Defending Philosophical Knowledge

Item Type	dissertation
Authors	Dixon, Jonathan
DOI	10.7275/20201463.0
Download date	2025-03-15 04:35:40
Link to Item	https://hdl.handle.net/20.500.14394/18545

Defending Philosophical Knowledge

A Dissertation Presented

by

JONATHAN DIXON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

MAY 2021

Philosophy

© Copyright by Jonathan Dixon 2021

All Rights Reserved

Defending Philosophical Knowledge

A Dissertation Presented

by

JONATHAN DIXON

Approved as to style and content by:

_____
Hilary Kornblith, Chair

_____
Christopher Meacham, Member

_____
Sophie Horowitz, Member

_____
Seth Cable, Member

_____
Phillip Bricker, Department Head
Philosophy

DEDICATION

For Charlie.

ACKNOWLEDGMENTS

I could not have completed this dissertation without the help and support of the many people I thank below.

I would first like to thank the members of my dissertation committee from the UMass Philosophy Department: Hilary Kornblith, Christopher Meacham, and Sophie Horowitz. Your comments, feedback, and helpful suggestions were invaluable in the development of my arguments, in revising this manuscript, and in making this document much better than it otherwise would have been. You are my epistemological dream team. Additionally, I would like to thank you all for the courses that I took from you during my time as UMass. They were vital parts of my philosophical education and development. I will miss taking classes from you all. I would also like to thank Seth Cable from the UMass Linguistics Department for agreeing to be the outside member of my dissertation committee. I hope my defense of philosophical knowledge also serves to defend linguistic knowledge in the way you hoped.

Of my committee members, I would especially like to thank my advisor, Hilary Kornblith, for being the best advisor a graduate student could hope for. Hilary's help through my graduate career spanned both time and distance. His timely and thoughtful feedback was integral to the development of my arguments, and I could not have written this dissertation without his sage advice. Hilary also went beyond the call of duty to ensure that my complicated path through graduate school was as painless and as productive as possible. Your support and guidance as my advisor will be missed.

I also want to acknowledge and express my appreciation for the support I received from the UMass Philosophy Department during my time there. Thanks to the department

chairs during my time, Joe Levine and Phillip Bricker, for their flexibility and support in allowing me to take an unusual path through graduate school. Thanks to all the UMass Philosophy faculty to providing excellent courses that both broadened and deepened my love for our discipline. And thanks to all the UMass Philosophy graduate students for the many conversations, discussions, laughs, hang-outs, parties, and any gathering where we became good friends.

I owe a debt to my undergraduate professors at California State University Long Beach for originally inspiring me to pursue philosophy academically. And I thank my graduate professors at Virginia Tech for taking my potential and adding some academic rigor to it.

Over the preceding years I was fortunate to presented different versions of the argument in this dissertation at various conferences. Parts of chapter 1 were present at the 3rd Annual Philosurfer Convergence (April 2018), and at the 2019 Eastern APA (January 2019). Parts of chapter 2 were present at the 4th Annual Edinburgh Graduate Epistemology Conference (June 2014). Parts of chapter 3 were presented at the 2019 North Carolina Philosophical Society Conference (March 2019), at the 4th Annual Philosurfer Convergence (June 2019), at the What is Philosophy? And Other Topics in Metaphilosophy Conference at Johns Hopkins University (Oct 2019), and at the Disagreement and Philosophy Graduate Student Conference at Ryerson University (March 2020). Parts of chapter 4 were almost present at the 2020 Joint Conference of the North Carolina Philosophical Society and the South Carolina Society for Philosophy (March 2020) and at the Ohio Philosophical Association 2020 Annual Meeting at the University of Akron (April 2020). Unfortunately, both conferences were cancelled due to

the corona virus outbreak in Spring 2020. I would like to thank all the organizers and conference participants on these occasions for their valuable questions, objections, concerns, and suggestions.

And finally, thank you to all those in my life who encouraged me. My parents Joe and Sheila Dixon have given me an excess of love and support. I could not have gotten this far without them behind me through each phase of life. Thanks to my siblings Tracy and Dana, sibling-in-law Steve Carrol, and nephew Jonathan for keeping me laughing and feeling loved. Likewise, thanks to my step-family, Charles, Frances, Jay, Nicole, Caroline, Bryan, and my nieces and nephews, Eloise, Jack, Henry, Sam, and Sarah for making me a part of your loving family. Most importantly, thank you to my partner and philosophical co-conspirator Elizabeth Hupfer. She is one of the most loving and intelligent people I know. To me, she is the best person – in all possible worlds. Beth: thank you for living this life with me; I would not want to live it any other way. Sine qua non.

ABSTRACT

DEFENDING PHILOSOPHICAL KNOWELDGE

MAY 2021

JONATHAN DIXON, B.A., CALIFORNIA STATE UNIVERSITY LONG BEACH

M.A., VIRGINIA POLYTECHNICAL UNIVERSITY AND STATE UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Hilary Kornblith

This dissertation concerns whether philosophy as a discipline can, and does, produce philosophical knowledge. Specifically, this dissertation concerns several prominent arguments for philosophical skepticism. Some support philosophical skepticism by arguing that the philosophical practice of appealing to intuitions to justify philosophical beliefs is illegitimate because either intuitions are not a legitimate kind of evidence or intuitions are an unreliable source of justification. Others argue that philosophical knowledge is untenable because philosophers rarely, if ever, resolve their philosophical disagreements despite spending their professional lives attempting to do so. In brief, the purpose of this dissertation is to defend philosophical knowledge from these arguments by showing that philosophical knowledge is not threatened by either intuition or disagreement skepticism.

PREFACE

The primary aim of this dissertation is to save the discipline of philosophy from two different (but related) kinds of arguments for philosophical skepticism. The first concerns the supposed widespread philosophical practice of appealing to intuitions as evidence to justify philosophical theses. Many now argue that this practice is unjustified because either intuition cannot legitimately be counted as evidence, or even if they can be legitimately counted as evidence, they are an unreliable source of justification. The second concerns the undeniable fact that there is widespread disagreement among philosophers about all, or nearly all, substantive philosophical theses. Many now argue that this undeniable fact shows that philosophical knowledge is untenable because philosophers rarely, if ever, resolve their philosophical disagreements despite spending their professional lives attempting to do so.

To address and rebut these skeptical arguments, this dissertation will proceed as follows:

Chapter one concerns the preliminary question of whether reliability is a necessary condition for knowledge. This question is significant because if knowledge does not require reliability, then philosophical knowledge is not threatened by the unreliability of intuitions or by philosophy's poor track-record of resolving disagreements. In this chapter I argue that this way of avoiding philosophical skepticism is mistaken because reliability must be a necessary condition for knowledge. To accomplish this, I first respond to John Turri's recent arguments against the reliability condition for knowledge and then show that the reasons why Turri's arguments fail help to demonstrate why reliability must be a necessary condition for knowledge.

Chapter two concerns the challenge to philosophical knowledge posed by intuition skepticism. The intuition skeptic argues that because intuitions are not a proper kind of evidence or are unreliable, they cannot render knowledge. Philosophers like Pust, Bealer, and Bonjour have argued that any such argument is self-defeating because the premises of any such argument will be, at least in part, supported by some intuitions. While intuition skeptics are keenly aware of this danger, they either respond to this threat directly or take measures to avoid this threat by limiting the scope of their intuition skepticism. Examples of the former include Silva who argues that epistemic self-defeat is not a threat to an Unreliability Argument for a global kind of intuition skepticism. Examples of the latter include Machery who argues that experimental data supports an Unreliability Argument for a near-global skepticism against *philosophers' use* of the method of cases, which itself (allegedly) does not rely on intuitions about cases. The aim of this chapter is to show that the self-defeat challenge is not as easily dealt with or avoided as the above philosophers would have us believe, and that this challenge places severe limits on the kind of intuition or philosophical skepticism that an Unreliability Argument can establish. In short, I argue that the power and scope of the self-defeat challenge has been underappreciated in this literature.

Chapter three concerns whether disagreement among philosophers leads to philosophical skepticism. Conciliationism is roughly the family of views that hold that rationality requires agents to reduce confidence or suspend belief in p when epistemic peers (i.e. agents who are about as equally well-informed and intellectually capable) disagree about p. While Conciliatory views are *prima facie* plausible, they seemingly lead to a pervasive philosophical skepticism since a great many (if not all) philosophical

propositions are disagreed upon by philosophical peers. In this chapter I aim to save philosophical knowledge from this kind of disagreement skepticism by arguing that all plausible versions of Conciliationism are false because they are epistemically self-undermining, veridically self-undermining, and lead to a hitherto unrecognized *reductio ad absurdum.* I then explain how these results can be extended to other competing views to Conciliationism and conclude by reflecting on the counterintuitive consequences of these results.

Chapter four extends the arguments in chapter three about general philosophical disagreements to combat one of the oldest and most discussed challenges to moral knowledge: the argument from *moral* disagreement to moral skepticism. And while this chapter does discuss the general peer disagreement literature, addressing moral disagreement arguments is dialectically well-founded because there are many additional and varied kinds of moral disagreement to moral skepticism arguments that do not rely on peer disagreement. So ironically, focusing on moral disagreements arguments allows this chapter to have a wider scope and cover more argumentative ground than the previous chapter on the threat posed by peer disagreement to moral (and non-moral) knowledge. And while there are many different versions of this argument in the literature, it seems that all (or nearly all) moral disagreement arguments share an underlying structure. In this chapter I argue that all moral disagreement arguments that satisfy this underlying structure cannot establish moral skepticism because this underlying structure leads to a previously unrecognized *reductio ad absurdum*. In short, I argue that this reductio argument (very likely) refutes all versions of the moral disagreement to moral skepticism argument in one fell swoop.

In sum, my dissertation argues that, while reliability is a necessary condition for knowledge, philosophical knowledge is not threatened by either intuition or disagreement skepticism.

TABLE OF CONTENTS

Page

CHAPTER 1

RELIABLE KNWOELDGE: A REPLY TO TURRI

Many epistemologists agree that knowledge must be reliably produced. For example, Goldman holds that justification is necessary for knowledge and that justification "is a function of the reliability of the process or processes that cause it" (1979: 345); Sosa holds that knowledge is produced by a disposition "that would in appropriately normal circumstances ensure (or make very likely) the success of any relevant performance issued by it" (2007: 29); and Williamson claims that "no reason has emerged to doubt the intuitive claim that reliability is necessary for knowledge" (2000: 100).[1] Recently John Turri (2015) argues against this orthodoxy by providing two theoretical arguments for the possibility of unreliably produced knowledge. If either of Turri's arguments is sound then all accounts of knowledge that require reliability are false and most epistemologists have been on the wrong track in understanding the nature of knowledge. Realizing this, Turri (2015a, 2015b, 2016, 2019) outlines a new account of knowledge, called *abilism,* which allows for knowledge to be unreliably produced. Turri (2016) also supports his account of knowledge by providing some empirical data from experiments he conducted which seemingly show that unreliably produced knowledge is compatible with our folk intuitions about knowledge.

[1] See Turri (2015: fn 1) and Goldman and Beddor (2015) for a more complete list of epistemologists who think that knowledge must be reliably produced. However, there are a few contemporary philosophers who indirectly deny this claim. Sartwell (1991, 1992) argues that true belief alone is sufficient for knowledge; and because Hetherington (1998, 1999, 2016) argues that victims of Gettier-cases do possess the relevant knowledge, *a fortiori,* he holds that lucky (and so unreliable) processes can render knowledge.

After providing some background and clarifying terms in §1, in §2 and §3 I explain why each of Turri's (2015) theoretical arguments for unreliable knowledge fails. In §4 I explain Turri's account of knowledge and argue that while Turri's account of knowledge is false, it does help to explain why certain unreliable processes are nonetheless epistemically valuable. In §5 I provide reasons to doubt Turri's (2016) experimental evidence for his abilist account of knowledge.[2] And I conclude in §6 with reasons why reliability must be a necessary condition for knowledge.

1. Background and Clarifying Terms

I should first mention that Turri's (2015) theoretical arguments for unreliable knowledge rely on what is called an *achievement account of knowledge*. This is roughly the family of views which hold that an agent S has knowledge of P just in case S's true belief in P manifests S's cognitive achievement.[3] While there are many ways of spelling-out the details of this account of knowledge and there are many challenges to this family of views,[4] I will set these issues aside and grant for the sake of argument that knowledge is a kind of cognitive achievement because I claim that *even if* we grant this, both of Turri's (2015) arguments for the possibility of unreliable knowledge fail.

The next thing I should explain is what Turri means by 'reliability' and 'achievement'. Turri's definition of 'reliability' is in line with how it is standardly

[2] While Turri discusses his account of knowledge and its potential advantages elsewhere (Turri 2015a, 2015b, 2016a, and forthcoming), I will be focusing my discussion on Turri (2015 and 2016) because this is where his arguments that directly attempt to establish unreliable knowledge are found.

[3] This general account of knowledge is part of the ongoing research program in epistemology called *performance-based epistemology* which is exemplified by virtue epistemology. See Sosa (2007), Zagzebski (2009), and Greco (2010) for more details on this research program, virtue epistemology, and achievement accounts of knowledge.

