walk through the RESEs that Bob and Anna must independently perform to count each other as peers. The RESE is relative to the following three targets of evaluation: (1) the proposition  $p$  ("The total divided by two is \$43"), (2) Anna's evidence and epistemic virtues relevant to  $p$ , and (3) Bob's evidence and epistemic virtues relevant to  $p$ . So, importantly, Bob needs to check more than what is explicitly presented in the case. That is, he not only needs to check to see if he is "aware of any reason for suspecting that Anna is especially good, or bad, at the current reasoning task" to meet the virtue condition, he also needs to evaluate whether he has any reason to suspect that *he* is especially good, or bad, at the current reasoning task to meet the virtue condition.

For Bob's RESE in this case, then, Bob needs to ask himself what his confidence is that his math reasoning skills and Anna's math reasoning skills are sufficiently equivalent to satisfy the requirements for peerhood. To do this, Bob must separately evaluate his reasoning skills and Anna's reasoning skills and then, perhaps through some RESE-informed scoring system, Bob must compare his appraisal of his reasoning skills to his appraisal of Anna's reasoning skills. For the two to be epistemic peers (from Bob's

point of view, at least), these two evaluations of mathematical reasoning will determine if Bob can reasonably believe that he and Anna meet the virtue condition of peerhood. In the case as it is presented, Bob is not *aware* of any reason to think that Anna is better or worse than he is at performing the calculations required. From this we can infer that Bob does not believe that there is a virtue condition-based disagreement-relevant epistemic asymmetry between he and Anna. Therefore, at least from Bob's point of view, he and Anna are peers on the primary virtue condition.

But I added the word "aware" to the case for a reason. Consider the following morbid example: Bob could, in fact, be having a stroke when he comes to believe that he has no good reason to believe that Anna is any better or worse than he is at the task at hand. However, as is usually the case, Bob is not immediately aware that he is having a stroke, so he isn't aware that the stroke is affecting his judgment. It could be the case that the stroke is damaging his brain in such a way that he forgets that Anna is, in reality, his two-year old niece who has limited math reasoning abilities. He might not be *consciously aware* of any good reason to think that Anna is no better or worse than he is at mental arithmetic. This merely illustrates the point that we can be unaware of defeaters for our beliefs, so I take it that the awareness rider I added here isn't unusual or highly contentious in this case. So, let's see how the things we are *unaware* of might affect the accuracy of our RESEs.

#### **6.4: Subjective Confidence Estimation Biases and Epistemic Peer Evaluations**

Bob isn't aware that he has any reason to think that Anna is any better or worse than he is at performing mental arithmetic. However, there may be good reasons of which

Bob is *unaware* that will skew his evaluation of his own mathematical reasoning abilities. One such reason comes from what we learned about overconfidence and underconfidence biases in the previous chapter. Let's say in this case that Bob, during his RESE, is rating his mathematical reasoning on a scale of 1 to 10 with 10 being the best and 1 being the worst.<sup>40</sup> Let's further say that he rates his basic mental arithmetic skills as an 8 of 10, and that his confidence in that rating is 90% (on a half-scale). Let's say that he also rates Anna's basic mental arithmetic skills as an 8 of 10 and that his confidence in that rating is 85%.<sup>41</sup> Since, from Bob's point of view, he and Anna both scored an 8 of 10 on the key epistemic virtue targeted in his RESE, he believes that Anna is, thus far, his epistemic peer on the amount owed. However, this is where the importance of the studies on overconfidence and underconfidence bias become salient and significant for epistemic peer evaluations.

If he was aware of the Cognitive Bias Problem (CBP), Bob should suspect that he might have some sort of bias in his evaluation of his and Anna's respective arithmetic skills. Furthermore, Bob should also believe that he cannot detect when or to what extent that bias will affect his judgment. And, finally, he should believe that he is not immune to

---

<sup>40</sup> For purposes of exposition, I'm sketching a general epistemic scoring system here. However, a detailed scoring system would involve a deep discussion of what, exactly, is being modeled. The question of RESE precision is tightly related to the problems with strict vs loose definitions of peerhood; imprecise evaluations will lead to the same problems with loose definitions of peerhood while hyper precise evaluations will, like strict definitions of peerhood, be too burdensome for real-world interlocutors. My concerns here are aimed more at accuracy than precision since, regardless of the appropriate degree of precision we decide upon for RESEs, inaccurate RESEs cause us to get the wrong result in peer evaluations. If we get the wrong result in peer evaluations, we won't—other than by mere epistemic luck—get the right result when responding to “peer” disagreements.

<sup>41</sup> It would make sense that one's confidence in their own rating would be slightly higher than their confidence in their appraisal of another person's skills. This is reflected in Foley (2001) and in Fumerton (2010) (see Chapter 2 for my discussion on Fumerton's view).

cognitive biases, even if he believes that he is less prone to biases than other people (don't forget that bias blind spot). In short, if Bob read Chapter 5, he shouldn't completely trust his 90% confidence estimate that his arithmetic skills for this task are an 8 of 10. However, we'll assume that Bob hasn't read Chapter 5, and even if he had, he would understand that just knowing about biases doesn't mitigate their effects on our judgment-making processes.

Since Bob missed out on Chapter 5, he remains confident in his rating of his and Anna's arithmetic skills. Could I disprove Bob's claim regarding his confidence level simply by pointing to Chapter 5? No. But Bob's lack of exposure to research on cognitive biases isn't the real problem here; the problem is that for Bob to perform an accurate RESE, he needs some way to be sure that he is not over- or underconfident in his arithmetic abilities. The CBP tells us that, because Bob is a human, he does not have a way to be sure that he is not over- or underconfident in his mathematical abilities.<sup>42</sup> Here's a plausible example. It could be the case that Bob is slightly overconfident in his arithmetic abilities, and, perhaps objectively (say, according to some standardized testing), Bob is actually a 7 of 10 on whatever scale he was using to compare his and Anna's epistemic virtues in terms of arithmetic abilities. That is, his 90% confidence that he is an 8 is just plain wrong, as is the case for many subjects when asked about their confidence levels in their responses to test questions in studies on overconfidence bias. If

---

<sup>42</sup> Of course, there *is* a way to determine if one is over- or underconfident in their judgments— the studies on confidence biases depend on various methods to do just that. However, to be sure that one is not over- or underconfident about X, one must know: (1) the actual answer to X (to compare to their answer) and (2) their confidence level (half or full scale) in that answer. If this were the case in RESTAURANT RETOLD, there would be no disagreement since Bob would *know* the answer to X. If knowledge is what peers are comparing, we run into the problem discussed by Hawthorn and Srinivasan from their 2013 work that I discussed very briefly in Chapter 3.

Bob is actually a 7 of 10 rather than an 8 and he is unaware that his overconfidence in his self-evaluation is preventing him from having an accurate confidence estimate, he and Anna might not be epistemic peers after all. This version of the case is plausible given the CBP and what we have learned about overconfidence bias in particular. Note, too, that it is not important why Bob is overconfident—perhaps his overconfidence stems from an above-average bias (see Chapter 5) or perhaps all his life he has been told by authority figures that he is great at everything he does. The point here isn't that Bob is overconfident; the point is that he doesn't realize when he is overconfident or how overconfident he is. However, realizing when one is overconfident in their self-evaluations is exceptionally important for performing accurate RESEs. Furthermore, we can imagine a plausible addition to this case where Anna is affected by an *underconfidence* bias. Perhaps this is due to the hard-easy effect (see Chapter 5) and Anna believes the task of evaluating arithmetic ability is *deceptively* easy. Or perhaps Anna has been told all her life that girls aren't good at math. Regardless of the reasons she is underconfident, if she arrived at the same conclusions that Bob did—that they are both 8's—but was underconfident in her rating of her arithmetic abilities, then that underconfidence in her potential to be a 9 rather than an 8 would further separate the two as epistemic peers on *p*.