[4] For example, see Pritchard (2008, 2009) and Lackey (2007, 2009).

understood: a process, disposition, or ability is (epistemically) reliable when and only when (significantly) more than half of its produced beliefs are true; and a process, disposition, or ability is (epistemically) *un*reliable when and only when less than half of its produced beliefs are true (Turri 2015: 530).[5] Unfortunately, Turri (2015) does not provide a definition of 'achievement'. Instead, Turri provides several examples of what he takes to be achievements to illustrate how this notion should be understood. The important thing for Turri is that *achievements need not be reliably produced.* For example,

> Ted Williams is the best baseball hitter ever. But he normally failed to get a hit. The relevant ability could at best be counted on to produce a hit about four in ten times: his best yearly average ever was .407 and his lifetime average was much lower. But certainly many of his hits were achievements. And many hits by many lesser hitters are no doubt achievements too, despite the fact that these hitters fail to get a hit seventy-five percent of the time or more. (Turri 2015: 531)

Examples like these might give the impression that 'achievements' involve prolonged practice and considerable effort. However, Turri also gives examples of achievements that do not require much practice or effort. A novice musician who plays a chord for the first time, a child who takes his first step or speaks his first sentence, and a rookie golfer who makes par for the first time are all examples of achievements for Turri. Indeed, given these examples 'achievements' need not involve any special degree of pre-existing ability because "achievement can issue from even highly *un*reliable ability" (Turri 2015:

[5] Understood this way, 'reliability' is determined by the *frequency* of past events rather than by a *propensity* of events to occur. While Turri does acknowledge the propensity definition of reliability (cf. Alston 1993: 8-9, and Goldman 1979: 345), all of Turri's examples of reliable processes assume the frequency understanding of reliability. This will be discussed more in §2.

531). Turri (2015) also does not define 'unreliable ability' but I take it that an agent has an unreliable ability to Φ iff in using this ability to Φ the agent fails to Φ most of the time. So, for Turri, achievements involve simply attaining one's intended outcome through one's (un)reliable process, disposition, or ability.

Lastly, it is worth pointing out that this account of achievement is unique among those who hold an achievement account of knowledge because it does not require that achievements manifest one's *competence* (see Sosa (2007), Zagzebski (2009), and Greco (2010)). For these authors, achievements require competence and competence involves the reliability of processes, dispositions, or abilities. Therefore, their achievement accounts of knowledge require reliability. Turri (2016a) explicitly points out this omission and Turri (2015a, 2015b, 2016a, and forthcoming) endorses this as a beneficial feature of his achievement account of knowledge because it avoids problems Turri sees for these authors' accounts of knowledge. Turri's account of knowledge is further explained in §5 and directly challenged in §6 below.

2. Against Turri's First Argument

Turri's first argument for the possibility of unreliable knowledge is

1. Achievements don't require reliable abilities. (Premise)
2. If achievements don't require reliable abilities, then unreliable knowledge is possible. (Premise)
3. So unreliable knowledge is possible. (From 1 and 2) (Turri 2015: 531)

Turri supports the first premise by referencing the many examples he provides of achievements issuing from unreliable abilities provided above. Turri supports the second premise by saying,

> If athletic, artistic, social and academic achievements can manifest unreliable ability – it stands to reason that it is possible for knowledge to do so too. Knowledge is an intellectual achievement, so absent a special reason to think otherwise, we should expect it to share the profile of achievements generally. (Turri 2015: 532)

In other words, premise two is plausible because if knowledge is a kind of (intellectual) achievement and achievements generally do not necessarily issue from reliable processes, abilities, or dispositions, then it is possible that knowledge can be unreliably produced as well. In short, Turri's argument attempts to shift the burden of proof on those who believe reliability is a necessary condition of knowledge to show why knowledge, as an intellectual achievement, cannot issue from unreliable abilities. Therefore, unreliable knowledge is possible.[6]

Turri's first argument fails because it faces a dilemma: Either the first premise is false or the argument as a whole begs the question. The first premise is false if it is interpreted to mean "*all* achievements don't require reliable abilities." There are many achievements that require reliable abilities. More specifically, achieving some goal often requires reliably performing some action. For example, to win a competitive darts or archery tournament often requires one to reliably hit their intended mark.[7] Indeed, to achieve the goal of performing some action with 90%+ accuracy (e.g. hitting a bullseye in archery, hitting a baseball, playing a piece of music, or walking) requires performing

[6] Turri (2015, 2015a, 2015b, 2016, 2016 and forthcoming) never specifies how 'possibility' should be understood. I likewise will not assume any particular account of 'possibility.'

[7] However, in order to achieve some goal one need not reliably achieve that goal (e.g. to win the archery competition one need not reliably win the archery competition). Thanks to an anonymous reviewer for helping to clarify this.

this action with 90%+ accuracy. So, the proper interpretation of the first premise must be something like "*some* achievements don't require reliable abilities." However, if this interpretation is placed back into the argument above then it begs the question. The second premise would now read "if *some* achievements don't require reliabilities, then unreliable knowledge is possible." But since Turri has said nothing against the possibility that knowledge is the kind of intellectual achievement that requires reliability (like the ones listed above), Turri has not provided adequate reason to think that *knowledge* is the kind of achievement that can be unreliably produced – which is the purpose of the argument. So, in order for this argument to conclude "unreliable knowledge is possible," it must beg the question.[8]

Turri anticipates and responds to this dilemma[9] by claiming that it can be avoided if we interpret the first premise as a proposition "about dominant tendencies, or what is typical, or what is natural and normal for a kind" (Turri 2015: 534). For example, the propositions that "humans don't have eleven fingers" or "cats don't have two faces" express tendencies about how humans and cats anatomy are typically constituted (Turri 2015: 534). Although there are exceptions to these claims, these exceptions do not render these claims false when these claims express such tendencies. So, if premise one is understood as a tendency proposition, Turri claims his argument "would still be plausible because, as already mentioned, we would expect knowledge to fit the profile of

[8] An anonymous reviewer helpfully pointed out that Turri's account of achievement might allow for the manifestations of epistemic abilities like self-deception (i.e. I have good justification for P and P is true, but I deceive myself into believing ~P) to count as achievements. And since such achievements are clearly not knowledge, this provides an additional reason against Turri's claim that knowledge shares the general profile of achievements.

[9] Turri (2015: fn. 7) attributes this dilemma to Bruce Russell.

achievements generally, unless we're given a special reason to think otherwise" (Turri 2015: 534).

This response still fails for the reasons mentioned above. Even if we grant that premise one is a tendency proposition Turri has not established that achievements have a general tendency to be unreliable. As argued above, there are a large number of achievements that require reliability. Turri's few examples of unreliable achievements are insufficient to establish that premise one is a tendency proposition. Furthermore, Turri has provided no positive reason to think that *knowledge* is kind of achievement that can be unreliably produced – which (again) is the purpose of the argument. So, Turri's first argument fails because it either has a false premise or begs the question.

Against Turri's Suggestive Achievement Examples

Although Turri's first argument has been refuted, one might be seduced by Turri's suggestive examples into thinking that unreliable knowledge is still possible. That is, if knowledge is a kind of achievement and *some* achievements can be unreliable produced, then might there be examples of unreliable knowledge? For instance, it might be thought that when Ted Williams gets a hit on a particular occasion, using his unreliable hitting ability, that this hit shows that unreliably produced knowledge is possible; or it might be thought that when a novice musician uses their nascent ability to play a chord for the first time that this shows unreliably produced knowledge is possible. In order to dispel this lingering worry, I argue that all of Turri's examples are misleading and because of this they do not convincingly support the possibility of unreliable knowledge.[10]

[10] I will focus on Turri's novice musician and athletic examples (e.g. Ted Williams) and set aside Turri's examples involving infants because, since it is not clear that infants are cognitively developed enough to have knowledge bearing mental states, such examples cannot convincingly show that unreliable knowledge is possible.

To accomplish this, it is important to note two features that all of Turri's examples share. First, all of Turri's examples understand reliability as the *frequency* of achieving/successfully performing intended *physical actions* (e.g. Ted Williams' batting average, playing a chord for the first time). Second, the kind of knowledge at issue in all of Turri's examples is *knowledge-how* to perform physical actions (e.g. knowing-how to get a hit, play a chord, etc.). Taken together, all of Turri's examples use one's frequency of Φ-ing (i.e. one's track-record of achieving/successfully performing intended physical actions) to support the possibility that one unreliably knows-how to Φ. For instance, Ted William's .407 batting average is used to support the possibility that he unreliably knows-how to get a hit. It is these two features that, I argue, make all of Turri's examples misleading and unconvincing.

Understanding reliability as the frequency of Φ-ing is problematic because such frequencies do not take into account the distorting effects of Φ-ing in difficult circumstances. In general, difficult circumstances can hinder one's ability to Φ, and if we only consider one's track-record of Φ-ing in such circumstances, it will misleadingly appear that one is less reliable at Φ-ing generally.[11] For example, my bike-riding track-record in difficult circumstances (e.g. on a snowy mountain, or underwater) does not adequately reflect my bike-riding reliability.[12] Turri's Ted Williams example is likewise

[11] I am assuming that 'difficulty' is always relative to an agent's, or class of agent's, ability. See Bradford (2015, esp. 27-28) for a defense of this claim.

[12] Virtue epistemologists often stress this point (e.g. Sosa (2007: 29) argues that knowledge must be produced from an agent's competent disposition "that would in *appropriately normal conditions* ensure (or make highly likely) the success of any relevant performance issued by it" (my emphasis)). I am assuming, along with many virtue epistemologists, that the notions of 'difficult' and 'normal' circumstances are intuitive enough to not need further explication (cf. Alston (1995: 10)).

misleading because only looking to Williams' batting average to support that he is an unreliably hitter obscures the difficulty of getting a hit in the major leagues. One could object that because the achievement in question is Williams' ability to get a hit in the major leagues that Williams' batting average in these difficult circumstances is the relevant indicator of his hitting reliability. To be clear, I do not deny that frequencies are relevant for *assessing* whether something is reliable in some circumstances since frequencies are often the most direct way of doing so (cf. Alston 1995: 6). The point here is that frequencies *alone* cannot properly characterize the nature of something's reliability without taking into account the potential distorting influence of difficult circumstances. And since Turri's examples are not sensitive to this they are misleading and unconvincing.[13]

Understanding reliability as a frequency is also problematic since one can reliably Φ even if they have never Φ-ed. For example, while I have never baked a cake I may be reliably able to do so (in appropriate circumstances) if I memorized the recipe; or while I never have planted a tree I may reliably be able to do so if I watched a video tutorial. In such cases, if we use the frequency that I have performed these actions to reflect my reliability, then it will misleadingly appear that I am unreliable at performing these actions since my track-record is *zero*. Likewise, Turri's examples are misleading since it is very plausible that one can reliably get a hit or play a chord (in some circumstances) without ever having done these things.

This last point is additionally relevant to Turri's examples since, like reliability, one can *know-how* to Φ even if they have never Φ-ed. For example, I can know-how to

[13] Thanks to an anonymous referee, Christopher Meacham, and Sophie Horowitz for pushing me to be clearer on this point.

bake a cake and plant a tree, despite never having done these things, because I memorized the recipe and watched the video tutorial. In such cases, if we use the frequency that I have Φ-ed, as Turri does, to support that I unreliably know-how to Φ, then it will misleadingly appear that I am both completely unreliable and lack the knowledge-how to Φ since my Φ-ing track-record is *zero*.[14] Again, Turri's examples are misleading since it is very plausible that one can know-how to hit a baseball or play a chord without ever having done these things (e.g. by having knowledge-how by description rather than by acquaintance; or alternatively, having theoretical know-how rather than applied know-how).[15]

Ultimately, Turri's examples are misleading and unconvincing for two underlying reasons. Firstly, by using the frequency of Φ-ing to establish unreliable knowledge-how, they mask or gloss-over underlying *cognitive processes* that are potentially reliable and vital to an agent's knowledge. For instance, one can retain the knowledge-how to Φ even if one loses the ability to Φ. If a famous piano player were to lose her arms in a car accident, she would intuitively still know-how to play the piano (Stanley and Williamson 2001: 416; and Stanley (2011)). But if we were to use her Φ-ing track-record after the

[14] If you assume that all knowledge can be downloaded/implanted in one's mind or trained through simulation, then knowledge-how to Φ never requires having Φ-ed since agents with this knowledge will never have Φ-ed. But, in real world cases there are certain kinds of knowledge-how that can only be acquired by successfully performing the desired action (possibly many times). Regardless, frequencies do not necessarily indicate the (un)reliability of one's know-how, and so they cannot convincingly be used to show unreliable know-how.

[15] An anonymous reviewer helpfully suggested that Turri might be trading-off on Ted Williams' broader abilities to play baseball to misleadingly argue that he has unreliable knowledge-how. This amounts to a kind of generality problem (about which abilities are responsible for one's (un)reliable know-how) that is similar to the well-known generality problem (about which cognitive processes are responsible for one's reliability). I set these issues aside in this chapter but I follow Comesana (2006) and Bishop (2010) in thinking that this original problem will arise for any plausible theory of justification.

accident to indicate her reliability it will misleadingly appear that she has unreliable know-how since her Φ-ing track-record will now *approach zero*. Likewise, it seems Ted Williams would reliably know-how to hit a baseball as long as he retains the cognitive processes that allow him to get a hit (when he is able-bodied).