In this version of the case, we have two people who believe that they are respectively 8's on a scale of 1 to 10 when, objectively (our "God's eye view"), one is a 7 and the other is a 9 regarding the epistemic virtues relevant to the requisite evaluation. **If this is the case, then Bob and Anna are not epistemic peers since the gap in**

**epistemic virtues counts as a disagreement-relevant epistemic asymmetry.**

Furthermore, both Bob and Anna are unaware that this disagreement-relevant epistemic asymmetry exists because they cannot reasonably adjust for confidence-affecting biases in their RESEs. That is, their RESEs are inaccurate *and* they aren't aware that their RESEs are inaccurate.

If the preceding interpretation of the RESTAURANT RETOLD case was the only permutation where Anna and Bob fail to meet the requirements for peerhood, then my reader might accuse me of cherry picking the one and only failure mode for epistemic peerhood evaluations to support my thesis. However, when taking the ubiquity and opacity of the CBP into consideration and the various ways in which Bob and Anna could be over- or underconfident in their evaluations of their mathematical reasoning abilities, it is more likely that Bob and Anna will make some RESE-relevant mistake than not. This is true even if their mistakes don't tally exactly as I have described them so far. For instance, if Bob and Anna are *both*, say, overconfident in some RESE-relevant judgment, we would be mistaken to think that their biases "cancel out" in that case. While we can be sure that we are often prone to overconfidence bias, we don't have a quick and reliable way to measure the *magnitude* of our overconfidence in any given situation. So, even if both Anna's and Bob's biases are pointed in the same direction—in this case as overconfidence—that doesn't mean that they are *equally* overconfident. Hence, even in the case where they are both overconfident, there is little reason to believe that they are correctly judging each other as epistemic peers; that assumption disregards the magnitude problem in over- and underconfidence biases.

Additionally, we can imagine a plausible variation of RESTAURANT RETOLD where Bob and Anna—due to a mismatch in confidence biases—do *not* believe that they are peers when *they are* in fact peers, which is also problematic for transferring the advice offered in the peer disagreement literature to real-world disagreements. Bob and Anna will not respond in a rational way if they have misjudged each other as non-peers since epistemically asymmetrical cases are resolved differently than peer cases. The former case where they believe that they are peers when they are in fact not is a type I error (false positive) and the latter case where they are peers but do not believe that they are is a type II error (false negative). Both of these errors, which are likely given all the permutations that result in mismatches between estimated confidence and actual abilities, are highly problematic for performing accurate RESEs. If we cannot be confident that we are performing accurate RESEs, then we cannot properly use the methods prescribed in the literature to resolve disagreements.

I should pause here to discuss my own methodology since it might appear to the reader that I am committing the same mistakes that I earlier criticized the current literature's authors of making. While I am using an idealized case with plausible yet fictional applications of the heuristics and biases program's findings on cognitive biases, I am not merely stipulating that Bob or Anna are over- or underconfident without good empirically grounded reasons to do so. Recall that my use of an idealized case of peer disagreement is deliberate; I am using a modified version of Christensen's RESTAURANT precisely because it falls in the Goldilocks zone of reasonable (i.e., could happen) idealized peer disagreement. However, even though I am starting with a

paradigmatic idealized case, I am not merely stipulating that our interlocutors are being affected by cognitive biases; we have robust evidence that tells us it is likely that any particular interlocutor will be affected by various cognitive biases when performing a RESE. If the interlocutors in the idealized cases of peer disagreement are meant to represent real people even remotely, then our idealized cases of peer disagreement ought to consider the potential for disputants to fall prey to RESE-affecting cognitive biases.

Indeed, caution about biases in general is built into the standard definition of epistemic peers. Recall that Kelly’s definition of epistemic virtues includes a “freedom from bias” condition. This condition indicates that Kelly and others had concerns about how biases affect our judgments about our own epistemic abilities and the abilities of others. Christensen’s (2007 parenthetical, expanded in 2014) often cited example about parents being biased in favor of their children tells us that common biases (as opposed to cognitive biases) are often sufficient to create a disagreement-relevant epistemic asymmetry.<sup>43</sup> My deployment of the CBP in RESTAURANT RETOLD is reasonable because, in general, we are less “free from bias” than we often believe we are (again, see: bias blind spot). So, meeting the freedom from bias requirement is not as simple as recognizing common biases and correcting them in our RESEs. By inserting specific examples of the CBP into the idealized cases, I’m practicing social epistemology that is

---

<sup>43</sup> By “common bias”, I mean the sort of well-recognized biases that we often attribute to others (but see Chapter 5’s review of bias blind spot). For example, we are often aware that we are biased towards our children, towards our favorite sports team, or when making judgments about “matters of taste.” The relevant difference between common biases and cognitive biases for my purposes here is that we are aware of many of our common biases; they are driven by conscious awareness of our preferences. Cognitive biases, on the other hand, do not always align with our preferences— in fact, they often motivate us to think and behave in ways that are contrary to our preferences (accidental unwarranted discrimination, irrational decision-making, behaviors that are harmful for our long-term well-being, etc.).



constrained by the tenets of bounded rationality, just as social epistemology ought to be practiced.

Perhaps my reader is not convinced of the seriousness of the damage that the CBP does to the peer disagreement project. Perhaps my example of one cognitive bias is an outlier; it could be the case that we only have to dodge that one type of bias to be “free from bias.” However, as I discussed in Chapter 5, there are a host of other cognitive biases that we should be concerned about, especially when attempting to perform accurate RESEs. In the next section, I provide another example using RESTAURANT RETOLD, but this time combined with the halo and horns effects.

### **6.5: Halo and Horns Effects and Epistemic Peer Evaluations**

Let’s consider another version of RESTAURANT RETOLD: As usual, Bob and Anna are out for dinner. However, on this particular night, Anna has plans to attend an art gala after dinner. She’s had her hair styled and she’s wearing a beautiful formal gown—she’s radiant. They finish dinner and the bill arrives. Bob and Anna calculate the total with tip, as usual, and they disagree about  $p$  as I described in RESTAURANT RETOLD. Bob performs his RESE as required to determine peerhood on  $p$ . He’s well aware that Anna appears exceptionally attractive this evening, but he is also aware that a person’s attractiveness has nothing to do with how well they perform mathematical calculations. During his RESE, he checks for all the required conditions: their evidence seems the same, their epistemic virtues seem the same, and he pays close attention to how much alcohol and coffee they’ve each had. As the result of his RESE (using a similar scoring

method he used in the case above), he rates Anna an 8 in mathematical reasoning.

Furthermore, he rates himself as an 8 in mathematical reasoning as well.

However, it is plausible that on this particular evening, his judgment is being affected by a halo effect due to Anna's pleasing appearance. It could be the case that Anna's appearance—through the lens of the halo effect—causes Bob to incorrectly consider the effects that Anna's last glass of wine had on her arithmetic abilities, etc. In other words, Bob may have accurately taken the effects that the wine had on Anna's epistemic virtues properly into account if Anna wasn't so surprisingly attractive that evening. Of course, if Bob became aware of an unwarranted upward nudge in his judgment of Anna's abilities, he would adjust his RESE to compensate for the halo effect. But remember that part of the CBP is that we aren't aware when our biases are affecting our judgments. So, Bob can't make the requisite adjustments to his RESE to counter the halo effect. Even if Bob read Chapter 5 and understood the risk that a halo effect might be skewing his judgment, he still couldn't be sure if the halo effect was altering his judgment about Anna's mathematical reasoning skills on that particular evening. As inconsequential as it may seem, the general appearance and perceived attractiveness of a person affects how we judge their abilities.<sup>44</sup> That is, despite our conscious aversion to