Secondly, and contra Turri, reliability as it relates to knowledge consists in a *propensity* of a *cognitive process* to produce many more true beliefs than false beliefs. On this more common view of reliability,[16] one's reliability involves whether one's beliefs on how to Φ are reliably produced, not whether one is able to Φ reliably. This understanding of reliability is preferable because it can, while the frequency understanding cannot, capture why some paradigm reliable processes are reliable. For example, practicing Φ is generally thought to be a reliable way of acquiring the ability and knowledge-how to Φ even though one's track-record will often not adequately represent this. A novice musician's frequency of playing a certain chord might consist of many failed attempts followed by one successful attempt where she "gets the knack of it" and can now play this chord whenever. So, if we understand her reliability as the frequency of her Φ-ing to show she has unreliable knowledge, as Turri does, then it will misleadingly appear that she is has unreliable knowledge. In contrast, it seems that practicing is a reliable way of gaining knowledge-how to Φ because it has a propensity to produce (and reinforce) true beliefs about Φ-ing. Presumably, Ted Williams and the

[16] For example, Goldman writes "*reliability* consists in the tendency of a [cognitive] process to produce beliefs that are true rather than false" (Goldman (1979: 345)) (also see Alston (1995: 5-10), Comesana (2006: fn. 1), and Goldman and Beddor (2015)). Goldman also says that our ordinary conception of justification is ambiguous between the frequency and propensity understandings of reliability because we usually assume the former will match the latter. My arguments in this section explain how these two understandings of reliability can come apart. I will set aside the many other difficulties in articulating what 'reliability' amounts to (see Frise (2018) and Tolly (forthcoming)).

novice musician practiced for many hours to acquire the underlying cognitive processes with the propensity to form more true than false beliefs about performing their desired actions.

Thus, this shift from frequencies to propensities and from physical actions to cognitive processes explains why Turri's examples are misleading and unconvincing. This shift also explains why Turri's first theoretical argument begs the question. Because it is problematic to understand reliability with frequencies (for all the reasons given above), any example that relies on them to establish unreliable knowledge will be misleading and unconvincing and so will likely beg the question against reliable knowledge.

A better strategy for Turri to establish that unreliable knowledge is possible is to take a more direct route by providing an example where one intuitively knows some proposition P even though one's true belief that P was formed by a cognitive process that has a propensity to render more *false* than true beliefs. This is what, I believe, Turri's second argument for unreliable knowledge attempts to do. In §5 I will take on the burden of proof and argue that reliability is a necessary condition for knowledge.

3. Against Turri's Second Argument

Turri's second and more direct argument for the possibility of unreliable knowledge involves explanatory inference (aka, inference to the best explanation or IBE). As Turri notes, IBE is used in scientific reasoning and in everyday life to provide probable explanations for a set of data or certain phenomena. What best explains the fact that humans and chimpanzees have so many anatomical similarities? We have a common ancestor. What best explains the appearance of a new jug of milk in the fridge? My

spouse bought it at the store. Turri claims that this kind of reasoning supports the possibility of unreliable knowledge:

> The epistemic efficacy of explanatory inference supports the view that unreliable knowledge is possible. Inference to the best explanation yields knowledge if the explanation that we arrive at is true. But even when it is true, the best explanation might not be very likely. So our disposition to infer to the best explanation might not be reliable. So unreliable knowledge is possible. (Turri 2015: 536)

That is, even though IBE is often unreliable, the explanations it provides (when true) can yield knowledge. More specifically, some hypothesis 'H' can best explain a set of data 'D' in our world even if there is a greater number of (nearby) possible worlds were D obtains and H is false (Turri, 2015: 536-537).[17]

To illustrate this argument, Turri provides a case study involving the television show *House M.D.* Gregory House (the protagonist) is a world renowned medical doctor who has an incredible ability to diagnose patients where other doctors have failed. Simply put, he is the best of the best; he is the Ted Williams of diagnosticians. However, despite being the best, House misdiagnoses patients a lot. Indeed, nearly every episode follows the same structure where House misdiagnoses the patient several times before coming to the right diagnosis just in the nick-of-time to save the patient's life. Turri contends that House's method for diagnosing patients is IBE – House infers a hypothesis/diagnosis that best explains the data/symptoms. And with each failed diagnosis House gains new insights to symptoms that inform his subsequent diagnoses. Given this description of

[17] There is controversy about whether IBE can provide explanations and/or render knowledge (e.g. van Fraassen (1989)). So, Turri's second argument has the important caveat that one must first accept that *IBE can produce explanations / knowledge* before this argument can be persuasive.

House's track record, Turri argues that House's reliability is considerably less than .5. But despite House's unreliability, when he ends up correctly diagnosing his patient "House knows what disease that patient has" (Turri 2015: 538). In short, this case study shows that IBE "can yield knowledge, even though it doesn't yield the correct verdict most of the time" (Turri 2015: 539). Turri summarizes his second argument as:

1. If House knows, then unreliable knowledge is possible. (Premise)
2. House knows. (Premise)
3. So unreliable knowledge is possible. (From 1 and 2)

> The argument is valid. Line 1 is supported by the fact that House's method usually produces false beliefs. Line 2 is supported by intuition, and by the fact that millions of viewers, including trained epistemologists, detect no incoherence in the story line, week after week, over many seasons. (Turri 2015: 539)

I believe that both premises of Turri's second argument are false because Turri misrepresents House's medical abilities and knowledge. While Turri is right that House's diagnostic track record is well below .5, Turri takes the lesson here to be that, despite his track-record, "House knows" the correct diagnosis when he gets it right via IBE because House has a special ability to figure out the right diagnosis more often than any other doctor. This misrepresents House's abilities because, contra Turri, House is remarkable at getting the right diagnosis not because he knows the correct diagnosis more often than any other doctor, but because he has a remarkable ability to propose novel diagnostic hypotheses worthy of consideration and testing. But this ability to come up with possible explanations of patient's symptoms does not itself allow House to know that his

diagnoses are correct until the treatment actually works (or when the reliable test results confirm his diagnosis).[18]

To illustrate these points, consider the following case that parallels Turri's House example:

> Jessica has very poor eye sight and is legally blind without her glasses. However, despite her eyesight, Jessica has a special ability to correctly identify pictures without her glasses. While others who are similarly handicapped can only identify pictures 5% of the time on average, Jessica is able to correctly identity such images 25% of the time on average. Now imagine that Jessica is presented with an image of a basketball that she and others with her eye sight phenomenologically describe a blurry spot of reddish orange. Without her glasses Jessica infers incorrectly three times in a row that the picture is of an orange fruit, the Sun, and then a Lego piece. After each incorrect answer or hypothesis Jessica is told new information about the image that reveals why her answers were incorrect e.g. it is not a fruit for her orange fruit hypothesis, it is an object you can touch for her Sun hypothesis, and it is an object that is bigger than a Lego piece. After all of this Jessica then answers correctly, but is not yet told that she is correct.

The crucial question to now ask is: At this point, does Jessica *know* what the picture is of? Intuitively, the answer is no. While Jessica, like House, has a special ability to get it right more often than her peers, this is not because she knows the correct answer more

[18] I am indebted to Hilary Kornblith for a discussion on these points.

often, but because she is better at coming up with worthy hypotheses.[19] And, like House, Jessica does not know her hypothesis is correct until it's confirmed. Thus, premise two of Turri's argument is false because before the proposed treatment works (or when a reliable test result confirms a diagnosis) House does not know whether his hypothesized diagnosis is correct. Premise one is also false because if we plug this understanding of what House knows back into the antecedent of this premise, it renders the consequent false. That is, if "House knows" is understood to be true only after his hypothesized diagnosis has been tested and confirmed, then House's knowledge is not an instance of unreliable knowledge.[20]

4. Against Turri's Account of Knowledge

[19] This misrepresentation of House's abilities is related to another debate concerning the nature of IBE. There is a tradition, going back to Pierce, of distinguishing abduction from IBE. Traditionally, abduction is concerned with hypothesis construction while IBE is concerned with selecting the hypothesis that is most likely to be true from a set of hypotheses. And traditionally, it is a feature of abduction to be unreliable in order to produce a variety of hypothesis to be tested. Furthermore, there has been a recent trend to conflate these two (see Mcauliffe (2015) for a defense of these points). On this understanding, House does not use IBE but uses abduction and Turri (2015) conflates these two when he writes "House and his team explicitly reason abductively" (537) and "House's method for trying to solve the case *just is* to employ inference to the best explanation" (540, his emphasis).

[20] Turri's claim that IBE is unreliable but can render knowledge also concerns two other recent debates about the nature of IBE. First, Turri's claim is related to the issue of whether van Fraassen's Bad Lot objection shows that IBE is an unreliable inference form. The Bad Lot Objection argues that IBE is an inadequate inference form because it has no way of discerning whether a set of hypotheses are all false and so would always lead to a false conclusion. Schupbach (2014) recently argues that van Fraassen's Bad Lot Objection does not establish that IBE is an unreliable inference form any more than we can show that modus ponens is unreliable by plugging in false premises (see Dellsen (2017) for a response). If this were true, then it would be inappropriate to characterize either IBE or modus ponens as unreliable. Second, many have argued that IBE is a heuristic that approximates objective Bayesian reasoning. On this understanding, IBE's primary function is to locate the "most probable available explanatory hypothesis to serve as a working hypothesis in an agent's further investigations" (Dellsen (forthcoming)). This understanding dovetails with my explanation of House's use of IBE above. In short, these issues concerning the nature and reliability of IBE impact and potentially undermine the soundness of Turri's argument that IBE can unreliably render knowledge.

After Turri (2015) presents his two theoretical arguments for the possibility of unreliable knowledge, he outlines an account of knowledge, called *ecumenical reliabilism* (elsewhere, *abilism* – see §5), which allows for this kind of knowledge. Although the theoretical motivation for proposing such a view has been eliminated, Turri's account of knowledge has some *prima facie* plausibility that needs to be accounted for by its detractors. My aim in this section is to explain Turri's account of knowledge and show how the intuitions that seemingly support it can be accounted for. Once these intuitions are adequately understood the *prima facie* plausibility of Turri's account dissipates.

Turri explains his account by first making a distinction between what he calls the *truth-conducive sense* of reliability and the *trustworthy sense* of reliability. The former sense of reliability is what Turri claims many philosophers use in their accounts of knowledge: having a true-belief track-record (significantly) greater than .5. Turri describes the latter sense of reliability by saying

> You're trustworthy (in some respect, in some circumstances) if it's appropriate to trust you (in that respect, in that circumstance). (I will henceforth drop the parenthetical qualifications.) In short, you're trustworthy if you merit the emotion of trust. Or to put in a way that makes explicit the underlying connection between reliability and trust to which I am calling attention, you're *trustworthy* if it's appropriate to *rely* on you. You might merit trust even if you're more likely to fail than succeed. (Turri 2015: 541)

For example, if a doctor like House can discern the correct diagnosis about 30% of the time, while other doctors only get it right 15% of the time, then "it is eminently

reasonable to trust the diagnostician who gets it right thirty percent of the time" (Turri 2015: 542). Likewise for Turri, Ted Williams merits our trust even though he does not get a hit most of the time.

Turri believes that accounts of knowledge that only appeal to the truth-conducive sense of reliability[21] fail to capture the *binary* nature of knowledge. That is, they fail to capture that knowledge can come from *either* reliable or unreliable (but trustworthy) sources. Turri defines ecumenical reliabilism to allow for either sense of reliability: "knowledge must proceed from either truth-conducive or trustworthy abilities" (Turri 2015: 542). He further simplifies his account of knowledge by saying that "if we suppose that truth-conducive abilities merit trust, then … ecumenical reliabilism could be simplified: knowledge must proceed from trustworthy abilities" (Turri 2015: 542).

I believe that there is something intuitively right about Turri's ecumenical reliabilism. Turri is right that we should *trust* House to get the correct diagnosis and for Williams to get a hit even though they are unreliable. They both have some ability to achieve their aims. But I do not think that such trustworthy abilities render knowledge (for all the reasons given in the previous two sections and next two sections). Instead I believe that what adequately captures the intuitive pull to attribute knowledge rendering capabilities to trustworthy abilities has to do with such processes having *epistemically valuable consequences over rival processes*. Truth-conducive processes are straightforwardly epistemically valuable because they arrive at true beliefs at a rate significantly greater than chance, without comparison to other processes – i.e. they are *non-comparatively reliable*. In contrast, trustworthy abilities are epistemically valuable

[21] See fn. 1 above.

because they have a high *comparative reliability*. A process is comparatively reliable when it is among the best available options for getting at certain results (e.g. true beliefs) because it is significantly more reliable than other available (non-truth-conducive) processes. All things considered, a life-saving surgical procedure that only has a 40% chance of successfully curing disease X is to be trusted over another surgical procedure that only has a 10% chance of success of curing disease X because the former is comparatively reliable. Similarly, if only one in a hundred pharmaceutical drugs tested on mice is approved for human trials, this process is epistemically valuable because it renders more true beliefs than many other available options. Indeed, while many scientific practices of discovery are only comparatively reliable, they have epistemic value because they are able render true beliefs at a higher rate than other available processes.[22] Likewise, we should trust House to get the right diagnosis because his comparative reliability makes him epistemically valuable.[23] In short, while Turri's account fails to capture knowledge, it's emphasis on trusting processes does help to explain why certain unreliable processes are nonetheless epistemically valuable.

5. Against Turri's Experimental Evidence for Abilism

Turri (2015a, 2015b, 2016, forthcoming) also labels his account of (un)reliable knowledge as *abilism* instead of *ecumenical reliabilism.* While there are small differences

[22] I am limiting 'available' processes to include only processes that are currently known, economically feasible, and ethically permissible. See Harris (2017: 82-92) for the prospect of replacing mice in biomedical research.

[23] Indeed, Turri says as much when he writes, "Ted Williams merits trust when he is at the plate. He might not get a hit most of the time, but he is the *best option*, so it's appropriate to trust him, to reply on him. House merits trust when he is on the case. He might not get the diagnosis right most of the time, but he is the *best option*, so it's appropriate to trust him, to rely on him" (Turri 2015: 541, my emphasis).

between these accounts, they can safely be considered together since they both hold that unreliable knowledge can occur when agents successfully achieve their desired ends *through* their abilities. It is this claim that will be the focus of the rest of this chapter.

And while I have argued that the theoretical motivation for such a view has been eliminated, Turri (2016) supports abilism by providing some empirical data from experiments he conducted which seemingly show that unreliably produced knowledge is compatible with our folk intuitions about knowledge. If true, these findings would offer powerful support to abilism and the possibility of unreliable knowledge, and for this reason it is worth examining them in detail. After explaining abilism and how Turri's experiments seem to support the possibility of unreliable knowledge, I argue that Turri's experimental evidence is unconvincing and does not support abilism or unreliable knowledge.

Turri defines *abilism* in the following ways:

> Abilism defines knowledge as true belief manifesting the agent's cognitive ability or powers (Turri 2016: 225);
>
> Knowledge is approximately true thin belief manifesting *cognitive ability* (Turri 2015a: 321; and 2015b: 164);
>
> Knowledge is an accurate representation produced by cognitive ability (Turri: forthcoming).[24]

Turri frames his account to be about one's cognitive abilities "producing" or "manifesting" true beliefs. This terminology serves to explain why certain unreliable processes can render knowledge. Turri (2016a) takes the following example from Sosa

[24] This last definition indicates that Turri now believes that neither belief nor truth is a necessary condition for knowledge. I will ignore these aspects of Turri's account in this chapter.

(2007) to elucidate these concepts: An archer hitting a bullseye manifests her athletic ability only when her hitting the bullseye is *based on* or *the result of* or *because of* her abilities. If a gust of an unexpected wind interferes with the arrow's path and causes the arrow to hit the bullseye, then the bullseye was not a result of the archer's abilities. But unlike Sosa, Turri does not require that our cognitive abilities be reliable (see §1). This also fits with his account of achievements explained in §1 above: Achievements involve simply attaining one's intended outcome *through* one's (un)reliable ability. In my own words, Turri holds that S knows or intellectually achieves P iff P is true, and S believing P is the result of or manifests S's (un)reliable cognitive abilities.

Turri (2016) presents nine experiments that, he argues, support abilism and show that unreliably produced knowledge is compatible with our everyday folk intuitions about knowledge. While much can be said in response to each of these experiments, I will only focus Turri's seventh experiment because Turri places significant weight on this experiment's results in supporting abilism and showing that our folk intuitions allow for unreliable knowledge.

Turri's seventh experiment took 226 participants "aged 18-72 years, mean age = 32 years; 97% reporting English as a native language; 89 female" and "randomly assigned them to one of three groups, Control/Ability/Guess" (Turri 2016: 210). Each of these groups read a passage where an agent named Carolyn has been examined by cognitive scientists to see how well she can detect words that are flashed quickly on a screen. In all groups, the passage says that Carolyn has a "special ability to detect words flashed" (Turri 2016: 207) on a screen for only 120 milliseconds whereas most people can only correctly report the right word that is flashed on the screen less than 1% of the

time. In the Ability group the passage says that when Carolyn exercises her special ability she has a 30% track-record of getting the right answer. In the Guess group the passage says that when Carolyn does not exercise her special ability she has a 30% track-record of getting the answer right. And in the Control group the passage does not indicate whether Carolyn used her ability or guessed when she gets the right answer. All versions of the passage say that Carolyn took the word flash test and Carolyn correctly identified the word as "Corn." After reading their version of the passage, participants were given either of the following statements "Carolyn ____ that the word is 'Corn' (knows/only thinks)" (Turri 2016: 207) and asked to rate on a scale from 1-6 (where 1 is "not at all confident" and 6 is "completely confident") whether they agreed with that statement.

The results show that, in both Control and Ability, participants were significantly more likely to attribute knowledge to Carolyn (note: they nearly matched) while participants in Guess were significantly disinclined to attribute knowledge to Carolyn. Turri argues that these results support abilism for three reasons: First, these results show that our folk intuitions allow for unreliable knowledge since most participants readily attributed knowledge to Carolyn despite her unreliable ability; second, these results show that our knowledge intuitions are keyed to knowledge produced through *ability* since, despite having the same degree of reliability in Ability and Guess (i.e. 30%), most participants attributed knowledge to Carolyn when her ability *produced* her true belief but did not when she merely guessed; third, this experiment shows that "most people do not share the intuition that reliability specifically is what rules out problematic luck" (Turri 2016: 213) which Turri cites as the main motivation for requiring reliability as a

necessary condition for knowledge.[25] Turri concludes that these "results were exactly as we would expect if abilism adequately captured the ordinary conception of knowledge" (Turri 2016: 213).

Against Turri's Empirical Results and Suggestive Carolyn Example

I believe this experiment does not convincingly support abilism or that unreliable knowledge is compatible with our folk intuitions because there is an alternative explanation that better explains these results without supporting abilism or the intuitive possibility of unreliable knowledge. This alternative explanation comes from Sripada and Stanley (2012: 7) who argue that the folk often mistake true belief for knowledge because "in many contexts, all we care about is that an agent had a true belief…[and] in such contexts, we don't care about the justificatory or basing requirements of knowledge." While this alternative explanation is not directed at Turri's experiments, if true this would undermine Turri's explanation of these results since on this explanation the folk would only be looking to whether agents like Carolyn had true beliefs rather than considering their (un)reliability or justification. Turri cites this alternative explanation and responds that:

> [T]he finding from Experiment 7 suggest a … more charitable explanation of any such tendency [to treat knowledge as equivalent to true belief]: when someone forms a true belief and has a cognitive ability relevant to acquiring truth, the default assumption is that she formed the true belief *through* ability and thus knows. (Turri 2016: 225, my emphasis)

[25] I will postpone further discussion of this third point concerning luck until the following section of this chapter since my argument against this supposed feature of abilism helps constitute my general argument against the possibility of unreliable knowledge.

In order to break this stalemate, I argue that there are two general ways in which Turri's Carolyn case is misleading, which show that this experimental evidence does not convincingly support abilism or the intuitiveness of unreliable knowledge.

First, the Carolyn case is unconvincing because it characterizes Carolyn's ability to recognize words flashed quickly on a screen with misleading language. The phrase "special ability" occurs in both the general description of Carolyn's abilities that the groups (Control/Ability/Guess) see and it additionally appears in the Ability group's version of the case as Carolyn "was exercising her special ability when she got it right" (Turri 2016: 207). It is also hypothesized in the case that "scientists think [Carolyn's special ability] has something to do with a *unique* feature of her optical nerves" (Turri 2016: 207, my emphasis). Furthermore, "astonishingly" is used to characterize Carolyn's ability over others to achieve the right answer in all versions of this case. While I do not deny that Carolyn's ability is "special" in the sense that she is exceptionally better than others at discerning the right word, these characterizations are problematic because this language could bias study participants into thinking that Carolyn has more knowledge than is warranted by the description of her case. After all, why would the narrator repeatedly characterize her ability as "special", "unique", and "astounding" if this is not important to the case? Sripada and Stanley (2012: 9) call this well documented biasing effect "the problem of narrator cues" and I think this biasing effect applies to the Carolyn example.[26]

[26] An anonymous reviewer helpfully suggested that Turri's use of "knows/only thinks" in his experiments might also bias some respondents because the 'only thinks' might suggest that some target proposition is of unknown truth value (unknown to themselves and the person being evaluated) or false. In other words, the pragmatics associated with 'She only thinks ...' might be creating some noise that would incline some to choose 'knows' when forced to choose among these options.

Second, the Carolyn example is misleading because it contains some of the problematic features that Turri's other examples (i.e. Ted Williams and House) have. For instance, the Carolyn example fails to take into account the distorting influence of difficult circumstances on one's reliability. Like Ted Williams' batting average in the major leagues or my bike-riding reliability on a snowy mountain, it is misleading to use Carolyn's track-record in difficult circumstances to establish her (un)reliability. Turri could object that because we are considering Carolyn's ability to identify words in difficult circumstances her track-record in these circumstances is the relevant indicator of her reliability. In response, I do not deny Carolyn is unreliable in these difficult circumstances, but my point here is that the Carolyn example does not take into account the distorting influence that difficult circumstances can have on *determining* one's reliability.

Indeed it seems the plausibility of Carolyn having unreliable knowledge depends on her being in circumstances *any* human agent would find difficult. To see why, consider if we change the Carolyn example such that her track-record of recognizing words is unreliable in more normal circumstances for the average agent. In this case, Carolyn can only accurately identify words flashed on a screen for a second or half-second 30% of the time (say because of poor eyesight or slow cognitive functioning), while most people are able to get the word right 90% of the time. This version of the example preserves Turri's abilism since Carolyn still has *some ability* to get at the right answer, but now she is unreliable in comparison to others in more normal circumstances. But intuitively in this case Carolyn fails to know the word she correctly identifies on the screen because she is unreliable. In short, characterizing Carolyn as getting it right more

often than most humans by using her "special ability" in such difficult circumstances likely biased study participants in to attributing knowledge to her.

This last argument indicates that that the Carolyn case is also problematic in the same way that Turri's House example is. Both Carolyn and House are in very difficult circumstances for any human agent and both are characterized as having incredible abilities to achieve things that the average person cannot. And like House, you might think Carolyn's special ability gives her the remarkable ability to identify and know the correct word more often than the average person. But as was explained in §3 with the Jessica example, this mischaracterizes House's abilities, and the same applies to Carolyn's abilities. Carolyn is remarkable at identify the right word not because she knows the correct word more often than the average person, but because she has a remarkable ability to propose novel words that are worthy of consideration. And like House, Carolyn does not know that her hypotheses are correct until they are confirmed.

To illustrate these points, imagine that you watched Carolyn incorrectly answering "arrow" when the word flashed was 'archer', "barn" when the word flashed was 'yarn', and "fork" when the word flashed was 'ford' before finally answering "corn" when the word flashed was 'corn'. Imagine further that Carolyn is not told that her last answer is correct after she provides it. At this point does Carolyn know that her word is correct? Like House (and Jessica), it seems that Carolyn would only know that her answer is correct *after* it is confirmed.

Thus, Turri's Carolyn example is misleading and cannot convincingly support abilism or the possibility of unreliable knowledge. Lastly, my arguments against Turri's seventh experiment involving Carolyn also apply to Turri's (2016) sixth, eighth, and

ninth experiments because all of these versions involve Carolyn and the misleading features of this case. Additionally, I believe my arguments also serve to indirectly cast doubt on Turri's first five experiments since by undermining the plausibility of Turri's abilist explanation my arguments indirectly support Sripada and Stanley's alternative explanation - which can also explain the results of these five experiments. In short, I believe my arguments in this section (and the next) provide convincing reasons to doubt *all* of Turri's (2016) experimental results.

6. Why Reliability is a Necessary Condition for Knowledge

So far I have argued that Turri has not provided adequate reasons to reject the orthodox view that knowledge requires reliability. In this final section, I will provide some positive reasons for this orthodoxy. I will argue that Goldman's (1979) "master argument" that gives "powerful *prima facie* support" to the reliability condition for knowledge is correct (Goldman 2012: 4). To accomplish this I will first show how Turri's recent objection to this argument using the abilist account of knowledge is flawed; and second I will further bolster Goldman's argument by briefly showing how my arguments in the previous sections constitute additional reasons for why reliability must be a necessary condition for knowledge.

Goldman's (1979) argument can be formalized as

1. Archetypal cognitive processes that do not confer knowledge all share the feature of unreliability.
    a. E.g. "confused reasoning, wishful thinking, reliance on emotional attachment, mere hunch, or guesswork and hasty generalization"

2. Archetypal cognitive processes that do confer knowledge all share the feature of reliability.
    a. E.g. "Standard perceptual process, remembering, good reasoning, introspection"
3. Therefore, knowledge proceeds from the reliability of the process or processes that cause it. (IBE) (Goldman 1979: 345)[27]

In other words, the reliability of a process is what explains why the processes listed in (2a) do render knowledge and why the processes listed in (1a) do not. A natural way to understand the intuitive pull of this argument is that the reliability of a (cognitive) process explains why knowledge cannot be *lucky*. Indeed, one tempting argument to make against accounts of knowledge that allow for the possibility of unreliable knowledge is that such accounts would implausibly allow for lucky knowledge. Turri's account of knowledge seems especially vulnerable to this objection since it seems the novice archer who achieves a bullseye on her first try has beginner's luck even though she achieved the bullseye, in some sense, *through* her abilities. So, contra Turri, Goldman's argument provides good reason to think knowledge requires reliability.