---

<sup>44</sup> Of special interest to my anticipated audience here, try Forgas' 2011 "She just doesn't look like a philosopher...?" and Talmas et al.'s 2016 "Blinded by Beauty: Attractiveness Bias and Accurate Perceptions of Academic Performance". The former study investigates how a person's emotional affect feeds into halo effects, but it also reveals that undergraduates use physical appearance to judge if a person is likely to be a philosopher or not. For me and my colleagues, it is important to understand that halo effects contribute to the problems with how students perceive our authority or competence (especially when considering this issue in the conversation about how gender and race affects student's perceptions of their professors). The Talmas et al. study shows in a sample from the general public (drawn from M-Turk), we judge the expected academic performance of others according to perceived attractiveness. If these results extend to judgments made by professional philosophers – and why would they not? – then it is likely that halo effects alter our evaluations of our students and our colleagues in ways that we'd rather they didn't.

prejudice and bigotry, there is little hope that we can completely evade these unconscious judgment-altering biases.<sup>45</sup>

Since I'm not using a definition of peerhood that is too strict (recall COMPUTERS) and since RESTAURANT RETOLD represents a clear, paradigmatic case of idealized peerhood (it also not too loose, as it was in POLICY), the degree of sameness or "equality" of Bob's and Anna's respective mathematical reasoning abilities need not be *exactly* equivalent. However, as was the case with confidence-affecting biases, it only takes a little imagination to construct variations of judgment-altering halo and horns effect biases that will skew Bob's and Anna's respective RESEs in such a way that there is a good chance that their RESEs will be inaccurate. So, even though I am not requiring exact equality of virtues here, it is important to remember that we only get the problem of *peer* disagreement when interlocutors are very nearly equivalent in evidential and virtue states.

Thus far, I've been focusing on how attractiveness and appearance might affect Bob's and Anna's judgments in one case. Here, my reader might be tempted, once again, to dismiss these potential instances of halo and horns effects as something that only affects relationships among strangers (but recall that those we would categorize as unknown statuses make up a significant portion of those whom we disagree with: see Chapter 6, section 6.1). After all, Bob and Anna are longtime friends and have played

---

<sup>45</sup> Presciently, I can sense that some readers will not find my application of the CBP through halo effects convincing in this version of RESTAURANT RETOLD. There is little more I can do to be more convincing that the halo effect *really does* affect us in the ways I've described in the case. The empirical data on halo and horns effects is robust and replicated (even cross-culturally, see Batres & Shiramizu, 2022).

their math game many times. So, one might reasonably think it could be that Anna and Bob have known each other long enough that these “first impression” sorts of halo and horns effects become muted over time. Along that line of reasoning, we might think that once we spend time getting to know people better, we become more familiar with their *actual* competencies and flaws. That is, perhaps track records and familiarity with our interlocutors provide us with some way to salvage epistemic peer evaluations from my attacks. This is not the case.

First, recall from the halo and horns effect section in Chapter 5 that a longitudinal study on the halo and horns effects showed that familiarity breeds contempt (see Jacobs and Kozlowski, 1985). That is, of the scant research-grade longitudinal evidence that we have, halo and horns effects *do not* seem to diminish significantly over time (six weeks, at least). Second, if a single study is not entirely convincing (as it shouldn't be), consider the etiology of the old adage about first impressions being extremely important; anecdotally, halo and horns effects seem to persist over subsequent meetings and future judgments. Going through a job interview with your fly down or a milk mustache is going to take a while to overcome. Showing up to a first date with lipstick on your teeth is going to raise a red flag that can take time to lower. These faux pas' matter to us for a reason—we are intuitively aware of the consequences of bad first impressions since they often affect how we are perceived long after we've committed them.

In this section (6.5) and the previous section (6.4), I've discussed two cognitive bias types: over- and underconfidence and halo and horns effects. I've demonstrated how CBP is a problem for performing accurate epistemic peer evaluations even in

paradigmatic cases of idealized peerhood. While my reader may not be convinced by my potential examples of the ways one bias can affect the accuracy of our RESEs, it is important to remember that these biases can stack; being affected by one doesn't mean that no others are in play. That is, different types of biases can simultaneously affect our judgments about the epistemic abilities of ourselves relative to others. Furthermore, these two bias types have a massive family of siblings and cousins in the HBP literature, many of which could cause us to miss or merely imagine disagreement-relevant epistemic asymmetries. However, I have not discussed the possibility that long track records among peers might mitigate the various effects that biases have on the accuracy of our RESEs. In the next section, I respond to that possibility.

#### **6.6: Confirmation Bias in Epistemic Peer Evaluations with Track Records**

A key feature of RESTAURANT RETOLD is that Anna and Bob have long and equally good track records at splitting the check in their heads. Here, we're imagining something like what Bundy would call an extraordinary track record. The case also tells us that Bob and Anna are disagreeing about an alethic matter and that when they've disagreed in the past, they've checked to see who is correct with a calculator. So, Bob and Anna have an *ideal* track record with regard to their after-dinner math game. Recall my anecdote in Chapter 3 about the *Great British Baking Show* as an illustration of how a real-world alethic track record might develop and how rare these extraordinary track records are. But since I am deliberately using a paradigmatic case of idealized peerhood here, we'll assume that Bob's and Anna's respective track records for their math game are in fact extraordinary.

Let's observe our subjects' verifications of their track records in slow motion. We'll start with Bob at the point where he is performing his RESE to check for asymmetries after discovering Anna disagrees with him about  $p$ . In his mental inventory, Bob pulls up his version of the track record for he and Anna's math game. He has a rough idea, from memory, of how many times he's been right (and wrong) and how many times Anna has been right (and wrong). His estimation of how many times he's been correct compared to how many times Anna has been correct seems roughly equal. So, if they've played this game 100 times and he's been correct 87 times while Anna has been correct 89 times, this track record seems sufficient to support the following: (1) individually, Bob and Anna are pretty good at this game—they are each correct nearly 90% of the time, and (2) their track records are similar when compared to each other's. So, there is a decent history of "trials" in Bob's track record, and his record indicates that they are both fairly reliable at arriving at correct answers in the math game. At least, this is Bob's recollection of the track record. However, let's not forget the studies on confirmation bias from Chapter 5.

Recall that confirmation bias is highly prevalent and that studies on that bias are robust. It seems plausible, if not highly likely, that Bob will be affected by confirmation bias. Let's think about what Bob is doing when he is recollecting the track record for the math game. He's searching for evidence about the times he was correct and the times that Anna was correct. However, confirmation bias causes us to "count the hits and miss the misses" when we search for evidence. Unless Bob keeps a written record of the hits and misses, it is likely that, if he believes that he is good at the math game (see

overconfidence bias), he'll recall more of his "wins" than he will his "losses" when compared to an objective scorecard. Also, recall Ross and Sicoly's (1979) study on the availability heuristic that found that domestic partners each claim to contribute to more than 60% of the housework. Another plausible way to interpret the results of that study is to say that we are more likely to recall our "successes" than our "failures" when we compare our accomplishments to others. So, if Bob is trying to think of the occasions that he was correct, he'll find plenty of instances when he was. However, he'll also forget to include some of the occasions when he was wrong (confirmation bias), and he'll more easily recall his successes than he will Anna's successes (availability heuristic). In short, Bob's track record is probably wrong even if he believes it is correct. The same could be said for Anna in this case as well. Though, without a well-documented objective record available, we don't have a reliable way to determine if the inaccuracies in their RESEs cancel each other out. Again, even though we have good evidence that real-world interlocutors will be influenced by cognitive biases, that doesn't mean that we can determine the magnitude of the effects of those biases in any particular instance. Even long track records do not indemnify us for the damage that cognitive biases can do to the accuracy of our RESEs. In real-world cases, affirming that we have similar and similarly good track records would require having objective records of those track records literally *recorded* somewhere other than in the memories of the interlocutors.