In response, Turri argues that Goldman's argument is not persuasive because "the hypothesis that knowledge requires *ability*, reliable or not, can [also] explain the membership of these lists" (Turri 2016: 190). That is, Goldman's argument fails to

[27] Strictly speaking, Goldman's (1979) argument is not about knowledge requiring reliability but is intended to support process reliabilism – the view that a belief is *justified* just in case it is caused by a sufficiently reliable process. But since Goldman assumes that justification is necessary for knowledge, this argument also supports the claim that reliability necessary for *knowledge*. And as Turri (2016: fn. 1) points out, one need not be committed to this connection between reliability and justification in order for this argument to support the reliability condition for knowledge. See Kornblith (2008) for an argument against justification as a necessary condition for knowledge.

provide positive reasons to think that reliability, rather than *un*reliability *through one's ability*, is required for knowledge because when agents use the processes in (1a) they have *no ability* to achieve true beliefs, while the processes in (2a) give agents the ability to achieve true beliefs. And, if the intuitive pull of Goldman's argument comes from reliability forbidding lucky knowledge, then this argument just equates unreliable processes with lucky processes since the processes that fail to confer knowledge listed in (1a) are all examples of lucky processes. Indeed, Turri agrees that lucky knowledge is implausible and denies that abilism allows for lucky knowledge because

> [t]he fact that someone cannot reliably produce an outcome does not entail that it's "just luck" when she does produce it. Unreliable performers usually still have *some* ability or power to produce the relevant outcome. *Unreliability* does not equal *inability*. (Turri 2015: 533)

While Turri does not explicate the different kinds of luck at issue here,[28] the ideas are clear enough to be intuitively compelling. The novice archer who hits the bullseye through their unreliable abilities (e.g. through effort and concentration) does not succeed just by luck; while the archer who hits the bullseye because of a gust of wind does succeed by luck. Likewise for Turri, intellectual achievements that issue from one's unreliable cognitive abilities are not lucky in the way that achieving a true belief through, say, guessing is lucky. Despite his poor track-record, when House correctly diagnoses a patient through his great diagnostic ability he does so in a way that an avid fan of *House M.D.* does not when they guess the correct diagnoses. Because many unreliable processes

[28] See Prichard (2005) for an analysis of the different kinds of epistemic luck.

manifest one's ability while lucky processes do not, Turri argues that Goldman's argument fails because it equates all unreliable processes with lucky processes.

I agree with Turri that unreliability does not equal inability and that we should not think that all unreliable processes are just lucky processes. To deny these claims is to implausibly deny that there are nascent abilities. I also think that Turri is right that Goldman's argument fails to consider the alternative abilist hypothesis because it equates unreliable processes with lucky processes. However, *none of this helps Turri establish that unreliable knowledge is possible.* To see why, we can think of Turri as attempting to make the following argument:

1. Not all unreliable cognitive processes are lucky.
2. There are non-lucky but unreliable cognitive processes manifesting one's cognitive ability (i.e. there are nascent cognitive abilities).
3. Some of the processes in (2) can render knowledge.
4. Thus, unreliable knowledge is possible.

The key premise is premise three.[29] But as the previous sections above show, Turri has not provided adequate support for this premise since his two theoretical arguments, experimental results, and all of his examples of unreliable process conferring knowledge have been shown to be faulty. Indeed, my arguments in the previous sections serve to

[29] For various theoretical reasons, many authors would deny this premise. For example, Pritchard (2012) and Kelp (2013) argue that an ability condition on knowledge should be combined with a safety condition in order to deal with counterexamples typically leveled against ability conditions (e.g. fake barn cases). And since safety is a kind of reliability condition, Turri is presumably committed to rejecting these accounts of knowledge. Others have suggested that when the ability condition on knowledge is properly unpacked it entails an anti-luck condition precisely because it entails a reliability condition. For example, Sosa (2015), Carter (2016), and Beddor and Pavese (forthcoming) have suggested, on different grounds, that the best version of a cognitive ability condition for knowledge entails a safety condition. Presumably Turri is also committed to rejecting all of these arguments. My argument below undermines Turri's third premise directly via novel counterexamples without relying on any of these theoretical considerations.

both undermine premise three of Turri's argument and to bolster the premises of Goldman's (1979) argument. My arguments show that Turri's examples involving one's unreliable ability (e.g. Ted Williams, House, Carolyn) are either misleading and cannot convincingly show that such processes render knowledge or they should be included in the list of processes that do not render knowledge. Either way, the failures of Turri's examples help to further establish that reliability must be a necessary condition for knowledge. My arguments also support the reliability condition for knowledge by decisively undermining abilism. Once properly understood, the House, Jessica, and Carolyn examples serve as counterexamples to abilism because they each show that one can have a true belief that manifests or is the result of one's unreliable cognitive abilities without having knowledge.

To further illustrate these points against premise three, I will now explicitly show how the Jessica example in §3 is one instance of someone who fits Turri's definition of abilist/unreliable knowledge but intuitively fails to have knowledge. Jessica's true belief that the blurry picture in front of her is of a basketball is the result or manifestation of her unreliable cognitive ability to recognize such images (25%) but she fails to have knowledge until she is told her belief is true. Premise three is false because counterexamples like this can be generalized to show that unreliable/abilist knowledge is impossible. In short, I argue that this unreliable/abilist knowledge is impossible because any agent that is in a sufficiently favorable epistemic position to have unreliable/abilist knowledge will fail to have knowledge. And as was shown in §3, Jessica is in such a sufficiently favorable epistemic position for unreliable/abilist knowledge but she intuitively fails to have knowledge.

One might object that Jessica is not in a sufficiently favorable epistemic position to have unreliable/abilist knowledge. Firstly, an objector could argue that knowledge can be unreliably achieved only above some threshold of unreliability (e.g. above 40%). So, while Jessica is very reliable in comparison to her peers, she still only has 25% reliability and falls below this threshold for unreliable knowledge. Additionally, one could object that our intuitions about the Jessica case may be compromised by the fact that Jessica's unreliability is caused by her sub-par eyesight or malfunctioning ability to see. Indeed, what makes the House case compelling is that House's unreliability is not caused by a sub-par or malfunctioning ability (since he is the best of the best) but because of the difficulty of his job – i.e. diagnosing unusual patients. So, for these reasons one could argue that the Jessica case is not a convincing counterexample to abilism and the possibility of unreliable knowledge.

In response, I claim that additional examples can be constructed to avoid these pitfalls that nevertheless show that unreliable/abilist knowledge is impossible:

> Ashley is a professional singer. While Ashley does not have perfect pitch, after many years of studying, practicing, and performing she has gained some ability to accurately identify notes played on a piano. Specifically, Ashley is able to accurately identify what single note is played by listening alone with a 50% average accuracy. In contrast, the average lay person is almost never able to correctly identify the right note since they have no ability to recognize which of the 12 possible notes is played. Those with perfect pitch are able to recognize which note is played with near 100% accuracy. Imagine that you are watching Ashley practice her ability over the period of half an hour. In this time you see her

> correctly identify what note is played on average 50% of the time. Furthermore you notice that when Ashley is wrong she is never more than a musical half-step from the right answer (e.g. if the answer is A#, Ashley answers A; or if the answer is F, Ashley answers E).

Unlike Jessica, Ashley is much more reliable at 50% and, like House, does not have a sub-par or malfunctioning ability. You could say that she *nearly* has perfect pitch since her answers indicate that even when she is wrong she is still tracking the correct pitch. But even with this great ability to identify pitches by auditory means alone, imagine that Ashley is played a Db note on a piano and correctly answers Db, but is not yet told that her answer is correct. At this point, does Ashley *know* that the note is a Db? Intuitively, Ashley does *not* know the answer is Db, and I contend the only explanation for this intuition is that despite her nascent perfect pitch ability she is still unreliable at identifying pitches. Thus, abilism is false because examples like this show that one can have a true belief that manifests one's unreliable cognitive abilities without having knowledge.

So, to reiterate, examples like this also show that unreliable knowledge is impossible since such agents are in sufficiently favorable epistemic conditions to have this kind of knowledge, but intuitively still fail to have knowledge. Furthermore, I contend that many more examples can be constructed to support the intuition that unreliable agents like Jessica and Ashley fail to have knowledge. In summary, I am making the following argument:

1. If those in sufficiently favorable epistemic positions to have unreliable/abilist knowledge fail to have knowledge, then unreliable abilist/knowledge is impossible.
2. Ashley, Jessica, etc., are in sufficiently favorable epistemic positions to have unreliable/abilist knowledge but fail to have knowledge.
3. Thus, unreliable/abilist knowledge is impossible.

In conclusion, Turri has not established that unreliable knowledge is possible and there are decisive reasons for thinking knowledge requires reliability.

CHAPTER 2

THE LIMITS OF INTUITION SKEPTICISM, THE METHOD OF CASES, AND SELF-DEFEAT

It is widely believed that "intuitions" and "the method of cases" are central to philosophical inquiry. This belief reflects the long tradition going back to at least Socrates of philosophers appealing to thought-experiments (aka hypothetical cases or scenarios) to "test" certain philosophical theses and theories against our intuitive judgments of these cases. If, after considering a thought-experiment, our intuitions rule in favor of a philosophical claim, this is often considered support or evidence for this thesis. And if our intuitions do not rule in favor of a philosophical claim this is often considered support or evidence against this thesis.

Examples of this practice are ubiquitous in contemporary philosophy. For example, consider the large literature of different accounts of knowledge being tested against various Gettier-cases, principles regarding free will being tested against various Frankfurt-cases, different moral principles being tested against various trolley-cases, etc.[1] In each instance, the method of cases is supposed to be roughly analogous to the scientific practice of testing scientific theories against observations of the world via laboratory experiments. Indeed, in the same way that scientific claims are justified on the

[1] In this chapter I will set aside the challenges to this characterization of philosophy's methodology (e.g. Capellan (2012) and Deutsch (2010)). See Devitt (2015) for a critical response to these challenges. Furthermore, I will assume along with the authors to be discussed below (Silva and Machery) that intuitions are treated as evidence in philosophy. Cf. Earlenbaugh and Molyneux (2009).

basis of observations in laboratory experiments, many philosophical claims are often thought to be justified largely, if not exclusively, on the basis of intuitive judgments about thought-experiments.

In recent years there has emerged a new discipline in philosophy called experimental philosophy (aka x-phi) that has, among other things, challenged the legitimacy of this philosophical practice. Experimental philosophers have used the methods of empirical psychology to see whether philosophers' intuitive judgments about the above, and other philosophical cases, are widely shared amongst various demographic groups or are significantly influenced by such presentation factors as framing and ordering effects. By and large, the results of these x-phi studies are that some of the most cited thought-experiments in philosophy do not illicit the same intuitive judgements from various demographic groups or that such intuitions are easily manipulated by presentation factors.[2] As such, these x-phi studies seem to provide a strong case that, unlike observations in science, intuitions about cases in philosophy are not a reliable guide to the truth. Consequently, many experimental philosophers have argued for some form of philosophical skepticism: many, if not all, philosophical propositions are either not justified or not known. In other words, it does not seem that intuitive judgements about cases in philosophy can legitimately play the evidentiary role that observations in laboratory experiments do in science.

This chapter addresses the scope and limits of such Unreliability Arguments[3] for philosophical skepticism. Ultimately, I will argue that such arguments are severely

[2] See Machery (2017: chap 2) for an extensive review of these empirical findings.

[3] I am following Pust (2012), Silva (2013), Machery (2017), and others in using this label.

limited in scope by the threat of epistemic self-defeat. An argument is epistemically self-defeating *iff* at least one of this argument's components (i.e. premises, conclusions, or reasoning) applies to and epistemically defeats itself or another component. In the present case, Unreliability Arguments threaten to defeat themselves if they rely on intuitions about philosophical cases.

While experimental philosophers are keenly aware of this danger, they either respond to this threat directly or take measures to avoid this threat by limiting the scope of their philosophical skepticism. Examples of the former include Silva (2013) who argues that epistemic self-defeat is not a threat to an Unreliability Argument for a global kind of intuition skepticism. Examples of the latter include Machery (2017) who argues that the above experimental data supports an Unreliability Argument for a near-global skepticism against *philosophers' use* of the method of cases, which itself (allegedly) does not rely on intuitions about cases. The aim of this chapter is to show that the self-defeat challenge is not as easily dealt with or avoided as the above philosophers would have us believe, and that this challenge places severe limits on the kind of intuition or philosophical skepticism that an Unreliability Argument can establish. In short, I argue that the power and scope of the self-defeat challenge has been underappreciated in the x-phi literature.

I begin in §1 by explaining Silva's (2013) extreme version of the Unreliability Argument for a global kind of intuition skepticism. §2 shows that Silva's argument is mistaken because when the epistemic self-defeat challenge is properly understood it establishes that such a global intuition skepticism is not only unjustified but false. §3 explains Machery's (2017) near-global Unreliability Argument for philosophers' use of

the method of cases, and §4 argues that Machery's argument is self-defeating despite limiting his argument to only philosophers' use of the method of cases. And in §5 I conclude by explaining what I take the upshot of this chapter to be: that Unreliability Arguments for philosophical skepticism are severely limited by the threat of self-defeat, and so philosophical knowledge is not substantially threatened by such arguments.

1. Silva's Unreliability Argument(s)

It should be emphasized at the outset, that I am not the first to raise the self-defeat challenge to arguments for intuition skepticism – the view that philosophical propositions cannot be justified or known on the basis of intuitive judgements. For example, Bealer (1992), BonJour (1998), and Pust (2001, 2019) have defended intuitions by appealing to epistemic self-defeat. Among other things, they each argue that no global version of intuition skepticism can coherently be made since any such argument must, in some way, rely on intuitions.