## **6.7: Conclusions of Applying CBP and Specific Cognitive Biases to Epistemic Peer Evaluations**

In the three sections above (6.4 - 6.6), I took a paradigmatic case of idealized peerhood and sprinkled in some cognitive biases. These biases affected the RESEs of the disputants in such a way that it seems unlikely that their respective RESEs will be accurate enough to correctly determine their relative epistemic statuses. I chose a small number of cognitive biases as examples to support my thesis. However, those biases are the proverbial tip of the iceberg: many unconscious cognitive errors can skew the judgments we must make to accurately identify others as our epistemic peers. There are myriad additional biases that could skew our RESEs, and each could inconspicuously alter our judgments about the epistemic virtues of our interlocutors. If we also consider the virtual library of various perceptual and memory-affecting illusions—in addition to the cognitive illusions I’ve discussed above—it is doubtful that we can reliably perform accurate RESEs in a way that satisfies the peer disagreement framework’s demands for peerhood.

Recall from Chapter 4 that King thought it was unlikely that we will ever actually be able to meet the (dialectical) evidence and (dispositional) virtue conditions for peerhood. He drew support for that claim from the reasonable assumption that no two people will meet the parity requirements for the evidence condition or virtue condition on a reasonable definition (not too strict, not too loose) of peerhood. While I agree with King’s proposal, my thesis supports a further claim: Even if we believe that we have met the evidence and virtue parity requirements for peerhood, it is still unlikely that we can



reliably perform accurate RESEs in such a way that we can detect pertinent disagreement-relevant epistemic asymmetries. The opacity, ubiquity, and prevalence of cognitive biases make it likely that we are unaware of how inaccurate our RESEs are. So, even if King is wrong (which I doubt), the peer disagreement peer evaluation framework still suffers from not attending to the descriptive limitations of real-world interlocutors.

If we permit treating our idealized interlocutors, Bob and Anna, as though they were real-world people, then the lessons of bounded rationality apply in RESTAURANT RETOLD. Here's how: if the CBP and the specific biases I discussed above cause us to question whether it is likely that Bob and Anna can reliably perform accurate RESEs, messier real-world cases of disagreement over our cherished beliefs will fare much worse. The complex topics that we often disagree about typically involve much larger sets of evidence and the simultaneous recruitment of several epistemic virtues. Since how we process evidence (for instance, the total of a dinner tab) involves using our epistemic virtues (for instance, mathematical reasoning), and since our evaluations of our epistemic virtues and others' virtues are likely to be skewed by biases, our attempts to use the methods prescribed in the peer disagreement literature will likely be surreptitiously thwarted by those biases. Thus, we can conclude that:

**(P5) Cognitive biases can affect our real-world epistemic peerhood evaluations in disagreements about our cherished beliefs to such an extent that we should not be confident that we can perform those evaluations in a way that satisfies the requirements for peerhood stated in the peer disagreement literature.**

This is the final premise for my main argument. I present that argument below, first as a sketch, then as a more precise valid argument.

## **6.8: Main Argument**

Over the course of this dissertation, I have been developing an argument. At the end of each chapter, I offered a premise that summarized what the chapter had accomplished. Importantly, I tried self-consciously to argue for highly plausible claims—not overextending myself at any point. I believe each premise has a great deal of support. Below, I combine the premises to sketch an argument that shows how my project challenges key presumptions from the peer disagreement debate.

### **6.8.1: Sketch of Main Argument**

(P1) The normative advice offered in the peer disagreement literature is intended to be applied to real-world disagreements about our cherished beliefs. (Chapter 2)

(P2) Accurately and competently performing real-world epistemic peer evaluations requires the following: (1) using a definition of epistemic peerhood that appropriately detects disagreement-relevant epistemic asymmetries among potential peers and (2) using a definition of epistemic peerhood that is not overly strict in order to allow real world interlocutors to meet the epistemic symmetry requirements for epistemic peerhood while also not being so loose that it fails to capture pertinent disagreement-relevant epistemic asymmetries. (Chapter 3)

(P3) Criticisms of epistemic peerhood generally rely on doubts that the evidential parity or epistemic virtue parity requirements can be satisfied in real-world situations. (Chapter 4)

(P4) Cognitive biases affect our judgments in ways that we are not aware of when we evaluate our own epistemic abilities, others' epistemic abilities, the evidence we use to justify our beliefs, and whether we are more or less biased than others.

(Chapter 5)

(P5) Cognitive biases can affect our real-world epistemic peerhood evaluations in disagreements about our cherished beliefs to such an extent that we should not be confident that we can perform those evaluations in a way that satisfies the requirements for peerhood stated in the peer disagreement literature. (Chapter 6)

Therefore,

(C) The current requirements for peerhood evaluations built into the peer disagreement project render the project's prescriptions inapplicable to most real-world cases of disagreement.

These premises capture the main points of each chapter using ordinary language. Each chapter supported or discussed why each of these premises is at least plausible. However, my work here is not complete. Since the above sketch is just that—a mere sketch of an argument—it only gestures at the plausibility of its conclusion. In the next section, to drive home my main thesis, I take this sketch and mold it into a proper valid argument.

### 6.8.2: Valid Main Argument with Comments

Using the premises from my argument sketch above and some additional commentary, I present my main argument in a valid form. Should my reader require further clarification of any premise, I've provided commentary for each below.

(V-P1) Either the advice offered in the peer disagreement literature is suitable for real-world application in disagreements about our cherished beliefs or the peer disagreement literature is suitable for theoretical investigation of rationality in ideal cases of disagreement. (P1 / Chapter 2)

(V-P2) If the advice offered in the peer disagreement literature is suitable for real-world application in disagreements about our cherished beliefs, then it is plausible that real-world interlocutors can readily identify others as epistemic peers.

(Chapters 3, 4, and 6)

(V-P3) If it is plausible that real-world interlocutors can readily identify others as epistemic peers, then it is plausible that real-world interlocutors can detect disagreement-relevant epistemic asymmetries that are present during a disagreement. (P2 / Chapter 3)

(V-P4) If it is plausible that real-world interlocutors can detect disagreement-relevant epistemic asymmetries present during a disagreement, then it is plausible

that real-world interlocutors can reliably perform accurate relative epistemic status evaluations. (Chapters 3 and 6)

(V-P5) If it is likely that the Cognitive Bias Problem negatively influences the accuracy of our relative epistemic status evaluations of real-world interlocutors, then it is *not* the case that it is plausible that real-world interlocutors can reliably perform accurate relative epistemic status evaluations. (Chapters 5 and 6)

(V-P6) It is likely that the Cognitive Bias Problem negatively influences the accuracy of our relative epistemic status evaluations of real-world interlocutors.

(P4 and P5 / Chapters 5 and 6)

(V-P7) It is *not* the case that it is plausible that real-world interlocutors can perform accurate relative epistemic status evaluations. (MP: V-P5, V-P6) (also apparent in P5)

(V-P8) It is *not* the case that it is plausible that real-world interlocutors can detect disagreement-relevant epistemic asymmetries that are present during a disagreement. (MT: V-P7, V-P4)

(V-P9) It is *not* the case that it is plausible that real-world interlocutors can readily identify others as epistemic peers. (MT: V-P8, V-P3)

(V-P10) It is *not* the case that the advice offered in the peer disagreement literature is suitable for real-world application in disagreements about our cherished beliefs. (MT: V-P9, V-P2)

Therefore,

(V-C) The peer disagreement literature is suitable for theoretical investigation of rationality in ideal cases of disagreement. (DS: V-P10, V-P1)

**Comments on V-P1:** As it stands, this disjunction offers us a choice: either the literature is suitable (i.e., useful) for real-world application or it is suitable for theoretical investigation. One question we might ask: Is this a tautology? Am I assuming something that is trivially true (namely:  $p$  or not- $p$ )? The answer depends on whether we interpret suitability for theoretical investigation as not being compatible with real-world application. This doesn't seem obviously true in this case; it could be true that the literature's advice is suitable for application *and* the literature is suitable for theoretical investigation (this disjunction is, of course, inclusive).