Pace these authors, Silva (2013) argues that there is a global version of the Unreliability Argument that provides sufficient reason to believe that all intuitions are illegitimate sources of justification *while remaining epistemically self-defeating*. That is, Silva argues that *even if* the global Unreliability Argument is epistemically self-defeating, appealing to epistemic self-defeat does not undermine its conclusion that all intuitions are illegitimate sources of justification. If Silva is right then global intuition skepticism is a coherent position that potentially threatens most of philosophical discourse. Contra Silva, I argue that epistemic self-defeat does defeat the global Unreliability Argument when epistemic self-defeat is properly understood. In arguing for this, I will come to the result

that there logically must be some intuitions that are a reliable source of evidence that can justify our beliefs. I begin by explicating Silva's argument.

Silva defines intuition very broadly as "a contentful mental state wherein a proposition seems true, independently of introspection, sense perception, memory, testimony, and inference" (Silva 2013: 582). The purpose of such a broad definition is to avoid the controversies surrounding the nature of intuitions. I will follow suit and accept this definition in this chapter. Additionally, I will follow Silva in his definition of justification as *ultima facie* justification: a proposition is ultima facie justified just in case the proposition is justified and undefeated.[4] With these stipulations, Silva advances his Unreliability Argument against legitimacy of intuitions as a source of justification. The argument states that for a source of evidence to justify belief it is a necessary condition that this source be reliable. Formalized, this condition is $(Sx \rightarrow Rx)$[5] where 'S' is "source of evidence that justifies beliefs" and "R" is "reliable source of evidence." Next, the argument claims that intuitions are an unreliable source of evidence (~Ri), and concludes that intuitions are not a source of evidence that justifies belief (~Si) – where "i" means "intuition." Silva renders the Unreliability Argument as:

1. $(Sx \rightarrow Rx)$
2. (~Ri)
3. Thus, (~Si)

Next, Silva explains that intuition defenders like Bealer (1992), BonJour (1998), and Pust (2001) will claim that this argument is epistemically self-defeating. For, if the

[4] See Senor (1996) for more on the distinction between ultima facie and prima facie justification.

[5] I am following Silva (2013) in not including the universal quantifier in this formalism.

conclusion is granted that all intuitions are not a source of evidence that justify beliefs then since the first premise of the argument is seemingly justified by intuition, then the first premise is an illegitimate source of justification for the conclusion. That is, it is plausibly claimed by the defenders of intuition that this conclusion provides its own defeater since the first premise will need to be justified and the only apparent source of justification is intuition.[6] Therefore, the intuition defender concludes that the proposition that intuition is not a source of evidence that justifies beliefs is itself unjustified. I will call this argument the *First Self-Defeat Argument* which Silva renders as:

(1) If (~Si) is justified, then (Sx→Rx) is not justified.

(2) (Sx→Rx) is justified. (Dialectical Assumption, explained below)

(3) Thus, (~Si) is not justified.

To provide further clarification of the First Self-Defeat argument, Silva explains that intuition defenders are making what he calls a "Dialectical Assumption" (henceforth (DA)). That is, they assume (for reductio) that "the premises of the Unreliability Argument are both true and justified" (Silva 2013: 583) at the outset of this argument. And then, the intuition defenders use the (DA) to argue that the Unreliability Argument is epistemically self-defeating. It is important to note again that *ultima facie* justification is the notion of justification used in the (DA) and the First Self-Defeat Argument. The intuition defender wants to establish that (~Si) is not ultima facie justified because it has been defeated by itself.

Silva then explains the problem with this reasoning is that:

[6] As Silva (2013: 580) explains, the intuition defender can justify this claim in several different ways. For example, Silva explains that Bealer argues that intuitions are necessary to make basic epistemic classifications, and Pust argues that only intuition can justify epistemic principles.

> defenders of intuition believe that we are justified in thinking that intuition *is* a source of evidence that justifies beliefs, i.e. (Si) is justified. But this claim can be justified only if we lack justification to believe that intuition is unreliable. Thus, *if* there is sufficient reason to think (~Ri) true, then (Si) cannot be justified and the fact that the Unreliability Argument is epistemically self-defeating fails to show otherwise. (Silva 2013: 584)

Essentially, Silva claims that the (DA) that the intuition defenders used to undermine the Unreliability Argument in the First Self-Defeat Argument can be turned around and used against the intuition defender's belief that (Si) is true. That is, according to Silva the *mere assumption* of the (DA) by the intuition defender is enough to show that (Si) is not justified – i.e. all intuitions are an illegitimate source of justification/evidence.

To make this argument clearer, Silva provides what I will call the *First Anti-Self-Defeat Argument*:

(5) (Sx→Rx) and (~Ri) are each justified. ((DA) for conditional proof).[7]

(6) If (Sx→Rx) and (~Ri) are justified and one completely deduces (~Si) from (Sx→Rx) and (~Ri), then (Si) is not justified. (Defended below)

(7) Thus, (Si) is not justified. (from (5) and (6))

Silva then discharges the conditional proof and argues that *if* there is sufficient reason to believe that the (DA) (i.e. (5)) is true, then it follows that (Si) is not justified. That is,

(8) If (Sx→Rx) and (~Ri) are justified, then (Si) is not justified. (conditional reasoning from (5)-(7))

(9) Thus, (Si) is not justified. (If the (DA) is correct, and (8)) (Silva 2013: 585)

[7] Skipping premise "(4)" is intentional in order to follow Silva's numbering. In Silva's paper he explains that the intuition defender could argue for an additional premise that "If (Si) is justified, then (~Si) is not justified." I leave this digression aside since it does not affect either Silva's argument or my response.

As Silva notes, defending premise (6) is crucial for this argument to go through. Silva defends this crucial premise by argument for two principles: "(i) the assumption that the contradictory of any justified proposition is itself unjustified, and (ii) a reasonable deductive closure principle for justification" (Silva 2013: 585). I will not further explicate Silva's defense of premise (6) or these principles since my response to this argument in the following section will not dispute this premise. Additionally, it would not be prudent to dwell on this premise or its defense since Silva provides what I will call the *Second Anti-Self-Defeat Argument* that does not rely on (6) or its principles.[8] This second argument strengthens Silva's case that the mere assumption of the (DA) is enough to show that (Si) is not justified. As he claims, the "(DA) poses a direct threat [to (Si)] *if*, as I keep emphasizing, there is sufficient reason to think [(~Ri)] is true" (Silva 2013: 587).

This Second Anti-Self-Defeat Argument, like the First Anti-Self-Defeat Argument, relies on the (DA). However, this second argument only requires that the (DA) assume the second premise of the Unreliability Argument is justified; namely, (~Ri) is justified. Silva argues that assuming (~Ri) is justified, without assuming (Sx→Rx) is justified, is all that is needed to show that (Si) is not justified. This is because Silva claims that it is generally accepted (especially by intuition defenders) that if something is an unreliable source of evidence, then this source of evidence does not justify beliefs. To claim otherwise, according to Silva, amounts to claiming that something can be unreliable, while still being a legitimate source of evidence that justifies belief. On its face, it seems that if you are justified in believing that some source is unreliable, then this *undercuts* the belief that this source can provide evidence that

[8] It also does not depend on conditional proof or on (Sx→Rx). This is explained below.

justifies your beliefs.[9] For example, if you are justified in believing that some newspaper is unreliable in reporting the facts of world events, then this newspaper cannot be a source of evidence that can justify your beliefs about world events. The reliability of one's source directly affects the justificatory status of that source. "[J]ustification for thinking a source is unreliable is enough to prevent that source from generating (undefeated) justification, and thereby prevents that source from justifying beliefs" (Silva 2013: 587). I render the *Second Anti-Self-Defeat Argument* as:

(10) If (~Ri) is justified, then (Si) is not justified.

(11) (~Ri) is justified. ((DA))

(12) Thus, (Si) is not justified. (This is the same as (9))

To reiterate, both of Silva's Anti-Self-Defeat Arguments rely on the (DA). Essentially, Silva is turning around and using the (DA) that the intuition defenders used to undermine the Unreliability Argument in the First Self-Defeat Argument against the intuition defender's belief that (Si) is true. If Silva is right the mere assumption of the (DA) is enough to give sufficient reason to believe (~Ri) is true. That is, even if the Unreliability Argument is self-defeating, the assumption of the (DA) that intuition defenders make against intuition skeptics in the First Self-Defeat argument provides sufficient reason to believe that (~Ri) is true, and this is all the intuition skeptic needs to show that (Si) is not justified.

In sum, Silva has provided two arguments which show that if intuition defenders grant the assumption, for reductio, that intuitions are not a reliable source of evidence,

[9] For Silva (2013: fn. 21), "one has an *undercutting defeater* for *p* if one either believes or has reason to believe that their (supposed) source of justification for *p* does not make *p* likely to be true."

then this assumption can be used to against the intuition defender's reductio argument (i.e. the First Self-Defeat Argument) to show that intuitions are not a source of evidence that can justify our beliefs. Therefore, according to Silva, the self-defeat defense of intuitions does not block global skepticism about intuitions.

2. Against Silva's Unreliability Argument(s)

In this section I will show that both of Silva's Anti-Self-Defeat Arguments fail to undermine the intuition defender's reductio argument (i.e. the First Self-Defeat Argument). Contra Silva, the self-defeat defense of intuitions does refute the Unreliability Argument. In arguing for this I will show that global skepticism about intuitions is false and that there must be some intuitions that are a reliable source of evidence that justify our beliefs.

To begin, one possible response to Silva's Anti-Self-Defeat Arguments is to claim that his arguments are in some way based on intuition and so are self-defeating. For example, the unreliable newspaper example above is used to elicit intuitions to support premise (10). It is a virtue of Silva's arguments that this sort of reply will not be successful. The reason is because this reply essentially assumes the (DA) (or something like it) and as was explained above, the *mere assumption* of the (DA) is all that Silva needs to run his arguments. Again, Silva argues that *even if* the Unreliability Argument is self-defeating, it still provides sufficient reason to believe intuition is not a source that justifies beliefs provided that there is sufficient reason to believe that intuitions are unreliable. And the same goes for Silva's Anti-Self-Defeat Arguments. However, it should be understood that because the assumption of the (DA) is essential to both Anti-

Self-Defeat Arguments, a closer analysis of what goes into the (DA) is required in order to refute Silva's arguments.

To begin my analysis of the (DA), I would like to emphasize that the premises of the Unreliability Argument (i.e. (Sx→Rx) and (~Ri)) should be understood *globally*. That is, (~Ri) states that *all* intuitions are an unreliable source of evidence. The conclusion of the Unreliability Argument (i.e. (~Si)) should also be understood globally: *all* intuitions are not a source of evidence that can justify beliefs. This global understanding of the premises and conclusion of the Unreliability Argument was crucial in order for Silva to explain what I called the First Self-Defeat Argument that intuition defenders use against intuition skepticism. Understood globally, (~Si) was used in this argument to undermine the first premise of the Unreliability Argument (i.e. (Sx→Rx)) in order to show that global skepticism about intuition as a source of evidence that justifies belief is not justified.

With this emphasis in mind, it is further crucial to see that there are two ways that the intuition defender could make a self-defeat argument. The first way is to do what the First Self-Defeat Argument does above and attack the first premise of the Unreliability Argument. However, the intuition defender could also make a self-defeat argument to attack the second premise of the Unreliability Argument. Silva does not mention this alternate self-defeat argument, which I will call the *Second Self-Defeat Argument*.[10] I render this argument as:

(1*) If (~Si) is justified, then (~Ri) is not justified.

[10] It might be helpful for the reader to note that the First Self-Defeat Argument is so named because it attacks the first premise of the Unreliability Argument, and the Second Self-Defeat Argument is so named because it attacks the second premise of the Unreliability Argument.

(2*) (~Ri) is justified. (From (DA))

(3*) Thus, (~Si) is not justified.

This argument can be legitimately made if two conditions hold. The first is that, as explained above, (~Ri) must be understood globally. The second is the understanding that no evidence alone can justify the claim that all intuitions are an unreliable source of evidence; intuition must partly justify (~Ri). With these conditions in place the intuition defender can plausibly claim that (~Ri) is at least partially justified by intuition. The first condition seems uncontestable since understanding (~Ri) globally is crucial for Silva's argument. And denying the second condition is implausible if you accept the first condition. To argue otherwise would amount to claiming that one could infer from a finite amount of evidence that intuition is an unreliable source of evidence to *all* intuition is an unreliable source of evidence. But, it is commonly thought that such enumerative inductions are based on some Uniformity of Nature principle that itself derives its plausibility from intuition.[11] So, since it is very plausible that (~Ri) is at least partially justified by intuitions, if Silva wanted to contest this condition he would need to provide an argument to show that (~Ri) (understood globally) is not at least partially justified by intuitions.[12]

[11] I am of course alluding to Hume's famous "Problem of Induction." See Hume (1999).

[12] Additionally, it is not clear that Silva can provide such an argument if he wants to maintain his project described above. To see this, remember that the *mere assumption* of the (DA) is all that Silva claims he needs to run his Anti-Self-Defeat Arguments. His Anti-Self-Defeat Arguments intend to show that *even if* the Unreliability Argument is self-defeating appealing to epistemic self-defeat does not stop global intuition skepticism. But if he were to provide such an argument this would undermine his project of showing that *mere assumption* of the (DA) is all that is need to run his Anti-Self-Defeat Arguments. That is, by contesting this condition and providing reasons independent of intuitions to hold (~Ri) understood globally, it seems Silva would also thereby provide reasons that show the mere assumption of the (DA) is not all that is need to run his arguments. Furthermore, it is also not clear that Silva can contest this condition because this would seem to undermine his claim that *even if* the Unreliability Argument is self-defeating

So, given that both premises of the Unreliability Argument can be attacked using epistemic self-defeat by the intuition defender, it follows that the (DA) can be attacked by the intuition defender using epistemic self-defeat. This is because Silva's (DA) states that "the premises of the Unreliability Argument are both true and justified" (Silva 2013: 583). To be clear, this of course does not by itself show that Silva's (DA) is undermined. Silva's argument seems to be unaffected by this result because it seems he could plausibly run Anti-Self-Defeat Arguments that he made against the First Self-Defeat Argument against the Second Self-Defeat Argument.

Now that all of the preliminary clarifications and elaborations have been made, I can begin my response to Silva's argument. On the whole, I think Silva misunderstands both the First and Second Self-Defeat Arguments because when they are properly understood they undermine both parts of the (DA). That is, the Self-Defeat Arguments show that assuming the (DA) is illegitimate since assuming the (DA) is self-defeating and so the (DA) (understood globally) *cannot be true and cannot be ultima facie justified* (henceforth UFJ). More specifically, I argue below that an intuition denier cannot turn-around and use the (DA) against the intuition defender in either of the Anti-Self-Defeat Arguments since the (DA) *has already been defeated* by the previous Self-Defeat Arguments.

---

appealing to epistemic self-defeat does not stop global intuition skepticism. Given the way that Silva describes the (DA), the (DA) (and so (~Ri)) has to be partly justified by intuition in order for the Unreliability Argument to be epistemically self-defeating. That is, without the (DA) (i.e. the premises of the Unreliability Argument) at least partially justified by intuition, it is not clear how the argument could be epistemically self-defeating in the first place. Silva could avoid this problem by claiming that only the first premise of the Unreliability Argument (i.e. (Sx→Rx)) is justified by intuition. But, again as argued above since it is very plausible that (~Ri) understood globally is at least partially justified by intuition, Silva would have to provide some non-arbitrary reasons to treat (Sx→Rx) and (~Ri) differently. And, again, such reasons seem to undermine his project of showing that *mere assumption* of the (DA) is all that is need to run his Anti-Self-Defeat Arguments.

To see how the (DA) is undermined by the Self-Defeat Arguments, it will be helpful to see how each of the Self-Defeat Arguments should be rendered if we understand the (DA) as stating that both of the premises of the Unreliability Argument are true and UFJ. That is, by claiming that "both of the premises of the Unreliability Argument are true and justified" the (DA) is making four claims: (Sx→Rx) is true, (Sx→Rx) is UFJ, (~Ri) is true, and (~Ri) is UFJ. Putting both the Self-Defeat-Arguments side by side and including the global understanding of the Unreliability Argument's premises, we can render the arguments fully as:

First Self-Defeat Argument (Full)

(1) If (~Si)(global) is true and (~Si)(global) is UFJ, then (Sx→Rx) is not true and (Sx→Rx) is not UFJ.

(2) (Sx→Rx) is true and (Sx→Rx) is UFJ. ((DA))

(3) Thus, (~Si)(global) is not true and (~Si)(global) is not UFJ.

Second Self-Defeat Argument (Full)

(1*) If (~Si)(global) is true and (~Si)(global) is UFJ, then (~Ri)(global) is not true and (~Ri)(global) is not UFJ.

(2*) (~Ri)(global) is true and (~Ri)(global) is UFJ. ((DA))

(3*)Thus, (~Si)(global) is not true and (~Si)(global) is not UFJ.

In Silva's original formulation of the First Self-Defeat argument he only includes the word "justified" instead of providing the full version of the argument above. I believe this omission leads Silva to misunderstand this argument. For what is significant about the full version of these arguments is that they each individually attack the premises of the Unreliability Argument (and hence also the (DA)) but they also both conclude that *(~Si)(global) is not true and (Si)(global) is not UFJ*. This is significant because this means that above self-defeat arguments establish that the original Unreliability

Argument's conclusion (i.e. ~Si) both false and not justified by its premises. Therefore Silva cannot turn around and use the (DA) to support this conclusion. In other words, because the (DA) leads to a reductio of the Unreliability Argument via the above self-defeat arguments, the (DA) shows that Unreliability argument is self-defeating.

This is also significant because this result can be used by the intuition defender to prove that (~Ri)(global) is false. In other words, this result can show that the (DA) is defeated. If this is correct then Silva also cannot use the (DA) in his Anti-Self-Defeat Arguments, because the mere assumption of the (DA) will *not* provide sufficient reason to believe that (~Ri) is true. That is, if (~Ri)(global) can be proved false, then we will have undermined Silva's central claim that the "(DA) poses a direct threat *if*, as I keep emphasizing, there is sufficient reason to think [(~Ri)(global)] true" (Silva, p. 587).

How can (~Ri)(global) be proved false from the conclusion of both Self-Defeat Arguments above? Below is the first argument on the way to proving (~Ri)(global) is false:

a. If (~Si)(global) is not true, then (Si)(local) is true.

b. (~Si)(global) is not true. (From Self-Defeat Arguments).

c. Thus, (Si)(local) is true.

To be clear, the conclusion of this argument just means that at least one kind or instance of intuition is a source of evidence that can justify our beliefs. Why should an intuition skeptic believe premise a.? The justification for this premise lies in the claim that any contradictory of any false proposition is itself true.[13] That is, because (~Si)(global) is

[13] Recall that Silva (2013: 585) explicitly accepts a version of this principle in his article when he defends premise (6) of the First Anti-Self-Defeat Argument. His version of this principle is: (i) the assumption that the contradictory of any justified proposition is itself unjustified.

false then its contradictory must be true, which is to say that (Si)(local) is true. (~Si)(global) says that "No intuitions are sources of evidence that justify belief," and since this is false it follows that "Some intuitions are sources of evidence that justify belief" (i.e. (Si)(local)) must be true.

The result that (Si)(local) must be true is all the intuition defender needs to establish that both of Silva's Anti-Self-Defeat Arguments fail because (Si)(local) can be used to establish (~Ri)(global) is *false*. The intuition defender can undermine the First Anti-Self-Defeat Argument by arguing:

d. (Sx→Rx) is true. ((DA))

e. (Si)(local) is true. (Established above).

f. Thus, (Ri)(local) is true.

This conclusion amounts to the denial of (~Ri)(global). This conclusion says "Some intuitions are a reliable source of evidence" is true, which entails that "No intuitions are a reliable source of evidence" (i.e. (~Ri)(global)) is false because this is the conclusion's contradictory. The intuition defender can undermine the Second Anti-Self-Defeat argument in much the same way, except in this case they do not need the (DA) to assume (Sx→Rx) as in premise d. above. Instead, all the intuition defender has to do is argue that because (Si)(local) is true it follows, without the need of (Sx→Rx), that (~Ri)(global) is false. To see this, recall that premise (10) of the Second Anti-Self-Defeat Argument states that "If (~Ri) is justified, then (Si) is not justified." Also recall that this premise is one that Silva believes everyone will accepts since for example if some newspaper is not a reliable source of world news then that newspaper cannot be a source of evidence that justifies your beliefs. The contrapositive of this premise can be rendered as: If (Si) is

justified, then (~Ri) is not justified. Furthermore, we can replace "justified" with "true" in this premise without losing this principle's appeal.[14] With this replacement and using the global and local locutions, the intuition defender can use this contrapositive of (10) to argue against the Second Anti-Self-Defeat argument as follows:

g. If (Si)(local) is true, then (~Ri)(global) is not true. (contrapositive of (10))

h. (Si)(local) is true. (Established above)

i. Thus, (~Ri)(global) is not true.

What this all amounts to is an argument against Silva's use of the (DA) in his Anti-Self-Defeat arguments, or more pertinently against Silva's use of assuming (~Ri)(global). Silva misunderstands the Self-Defeat Arguments when he fails to include the global understanding of the premises of the Unreliability Argument and when he fails to see that when the intuition defender assumes these premises (for reductio) in order to run their Self-Defeat Arguments, they are assuming these premises are each *true and UFJ*. Having understood the conclusion of the Self-Defeat arguments in light of these considerations, the intuition defender can argue to the ultimate conclusion that (~Ri) global is not true. This effectively undermines Silva's claim that the "(DA) poses a direct threat *if*, as I keep emphasizing, there is sufficient reason to think [(~Ri)(global)] true" (Silva 2013: 587). Furthermore, as shown above, by undermining Silva's use of the (DA) that the proper use of the (DA) entails there *logically must* be some intuitions that are a reliable source of evidence that can justify our beliefs (i.e. (Si)(local) and (Ri)(local) are true).

[14] To be clear, I am not claiming that these premises are equivalent, all I am claiming is that replacing "justified" with "true" renders a principle that is also compelling and is unlikely to be challenged by intuition defenders or intuition deniers like Silva.

I would like to conclude this section by elaborating on the results that (Si)(local) and (Ri)(local) are true. I have nowhere above indicated what these reliable intuitions as a source of evidence that justify belief are. Although I will not provide a defense here, I think an argument can be made that the class of intuitions that are a reliable source of evidence that justifies belief must include our logical intuitions: for, if our logical conclusions were not included in this class then global skepticism about our intuitions might be true. My and Silva's arguments above depend on familiar logical rules and reasoning. If it were to be shown that logical intuitions are globally unreliable then my arguments above would fail and perhaps Silva's would too. However, I think that any such result is very unlikely to be true. But I and other intuition defenders can revel in the result that some intuitions must be reliable sources of evidence that justify belief and that global intuition skepticism is untenable.

And, as we will see in the next section, the fact that (Si)(local) and (Ri)(local) must be true is very problematic for Machery's (2017) near-global Unreliability Argument. In short, my arguments below show that certain epistemological intuitions about philosophical cases must be also reliable to avoid a similar kind of self-defeat.

3. Machery's Unreliability Argument

While Silva (2013) attempts to establish that a global kind of intuition skepticism is possible, Machery (2017) argues for a near-global kind of skepticism *for philosophers' use* of the method of cases. In brief, Machery argues that in order for philosophers to acquire the kind of knowledge they seek, they must appeal to a certain type of case to adjudicate between competing philosophical theses. But, Machery argues, the extensive x-phi literature shows that appealing to such cases has a strong tendency to illicit

different intuitive responses from non-philosophers. Therefore, since appealing to such cases elicits unreliable judgements, Machery concludes philosophers cannot attain the kind of knowledge they seek by appealing to this type of case. And because appealing to such cases is usually the only way for philosophers to acquire the knowledge they seek, an extensive philosophical skepticism follows. As such, Machery's target is both the kind of knowledge philosophers attempt to acquire, and the general type of cases that philosophers appeal to in their philosophical investigations to acquire this knowledge.

In more detail, Machery claims that philosophers typically attempt to acquire some *modal* knowledge about the *nature* of knowledge, reference, right action, responsibility, etc. This is reflected in the fact that philosophers typically state their theories in metaphysically necessary or metaphysically sufficient conditions. For example, hedonic theories of well-being roughly state that pleasure is both necessary and sufficient for a good life. And philosophers also typically evaluate such modal theories by posing thought-experiments to test these theories against our intuitive judgments. For example, many philosophers believe that such simple versions of hedonism are refuted by some version of Nozick's (1974) "experience machine" thought-experiment. These are cases where someone is given a choice to remain in the real world or a choice to "plug-in" to a virtual world for the rest of their life. In this version, the real world will be filled with the typical trials and tribulations of life, while the virtual world will be filled with immense pleasure. The choice is permanent and once it is made one will have no memory of being given or making this choice.[15] Because many philosophers intuitively judge that

[15] Nozick's (1974) original version of the case has agents unplug after two years and reevaluate whether they wanted to stay in or leave the experience machine.

a life lived in the experience machine is not a good life, they believe that such simple versions of hedonism are false.

Machery argues that the type of case philosophers typically appeal to involve what Machery calls "disturbing characteristics": they involve unusual hypothetical cases that pull apart what typically goes together. For instance, the experience machine is an unusual hypothetical case because this situation is, as far as we know, not actual and which we have never or infrequently encountered in our everyday lives and thinking.[16] And the experience machine pulls apart the properties of pleasure, well-being, and a good life – which typically co-occur in everyday life, e.g. gaining pleasure from helping one's friends and family achieve their goals intuitively contributes to a good life. Indeed, it is an essential part of this thought-experiment that it pulls these properties apart since philosophers want to adjudicate between competing philosophical theories and (allegedly) attain some modal knowledge, e.g. what is necessary or sufficient for a good life.

For instance, other theories of well-being (e.g. desire-satisfaction theories) typically agree with a hedonic version of the good-life when it comes to everyday cases. To adjudicate between these competing theories and test whether simple versions of hedonism are right when they claim that pleasure is *sufficient* for a good-life, philosophers typically must devise unusual hypothetical cases, like the experience machine, where these properties come apart. This philosophical practice is supposed to mirror scientific practice: because "natural occurring phenomena rarely discriminate

[16] "A situation or event (type) is unusual if and only if we encounter it infrequently or if we rarely read texts about it" (Machery 2017: 113) and "A situation or event is unusual if and only if it occurs infrequently" (Machery 2017: 117).

between existing scientific theories, and scientists are compelled to devise artificial experimental conditions about which existing scientific theories make different predictions" (Machery 2017: 118).