Additionally, the first disjunct gives us an important clue about the proposed scope of my argument: it pertains to our "cherished beliefs". Recall that Feldman (and many authors that followed in his steps) were concerned about the epistemic significance of peer disagreements about law, science, politics, philosophy, and religion—these

complex topics are usually the domains in which we form our cherished beliefs. However, it remains an open question whether the advice offered in the literature is applicable to very simple cases of disagreement that don't concern our cherished beliefs. For instance, both Feldman's QUAD and Elga's HORSE RACE are idealized cases of disagreement where the evidence and virtues involved only rely on visual acuity and level of attention. That is, they are merely disagreements on perceptual matters (see Lougheed 2020 for a further discussion about this). I am not arguing that real-world cases of disagreement that resemble those cases will face the challenges I discussed in this chapter. However, I am pessimistic about accurately identifying peers *even in those cases*. In addition to the data we have on cognitive biases, we also have a considerable amount of data on perceptual illusions. We are often fooled by lighting, shape, shading, line angles, etc. So, even the seemingly simplest cases should take note of the descriptive limitations of our perceptual abilities. But I am, for the most part, ignoring these cases of mere perception-based-evidence disagreements because they do not represent disagreements about our cherished beliefs.

**Comments on V-P2:** This premise expresses the conditional claim that if the literature is suitable for real-world application, then we'll need to be convinced that identifying our real-world peers is plausible. The slogan form for this premise is: No peers means no *peer* disagreement. Again, this seems uncontroversial, but let's test the conditional (in the scope of "cherished beliefs"). Could it be true in the actual world that the literature's advice is suitable for real-world application, but we *don't* need to identify others as

peers? Here's one possible way that could happen: I disagree with you and then I briefly consult the peer disagreement literature. I come to understand that my options are to split the difference or remain stubborn. Instead of carefully reading about what is actually required for peerhood, I just assume that you are my peer and decide, on a whim, to split the difference. The advice was "suitable" for real-world application in the barest sense: I used it in a real-world case. Furthermore, I didn't identify you as a peer—I didn't even try. In light of this counterexample, V-P2 seems false. However, if we look closer at what I did, I am not being rational if I act in this way. In the peer disagreement framework, the one and only reason I should conciliate is because I am treating the opinions of my interlocutor as though they have equal weight to my own opinions. Furthermore, the one and only reason I give equal weight to the opinions of others is because I believe they are my epistemic peers. In general, conciliating in cases when we aren't sure if our interlocutor is our peer requires some further explanation beyond the framework of the *peer* disagreement literature (see my discussion of unknown status above in 6.1). So, it cannot be the case that we can *reasonably* apply the advice from the literature without also identifying others as peers. I take it as granted that *reasonable* application is implied in my claims. So, any reasonable attempt to apply the advice in the literature will require that interlocutors identify their peers and non-peers.

**Comments on V-P3:** This premise expresses the claim that a necessary condition for identifying peers is that the interlocutors attempt to detect disagreement-relevant epistemic asymmetries ("asymmetries" henceforth) among themselves. That is, if it



seems unlikely that we can find all of the asymmetries that matter in a particular disagreement, then it is unlikely that we can identify others as peers in that disagreement. So, V-P3 is definitional in nature. Let's test it anyway. Could it be true that it is plausible that we could identify our peers without any sense of the asymmetries that could be present? Maybe. Here's how: You and I disagree, and a "master peer-finder" appears and tells us that we are peers. Since this person is a master peer-finder, I ought to believe her. Technically, the interlocutors—you and I—did not detect the asymmetries, but we could identify each other as peers. In this scenario, the antecedent is true and the consequent is false.

Furthermore, this scenario isn't even too far-fetched. If I were mentoring two students who were writing dissertations on the very same topic, I might be a master peer-finder for those students relative to the topic of their dissertation. Let's say that these students didn't know of each other's work. One day, they meet each other in my office for the first time and a disagreement about their dissertation topic emerges. I tell them that, because I am familiar with their evidence and epistemic virtues that they are indeed peers on the topic. They didn't need to have long conversations with each other to discover their peerhood. In fact, they didn't even need to discuss anything pertaining to peerhood. This seems like an open and shut case against V-P3. However, notice that *someone* had to evaluate the interlocutors for peerhood. So, while the consequent of V-P3 has the interlocutors doing the work to find the appropriate asymmetries, it could easily become much wordier and cover all of the cases in which *someone* or even *something* (imagine AI checking for peerhood) must search for and detect the relevant asymmetries.

To identify others as peers just means being able to detect (or have someone else detect) possible relevant asymmetries. If there isn't an investigation for asymmetries of some sort, then there isn't a way to identify others as peers.

**Comments on V-P4:** Similar to V-P3, V-P4 tells us what is required to detect relevant asymmetries. If we want to find asymmetries in evidential positions or epistemic virtues among interlocutors, someone (or something, see comments on VP-3) will need to perform some sort of evaluation to locate those asymmetries. In this chapter, I called this evaluation a *Relative Epistemic Status Evaluation* (RESE). Notice that the consequent of V-P4 also tells us that it needs to be plausible that we can both *reliably* and *accurately* perform a RESE. So, either it is plausible that we can reliably perform an accurate RESE or it isn't the case that it is plausible that we can detect the relevant asymmetries. Once again, let's look for a reasonable counterexample.

Could it be the case that we can detect asymmetries present among interlocutors without accurately performing a RESE? Again, we could conjure up a third party to do the detective work for us, but this would just mean adding a lot of pedantic hedges to the consequent. Someone or something must somehow perform an evaluation of some sort—no matter how simple, brief, or remote—to check for asymmetries. This is because the nature of epistemic asymmetries is such that they are not directly perceptible; as necessarily internal mental states, they do not just appear to others. Whether by inference, empirical data, test scores, observation of behavior, testimony, or careful armchair

theorizing, there is necessarily an evaluation present somewhere in the process of determining the evidential and epistemic virtue positions of our interlocutors.

The further requirement that our RESEs need to be accurate is tied directly to the requirement that the asymmetries we need to detect are disagreement relevant. That is, we might think that we could get away with the following counterexample: It could be true that it is plausible that we can detect disagreement-relevant epistemic asymmetries, but false that we need to be reliable at *accurately* performing our RESEs to do so. After all, a haphazardly performed RESE might detect all sorts of *irrelevant* asymmetries. Let's say you and I disagree about whether the moon is made of cheese. You perform a RESE and discover that I have never read any Shakespeare. Since you have read the complete works of Shakespeare several times, you've detected a (non-disagreement-relevant) epistemic asymmetry between us. So, you detected an asymmetry, but your RESE was inaccurate relative to the relevant targets (i.e., the evidence and capacities required to know what the moon is made of). Here, it is helpful to reflect on what it means for something to be accurate: it hits what it is aiming at. If we disagree about the material properties of the moon, your RESE must be aimed at the evidence and virtues that pertain to judgments about the material properties of the moon. In short, to detect *disagreement-relevant* asymmetries, your evaluation must be aimed at what is relevant to the disagreement. If your evaluation is inaccurate, then it will fail to detect the *relevant* asymmetries. So, not only are RESEs required to detect disagreement-relevant epistemic asymmetries, they also must be accurately aimed at detecting the relevant asymmetries. If

it's not the case that our RESEs are accurate, then we won't be able to find the appropriate asymmetries that would block peerhood from obtaining.

**Comments on V-P5:** This conditional represents the hypothesis I set out at the beginning of this project. If cognitive biases make us bad judges about the epistemic qualities of ourselves and others, then we won't be able to accurately determine who our peers are. Allow me to work through this analysis carefully.

First, let me be clear what I mean when I say that it is *likely* the Cognitive Bias Problem negatively influences us. In Chapter 5, I provided a smattering of data from the heuristics and biases program (HBP) that suggests we are prone to judgment-skewing cognitive biases. As the researchers I quoted there stated, cognitive biases are prevalent. However, it could be reasonable to think that the mere prevalence of bias in a *population* does not indicate that any *specific* instances of bias are likely—we all like to think that we are special (but don't forget bias blind spot). To make the inductive move from general to specific, I'm using a statistical syllogism that has this general form:

(SS-P1) Most Y are affected by X.

(SS-P2) I am a Y.

Therefore,

(SS-C) It is likely that I am affected by X.

Given the findings from the HBP (and I only discussed a small part of its totality), I take myself to be warranted to reason in that way. I also think that anyone who finds

themselves in the role of Bob or Anna in RESTAURANT RETOLD should, *once they've been made aware of the Cognitive Bias Problem*, reason in this way as well.

Let's think about this a bit more. If Bob and Anna represent real-world interlocutors and all of the conditions for peerhood that are stipulated in the case hold, this also implies that their RESEs were accurate. As I said when I explained the modifications I made to Christensen's original restaurant case, Bob and Anna are not *aware* of any reason to think that either is better or worse at the (epistemic) tasks required. Some real-world interlocutors may be vaguely aware that we have cognitive biases, but they may not be aware of how prevalent biases are or how much each bias can skew our judgments about ourselves or others. I'll say that these interlocutors are in a *pre-reflective* state about the CBP. If Bob is in a pre-reflective state about the CBP, he may not understand that the CBP could serve as a defeater for his beliefs about Anna being his peer. We could ask whether Bob has justification for his belief that Anna is his peer in this state, but what would we gain in doing so? The CBP affects people regardless of whether they are aware of it. Additionally, the introspective opacity of the CBP limits our ability to determine when and how it is affecting us in any particular instance—this is part of the reason why it is such a problem in the first place. So, the question of justification for many of our beliefs about peerhood obtaining is tangential to whether we are actually equipped to perform accurate RESEs.

However, I do think that once we learn about the CBP and we find ourselves in a *reflective* state about how it could affect the accuracy of our RESE's, we *should* treat that information as a defeater for beliefs whether peerhood actually obtains. But this is a

further step in an argument I am not making here; that normative advice I offered is also tangential to my main thesis. That is, my argument is descriptive: Cognitive biases are prevalent, and they restrict how well we actually judge the epistemic qualities of ourselves and others. Determining who our peers are requires judging the epistemic qualities of ourselves and others. Therefore, how well we judge those epistemic qualities is based on *descriptive facts* about cognitive biases. V-P5 is, in a nutshell, the following: If biases skew our judgments about others, we won't judge them accurately.

**Comments on V-P6:** This premise is the moneymaker. Once it is added to the argument, logic takes over and does the rest of the work. So, since it is an atomic proposition, we just need to decide if it is true or false. The arguments I presented in Chapters 5 and 6 (so, P4 and P5, respectively) are sufficient to at least seriously consider the idea that the premise is true. The key to thinking V-P6 is true will depend on whether the studies I presented in Chapter 5 were convincingly used in the paradigmatic case of peerhood in Chapter 6. My reasoning here is as follows: We have robust evidence that our judgments about ourselves and others are often biased. So, it's plausible that Bob and Anna *could be* influenced by those biases—this should be uncontroversial. I stipulated that they *were* influenced by those biases only to make my larger point; in a sense, nature has “stipulated” that we are all prone to those biases. So, if we treat Bob and Anna as real people, they are at least prone to the biases I attributed to them.

I think the more interesting question regarding V-P6 is whether our biases affect our judgments *enough* to make them inaccurate in a relevant way. To put that another

way, we should wonder if our biases are sufficient to alter our judgments about ourselves and others in such a way that we will *usually* get peerhood evaluations wrong. To answer this, at the bare minimum, we'd need more research on how people actually evaluate others when they are instructed to perform peer evaluations, especially during disagreements about our cherished beliefs. It could turn out that they do quite well, though I would currently bet heavily against that being true. Even if it was the case that our biases only have a minor effect on our RESEs, the claim in V-P6 is at least worth probing further.

If all else fails, consider my back-up thesis: It would be prudent if the peer disagreement literature included detailed earnest discussions about the descriptive limitations of real-world interlocutors. Even if I only demonstrated that thesis here and philosophers took that advice to heart, the literature would at least begin to address the problems involved in establishing peerhood that many authors either completely ignored or only hinted at (with the exceptions of Gelfert 2011, King 2012, and Kenyon 2020; they stand out as reasonable voices to me).

However, I think my stronger main thesis—even if it turns out to be false—presents the kind of challenge worth responding to should anyone wish to engage with my work. Furthermore, I also happen to think that my thesis is true. That is, cognitive biases can affect our real-world epistemic peerhood evaluations in disagreements about our cherished beliefs to such an extent that we should not be confident that we can perform those evaluations in a way that satisfies the requirements for peerhood stated in the peer disagreement literature.

V-P7 through V-P10 are logical outputs of V-P2 through V-P6, so I will say nothing about them that I haven't already said above. This, of course, doesn't mean that they are uninteresting to discuss further. However, for my purposes here, they are merely cogs that lead to my conclusion.

**Comments on V-C:** If V-P1 is true, then my conclusion demonstrates that the peer disagreement literature has some value. While I only needed the first disjunct to be false to prove my point, I also happen to think that V-P1 *is* true; using the method of cases—despite its limitations for general real-world application—helps us understand what is at stake, and it helps define the solution space for the problems that appear in those cases. The peer disagreement literature does both of those things on some level. By looking at ideal cases of peer disagreements, we understand our options as interlocutors better. Should I conciliate or should I remain steadfast? Do I have good reasons to think that my opponent is my superior or inferior and how much should that matter here? What other responses to a disagreement are reasonable besides conciliating or remaining steadfast? These are the kinds of questions that help us understand why and when particular moves are rational or not.

Unfortunately, because most authors contributing to the literature were not interested in discussing the descriptive limitations of real-world interlocutors in detail, the advice they offered was not apt to guide us. The literature still has a long road from theory to application.



## **6.9: Summary and Conclusion to Chapter 6**

In the first part of this chapter, I explained what is required to evaluate ourselves and others as peers. First, I explained what a relative epistemic status evaluation (RESE) is, and I discussed the possible epistemic statuses we can assign to ourselves and others. Then I discussed the connection between the evidence condition and the virtue condition, arguing that they do not come apart easily. I continued by selecting a suitable paradigmatic case of epistemic peerhood from the literature as an ideal case for testing my thesis. Then, I proposed specific examples of how the CBP could affect our hypothetical interlocutors' peer evaluations. Finally, I ended the last section of the chapter with a sketch of my argument and a sharpened, valid version of that argument to demonstrate my thesis.

I have one more item worth mentioning before I close this chapter. The attentive reader will note that I didn't use P-3 from my argument sketch in my valid argument. That premise expressed the claim that most of the worries about real-world peerhood in the literature seemed to be focused on whether it is likely that any two people could satisfy the evidence parity or virtue parity requirements. In other words, it seems to some philosophers—especially King (2012)—that, even with a reasonably liberal definition of peers, there will always be some disagreement-relevant epistemic asymmetry present. Let me explain why my thesis is relevant to that claim. I think that even in cases when we have good evidence that our interlocutors have the “same” evidence and virtues as us, we should still be cautious about claiming that they are in fact our peers. We cannot dismiss the Cognitive Bias Problem by carefully scrutinizing each other's evidence and virtues—

in doing so, we might be influenced by bias. Nor can we ignore the effects of the Cognitive Bias Problem even when we have long track records with our interlocutors—track records don't immunize us to the effects of the CBP. This is an important consequence of demonstrating my thesis: if we have the same evidence and virtues as someone else, we will still need to overcome the Cognitive Bias Problem to correctly (according to the literature) identify them as epistemic peers. At present, we do not have any good methods to overcome the Cognitive Bias Problem, though there is ongoing research to discover those methods. So, we should turn our attention towards efforts to mitigate or overcome our biases to better understand how we ought to rationally respond to real-world disagreements.

## **CHAPTER 7: CONCLUSION TO THE PROJECT**

Recall that I began this dissertation with a common question: How ought you and I respond to our disagreement on a topic we care about deeply? When we checked the internet to answer that question, our search results led us to the peer disagreement literature, which seemed promising—at first. We assumed the advice offered in that literature would be useful for us to figure out how we ought to respond to our disagreement; it appeared suitable for real-world application. In fact, as I argued in this dissertation, the founding authors in the literature were indeed concerned about the epistemic significance of disagreements in real-world disputes that involved our cherished beliefs about law, science, politics, philosophy, and religion (Chapter 2). We were left with a question: Are we epistemic peers or not?

I dug into the definitions of peerhood and attempted to list the requirements we would have to satisfy to be peers (Chapter 3). During my investigation, I found a few authors who were somewhat skeptical that peerhood could ever be realized (end of Chapter 3), and I discovered that at least one author, Nathan King, argued extensively that peerhood is highly unlikely to be found in real-world scenarios (Chapter 4). While taking into consideration King's points, I wondered if our cognitive biases could affect the way we evaluate others as peers. I discussed Herbert Simon's theory of bounded rationality and briefly charted the development of the heuristics and biases program. Then I looked closely at a handful of specific cognitive biases that could negatively influence our epistemic peer evaluations (Chapter 5). Finally, I combined what we learned about the requirements for peerhood as stated in the literature with what we learned about cognitive biases to argue that it is likely that we are often bad judges of who our epistemic peers are. I even provided a valid argument that demonstrated that the literature is not suitable for real-world application in disagreements over our cherished beliefs, which thereby established my main thesis (Chapter 6).

So, where does that leave us after this "brief" pause in our quest to answer the question: How ought you and I respond to our disagreement? We could try to evaluate our relative epistemic positions. We could sit for hours and share our evidence that pertains to the topic we disagreed about. (I'm afraid I've already forgotten what that topic was!) We could ask each other probing questions and recall instances where we've observed each other's reasoning abilities to determine if our epistemic virtues—the ones that relate to our topic—were "equal." Finally, we could recall all the times that we've

disagreed about a similar topic and count up the times that we were right and wrong about those topics. We could do all of this and arrive at a rough estimate of our relative epistemic standings.

Unfortunately, even if we discovered that our evidence and virtues were close enough to being in parity to satisfy the necessary and sufficient conditions for peerhood, and we individually counted up all the times we were right and wrong in disagreements over similar topics to establish a track record, we would still need to worry about how the Cognitive Bias Problem affected our evaluations. In order to evaluate each other for peerhood, we would have to accurately detect disagreement-relevant epistemic asymmetries among our respective epistemic virtues and evidence. If our evaluation didn't take this step, we would be getting ahead of ourselves—we could explain why we disagreed by appealing to the possible differential in our virtues or evidence. It could be the case that once we discovered precisely why we disagreed, we'd realize that we weren't having a peer disagreement at all. But detecting asymmetries is not as easy as it sounds when we reflect on the Cognitive Bias Problem. For example, your relevant virtues might actually be superior to mine, even if I was convinced that they weren't because I was being affected by overconfidence bias. And that is just one example—the list of biases that could skew our judgments about our epistemic virtues extends to the horizon. Worse yet, there's no way for us to determine which biases are affecting us and when.

We could be brave and bite the bullet here: we could just shrug off the Cognitive Bias Problem and do the best we can with the evidence we have. Indeed, most people get

through their lives just fine *despite* their biases. However, when it comes to deciding if we ought to change our minds about our *cherished beliefs*, those biases can interfere with our attempts to use advice that requires us to first evaluate our relative epistemic standings.

In our investigation, we've learned this much: We are particularly ill-suited to precisely evaluate our epistemic abilities and the abilities of others. More importantly, if we use the advice offered in peer disagreement literature and mistakenly revise our beliefs because we think we are peers when we are not, we may have altered our (cherished!) beliefs without cause. Furthermore, if we are peers but our biases prevented us from accurately identifying each other as such, we might end up holding on to bad beliefs that should be revised or discarded. Ultimately, it would be dangerous for us to use the advice offered in the literature to resolve our disagreement; as it stands, the process of evaluating others as peers doesn't properly take into account the psychological limitations of real-world interlocutors. Until that problem is resolved, you and I will have to search for another method to determine how we ought to respond to our disagreement.

## BIBLIOGRAPHY

- Aumann, R. (1976). Agreeing to Disagree. *The Annals of Statistics*, 4(6), 1236–1239.
- Bacon, F. (1620). *Novum Organum*.
- Bahník, S., Englich, B., & Strack, F. (2017). Anchoring effect. In R. F. Pohl (Ed.), *Cognitive Illusions: Intriguing phenomena in thinking, judgment, and memory* (second edition) (pp. 223-241). Routledge.
- Ballantyne, N. (2015). Debunking biased thinkers (including ourselves). *Journal of the American Philosophical Association*, 1(1)141-162.
- Batres, C., & Shiramizu, V. (2022). Examining the “attractiveness halo effect” across cultures. *Current Psychology*, 1-5.
- Bishop, D. V. (2020). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 47th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, 73(1), 1-19.
- Bundy, J. (2013). *The epistemic significance of disagreement with others*. Dissertation, UC Santa Barbara.
- Chandrashekar, S. P., Yeung, S. K., Yau, K. C., Cheung, C. Y., Agarwal, T. K., Wong, C. Y. J., ... & Feldman, G. (2021). Agency and self-other asymmetries in perceived bias and shortcomings: Replications of the Bias Blind Spot and link to free will beliefs. *Judgment and Decision Making*, 16(6), 1392-1412.
- Chapman, G., & Johnson, E. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In Gilovich, T., Griffin, D., & Kahneman, D. (Eds.), *Heuristics and biases: The Psychology of Intuitive Judgment* (pp. 120-138). Cambridge University Press.
- Christensen, D. (2007). Epistemology of disagreement: The good news. *Philosophical Review*, 116(2), 187-217.
- Christensen, D. (2011). Disagreement, question-begging and epistemic self-criticism. *Philosopher's Imprint*, 11(6), 1-22.
- Christensen, D. (2014). Disagreement and public controversy. in Lackey, J. (ed.) *Essays in collective epistemology*. Oxford University Press.
- Christensen, D., & Lackey, J. (Eds.). (2013). *The Epistemology of Disagreement: New Essays*. Oxford University Press.

- Cohen, S. (2013). A defense of the (almost) equal weight view. In Christensen, D., & Lackey, J. (Eds.), *The epistemology of disagreement: New essays* (pp. 98-117). Oxford University Press.
- Comesana, J. (2012). Conciliation and peer-demotion in the epistemology of disagreement. *American Philosophical Quarterly*, 49(3), 237-251.
- Conee, E. (2010). Rational disagreement defended. In Feldman, R., & Warfield, T. (Eds.), *Disagreement* (pp. 69-90). Oxford University Press.
- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1), 20.
- Ehrlinger, J., Gilovich, T. & Ross, L. (2005). Peering into the bias blind spot: People's assessments of bias in themselves and others. *Personality and Social Psychology Bulletin*, 31, 680-92.
- Elga, A. (2007). Reflection and disagreement. *Nous*, 41(3), 478-502.
- Elgin, C. (2010). Persistent disagreement. In Feldman, R., & Warfield, T. (Eds.), *Disagreement* (pp. 53-68). Oxford University Press.
- Enoch, D. (2010). Not just a truthometer: Taking oneself seriously (but not too seriously) in cases of peer disagreement. *Mind*, 119(476), 953-997.
- Evans, J. (1989). *Bias in human reasoning*. Lawrence Earlbaum Associates.
- Feldman, R. (2006). Epistemological puzzles about disagreement. In Hetherington, S. (Ed.), *Epistemology Futures* (pp. 216-236). Oxford University Press.
- Feldman, R., & Warfield, T. (Eds.). (2010). *Disagreement*. Oxford University Press.
- Foley, R. (2001). *Intellectual trust in oneself and others*. Cambridge University Press.
- Forgas, J. P. (2011). She just doesn't look like a philosopher...? Affective influence on the halo effect in impression formation. *European Journal of Social Psychology*, 41, 812-817.
- Forgas, J.P., & Laham, S. M. (2017). Halo effects. In R. F. Pohl (Ed.), *Cognitive Illusions: Intriguing phenomena in thinking, judgment, and memory* (second edition) (pp. 276-290). Routledge.

- Frances, B. (2014). *Disagreement*. Polity Press.
- Frances, B., & Matheson, J. (2018). Disagreement. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/disagreement/>
- Fumerton, R. (2010). You can't trust a philosopher. In Feldman, R., & Warfield, T. (Eds.), *Disagreement* (pp. 91-110). Oxford University Press.
- Gelfert, A. (2011). Who is an epistemic peer? *Logos & Episteme*, 2(4), 507-514.
- Gigerenzer, G., Czerlinski, J., & Martignon, L. (2002). How good are fast and frugal heuristics? In Gilovich, T., Griffin, D., & Kahneman, D. (Eds.), *Heuristics and biases: The Psychology of Intuitive Judgment* (pp. 559-581). Cambridge University Press.
- Gilovich, T. & Griffin, D. (2002). Introduction. In Gilovich, T., Griffin, D., & Kahneman, D. (Eds.), *Heuristics and biases: The Psychology of Intuitive Judgment* (pp. 1-18). Cambridge University Press.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.), *Heuristics and biases: The Psychology of Intuitive Judgment*. Cambridge University Press.
- Gutting, G. (1982). *Religious belief and religious skepticism*. University of Notre Dame Press.
- Harari, H., & McDavid, J. W. (1973). Name stereotypes and teachers' expectations. *Journal of educational psychology*, 65(2), 222.
- Hawthorne, J. & Srinivasan, A. (2013). Disagreement without transparency: Some bleak thoughts. In Christensen, D. Lackey, J. (Eds.), *The epistemology of disagreement: New essays* (pp. 9-30). OUP.
- Hoffrage, U. (2017). Overconfidence. In R. F. Pohl (Ed.), *Cognitive Illusions: Intriguing phenomena in thinking, judgment, and memory* (second edition) (pp. 291-314). Routledge.
- Jacobs, R., & Kozlowski, S. W. (1985). A closer look at halo error in performance ratings. *Academy of Management Journal*, 28(1), 201-212.
- Kahana, M. J. (2012). *Foundations of human memory*. OUP.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.



- Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 363-391.
- Kahneman, D., Slovic, P. & Tversky, A. (Eds.). (1982). *Judgment under uncertainty*. Cambridge University Press.
- Kahneman, D., Knetsch, J., & Thaler, R. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1), 193–206.
- Kelly, T. (2005). The epistemic significance of disagreement. In Gendler, T., & Hawthorne, J. (Eds.), *Oxford Studies in Epistemology Volume 3* (pp. 167-196). Oxford University Press.
- Kelly, T. (2010). Peer-disagreement and higher-order evidence. In Feldman, R., & Warfield, T. (Eds.), *Disagreement* (pp. 111-174). Oxford University Press.
- Kenyon, T. (2014). False polarization: Debiasing as applied social epistemology. *Synthese*, 191, 2529-2547.
- Kenyon, T. (2020). Peer idealization and internal examples in the epistemology of disagreement. *Dialogue: Canadian Philosophical Review/Revue canadienne de philosophie*, 59(1), 69-79.
- King, N. (2012). Disagreement: What’s the problem? or A good peer is hard to find. *Philosophy and Phenomenological Research*, 85(2), 249-272.
- Kornblith, H. (2010). Belief in the face of controversy. In Feldman, R., & Warfield, T. (Eds.), *Disagreement* (pp. 29-52). Oxford University Press.
- Kusev, P., van Schaik, P., Ayton, P., Dent, J., & Chater, N. (2009). Exaggerated risk: Prospect theory and probability weighting in risky choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1487.
- Lackey, J. (2007). What should we do when we disagree? In Gendler, T., & Hawthorne, J. (Eds.), *Oxford Studies in Epistemology: Volume 3* (pp. 274-293). Oxford University Press.
- Lackey, J. (2010). A justificationist view of disagreement’s epistemic significance. In Haddock, A., Millar, A., & Pritchard, D. (Eds.), *Social Epistemology* (pp. 298-325). Oxford University Press.
- Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of experimental psychology: Human learning and memory*, 4(6), 551.

- Lougheed, K. (2020). *The epistemic benefits of disagreement*. Springer.
- Matheson, J. (2014). Disagreement and epistemic peers. *The Oxford Handbook of Topics in Philosophy* <https://academic.oup.com/edited-volume/42642/chapter/358143541>
- Matheson, J. (2015). *The epistemic significance of disagreement*. Palgrave Macmillan.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press.
- Mercier, H. and D. Sperber. (2017). *The Enigma of Reason*. Harvard.
- Murphy, A. H., & Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 26(1), 41-47.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175-220.
- Persky, J. (1995). Retrospectives: The ethology of homo economicus. *Journal of Economic Perspectives*, 9(2), 221-231.
- Peter, F. (2019). Epistemic self-trust and doxastic disagreements. *Erkenntnis*. pp. 1189-1205.
- Pohl, R. (Ed.). (2017). *Cognitive illusions*. 2nd ed. Routledge.
- Priest, M. (2016). Inferior disagreement. *Acta Analytica*. 31: pp. 263-283.
- Pronin, E., Lin, D., & Ross, L. (2002). The bias blind spot: Perception of bias in self versus others. *Personality and Social Psychology Bulletin*. 369-381.
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, 111(3), 781-799.
- Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in cognitive sciences*, 11(1), 37-43.
- Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, 43(4), 565-578.

- Reber, R. (2017). Availability. In R. F. Pohl (Ed.), *Cognitive Illusions: Intriguing phenomena in thinking, judgment, and memory* (second edition) (pp. 185-203). Routledge.
- Reeder, G., Pryor, J., Wohl, M., & Griswell, M. (2005). On attributing negative motives to other who disagree with our opinions. *Personality and Social Psychology Bulletin*, 1498-1510.
- Reyes, R. M., Thompson, W. C., & Bower, G. H. (1980). Judgmental biases resulting from differing availabilities of arguments. *Journal of Personality and Social Psychology*, 39(1), 2.
- Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of personality and social psychology*, 37(3), 322.
- Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imagery. *Personality and social psychology bulletin*, 11(1), 118-127.
- Shi, F., Teplitskiy, M., Duede, E., & Evans, J. A. (2019). The wisdom of polarized crowds. *Nature human behaviour*, 3(4), 329-336.
- Shogren, J. F., Shin, S. Y., Hayes, D. J., & Kliebenstein, J. B. (1994). Resolving differences in willingness to pay and willingness to accept. *The American Economic Review*, 255-270.
- Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 99-118.
- Sosa, E. (2011), The epistemology of disagreement. In Haddock, A., Millar, A., & Pritchard, D. (Eds.), *Social Epistemology*, OUP.
- Stanovich, K. E., & West, R. F. (2008). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking & Reasoning*, 14(2), 129-167.
- van Inwagen, P. (2010). We're right. They're wrong. In Feldman, R., & Warfield, T. (Eds.), *Disagreement* (pp. 53-68). Oxford University Press.
- Wald, B. E. (2009). Dealing With disagreement: Distinguishing two types of epistemic peers. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, 3(1), 113-122.

- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology*, 12(3), 129-140.
- Wedgewood, R. (2010). The moral evil demons. In Feldman, R. & Warfield, T. (Eds.), *Disagreement* (pp. 216-246). Oxford University Press.
- West, R., R. Meserve, and K. Stanovich. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*. 103(3), 506-519.
- Wheeler, G. (2018). Bounded rationality. *Stanford Encyclopedia of Philosophy*.  
<https://plato.stanford.edu/entries/bounded-rationality/>