Machery challenges this philosophical practice by providing a systematic review of the x-phi literature that bears on this practice's reliability. Machery argues that this large literature shows that philosophers' use of the method of cases is unreliable since nearly all of the most cited philosophical cases of the past 80 years that are used to support or undermine philosophical theses are largely influenced by demographic effects (e.g. one's culture, age, gender) or presentation effects (e.g. the order or framing of the case). For example, this literature shows that there is a large variation in how non-philosophers respond to the experience machine thought-experiment. For demographic effects, the data from Olivola et al. (ms) shows that South Koreans are over 200% more likely to plug into the experience machine, men are about twice as likely to plug into the experience machine as women, and the older you are the more likely you are to plug in. For framing effects, the data from de Brigard's (2010) shows that how non-philosophers react to the experience machine largely depends on how the case is described. While people in general are unwilling to plug in to the experience machine, they are also unwilling to *unplug* if they are told to imagine they are already in the experience machine. And because it appears that there is bias for whatever one's current situation is, this thought-experiment does not obviously refute the simple kinds of hedonism that many philosophers believe it does.[17]

[17] See Machery (2017: 58, 65, 83, 86-87) for Machery's discussion of these studies.

According to Machery, what best explains this variance in intuitive judgments is that the philosophical cases themselves are defective. Specifically, philosophical cases are problematic because they typically have the "disturbing characteristics" mentioned above: they are unusual hypothetical cases that pull apart what typically goes together. For Machery, it is precisely because philosophical cases typically have these features that makes them unreliable guides to such modal knowledge. In contrast, Machery claims everyday uses of the method of cases does not appeal to this type of case:

> The issue that Unreliability brings to the fore is not that there is something intrinsically wrong with using cases in philosophy. Indeed, how could that even be the case? And surely there is nothing suspicious *in general* in judging about situations that are described. We do it all the time, and our judgements are warranted. (Machery 2017: 91)

Additionally, Machery writes:

> [T]here is in fact little reason to believe that judgments about hypothetical situations are in general unreliable: Simply consider the judgment that this book will fall instead of going up if you released your grip on it. (Machery 2017: 113)

So, Machery is not a general skeptic about all our general judgements regarding knowledge, right action, etc. because he claims that we often do make reliable judgments about such things.[18] Machery is just a skeptic regarding the type of cases philosophers typically have to appeal to in order to analyze the modal nature of their philosophical theses. It is this kind of knowledge that is beyond our capacities to attain by using the

[18] Machery (2017: 103) explicitly says that he is assuming for the sake of argument that certain error theories about the nature of things that philosophers are interested in (e.g. normativity, right action) are false.

method of cases. So, according to Machery, the extensive x-phi literature shows that philosophers cannot use the method of cases to acquire the modal knowledge they are after.[19]

Indeed, Machery goes further and argues that this large literature supports a general inductive pessimism about all current and potential philosophical cases. Machery holds that unless a philosophical case has been shown by empirical means to elicit the same intuitive judgment among philosophers and non-philosophers, regardless of demographic and presentation effects, that we cannot trust any such philosophical case to support or undermine a philosophical theory. In short, philosophical cases are guilty until proven innocent.

Machery helpfully formalizes his argument as:

Unreliability

1. Unreliable judgements are severely deficient from an epistemic point of view.
2. Judgements elicited by most of the philosophical cases that have been examined by experimental philosophers are unreliable.
3. If the judgements elicited by most of the philosophical cases that have been examined by experimental philosophers are unreliable, then the judgements elicited by most philosophical cases are plausibly unreliable.
4. We ought to refrain from making a judgement of a particular kind K (i.e., we ought to suspend judgement of kind K) when most judgements of this kind are

[19] Machery (2017: 118-120) also argues that philosophical cases are problematic because the "superficial and target content" of a case are often inextricably entangled and that often people respond to the superficial content in ways that account for the variability in intuitive responses. In other words, it is the narrative elements of a case that people often respond to and not what is relevant for philosophical purposes. I will set this complication aside in this chapter.

plausibly severely deficient from an epistemic point a view, except when this judgement is known to be an exception.

5. Hence, except when a philosophical case is known to elicit a reliable judgement, philosophers ought to suspend judgement about the situations described by philosophical cases. (Machery 2017: 102-103)

Premise 1 states a reliability condition (more on this in the next section). Premise 2 articulates, according to Machery at least, the findings of the empirical x-phi literature that nearly all of the most cited philosophical cases are significantly influenced by demographic or presentation effects. As such, this premise forms what Machery calls the "inductive base" for this Unreliability Argument. Premise 3 attempts to go beyond the inductive base and support the claim that all potential cases philosophers might construct (to test philosophical claims to their intuitions) are suspect. This premise forms what Machery calls the "inductive step" of his Unreliability Argument. Premise 4 articulates a normative claim that, in the present case, amounts to the claim that because the *type* of cases philosophers typically appeal to (i.e. ones with distorting characteristics) are unreliable, we ought to suspend judgment about them. Premise 5 concludes with a near global skepticism about philosophers' use of the method of cases. And according to Machery, because appeals to such cases are both common and usually necessary for philosophers to investigate and (allegedly) attain the modal knowledge they seek, a pervasive kind of philosophical skepticism follows.

4. Against Machery's Unreliability Argument

Machery's Unreliability Argument is very radical. If this argument is sound, then one of the central methods philosophers use to attempt to acquire philosophical

knowledge is bunk and a pervasive philosophical skepticism likely follows. As Machery puts it, the aim of this argument is to curb "philosophers' flights of fancy" by showing that "resolving many traditional and contemporary philosophical issues is beyond our epistemic reach" (Machery 2017: 1).

However, there are many ways to challenge Machery's Unreliability Argument. For example, one could challenge his characterization of philosophy and its use of the method of cases, one could challenge whether or the extent to which the x-phi literature supports the unreliability of current philosophical cases (premise 2), one could challenge whether the current x-phi literature warrants an inductive generalization or general pessimism against all potential philosophical cases (premise 3), or one could challenge Machery's characterization of philosophical cases as often requiring the disturbing features he mentions. While I believe that these, and other, challenges to Machery's Unreliability Argument are worthwhile, I will set these challenges aside and assume for the sake of argument that these aforementioned features of Machery's argument are correct.[20]

Instead, I want to argue that, like Silva, Machery has failed to appreciate the extent to which his Unreliability Argument is epistemically self-defeating: in the present case, if its conclusion is true, then this defeats the justification for some of its premises. In short, I argue that premise 1 and 4 of Machery's argument are either the kind of modal claims that this argument denies we philosophers have access to or these premises have to be supported by intuitions about the very kind of cases this argument deems illegitimate.

[20] However, I should note that my argument in §2 against Silva's Unreliability Argument (that inductive generalizations are largely supported by intuitions about cases) also potentially undermines Machery's 3rd premise (i.e. the inductive step).

Either way, the conclusion of Machery's own argument defeats these premises. To be clear, the issue here is not whether these premises are true; indeed, with proper refinements, I believe something like them is true. This issue is whether, by Machery's own lights, he can legitimately accept and use them in his Unreliability Argument. Furthermore, I argue that for Machery to avoid this deeply problematic feature of his argument, he must severely limit the scope of his Unreliability Argument and the philosophical skepticism it attempts to establish.

Let me begin by noting one common and problematic feature of both premise 1 and 4: Both premises include the phrase that unreliable judgements are "severely deficient from the epistemic point of view." This phrase is problematic because it does not precisely articulate what is wrong with unreliable judgments (in premise 1) or why we should suspend judgment about unreliable judgements (in premise 4). Typically, philosophers would state such premises in terms of 'justification' or 'knowledge' to make such claims more precise and informative. For example, reliabilists like Goldman (1979) would argue that unreliable judgments are insufficient for justification and knowledge. But Machery contends that this should be avoided to circumvent the discussions and controversies surrounding such epistemic notions, especially discussions about "what is sufficient or necessary for justification and knowledge" (Machery 2017: 97). Specifically, Machery mentions the often-discussed counterexamples to reliabilism (i.e. Trutemp and Norman the clairvoyant), the internalism/externalism about justification debate, and the generality problem.

But the crucial question is: why we are warranted in avoiding such discussions and allowing these premises to be expressed in an imprecise (and potentially

uninformative) way? Machery only says that many proponents of intuitions and the method of cases assume something like premise 1 and 4 (Machery 2017: 103), and so he is following their lead. But this won't do, since presumably proponents of intuitions and the method of cases accept these premises on the basis of considering intuitive judgments elicited from philosophical cases. Indeed, such proponents are likely to spell-out premises 1 and 4 in terms of reliability being *necessary* for justification or knowledge. And this also won't help Machery support these premises for two reasons. Firstly, this appeals to exactly the kind of modal knowledge that he argues philosophers cannot attain by using the method of cases. Second, proponents of intuition usually still support this kind of modal knowledge by appealing to cases Machery's argument forbids. For instance, reliabilists like Goldman (1979) hold that reliability is necessary and sufficient for justification and knowledge, and supports the centrality of reliability for these epistemic notions by appealing to the following cases: paradigm unreliable processes like confused reasoning, wishful thinking, and guesswork also do not provide adequate justification or knowledge; while paradigm reliable processes like sense perceptions, remembering, and good reasoning do provide adequate justification or knowledge (Goldman 1979: 95). It is not clear that these cases are non-philosophical in the way that Machery allows.

Indeed, to make matters worse, Machery relies on the following definition of reliability he approvingly takes from Goldman (1979) and Alston (1995):

> T, a psychological process outputting judgments, is reliable in environment E *if and only if* in E either T has the disposition to produce a large proportion of true judgments or, if T is an inferential process, T has the disposition to produce a

> large proportion of true judgements if its inputs are true. (Machery 2017: 96, emphasis added)

But this is also the kind of modal knowledge that Machery's argument forbids or is likely supported by the kind of cases Machery disavows. Either way relying on this definition seems to show another way that Machery's argument is self-defeating. Indeed, to support the dispositional aspects of this definition Machery writes that "reliability is a dispositional property and a process used only once can still be assessed for reliability" (Ibid). Additionally, Machery gives the following example to demonstrate the environment clause of this definition: "A fake-bill detector may be reliable (it may detect a very large proportion of fake bills) when the fake bills are of low quality, but unreliable when a super-duper counterfeiter made them" (Ibid). But it is not clear that Machery can appeal to cases like fake-bill detector since they seem to have the disturbing characteristics Machery identifies as problematic That is, Machery has to explain why cases like fake-bill detector are not the forbidden kind of unusual hypothetical case that pulls things apart that typically go together.

So, what warrants premises 1 and 4 for Machery? Another answer Machery could provide is that it is *obvious* that something like these premises are true. Indeed, Machery might argue that these premises can be supported by intuitive judgments about non-philosophical or everyday cases. As mentioned in the last section, Machery is not a skeptic about everyday uses of the method of cases because he holds that we often make reliable judgments about such cases. So maybe premise 1 and 4 are supported by such cases. While Machery does not explicitly say this is what warrants premise 1 and 4, he does provide cases that seem to support these premises. For premise 1, he writes:

> If judgments produced by an unreliable process were not severely deficient from an epistemic point of view, then choosing what to believe by random (by, e.g. throwing a coin to decide what to believe) or choosing on the basis of a process that works as designed but does not do better than a random process would result in epistemically appropriate or only moderately deficient beliefs. (Machery 2017: 97)

Furthermore, Machery writes that he "will take Premise 4 for granted" but explains that what it recommends is in line with the following practice:

> If I know that most eggs in a pack are rotten, the reasonable thing in the absence of further information is to throw out the *whole* pack. (Machery 2017: 93)[21]

But this also won't do, and it is the purpose of the rest of this section to explain why.

Starting with premise 1, which states: "Unreliable judgments are severely deficient from an epistemic point of view" (Machery 2017: 2012). As stated, this premise is a generic statement of the form 'Fs are Gs.' Generic statements are notoriously difficult to interpret semantically since they allow for different interpretations about their degree of quantification. For example, 'dogs are mammals' is correctly interpreted as '*all* dogs are mammals' but 'ticks are carriers of Lyme disease' must be interpreted as '*only a very few* ticks are carries of Lyme disease' since approximately only 1% of ticks carry this disease.[22] These examples illustrate the wide quantification spectrum and possible interpretations of generic statements.

[21] Machery (2017: 110, 183) also uses similar examples to support premise 3 (i.e. the inductive step) of his argument.

[22] These examples are taken from Leslie and Lerner (2016).

How are we to interpret the degree of quantification in premise 1? It appears that for Machery to reach his radical skeptical conclusion, this premise must be interpreted as having a universal quantifier (i.e. '*All* unreliable judgments…"). For, if the kind of quantification is anything less than this, then Machery's argument would allow for some (however few) unreliable judgments to not be severely defective. This would prevent Machery's argument from reaching its radical conclusion – i.e. a general skepticism about philosophers' use of the method of cases. But Machery cannot lend himself to this interpretation of premise 1 because, with such universal quantification, this premise is essentially saying that 'unreliable judgments are *sufficiently* severally defective from an epistemic point of view.' But this is exactly the kind of modal knowledge that Machery's Unreliability Argument prohibits. Indeed, this interpretation is equivalent to 'reliable judgments are *necessarily* non-defective from an epistemic point of view.'

And, I must confess, I am not sure what this equivalent version of premise 1 means precisely. If this premise was stated as 'reliable judgments are necessary for knowledge or justification' then I would have a better idea of what Machery's argument requires. But, recall that Machery explicitly chooses to use the less precise phrase 'severely deficient from an epistemic point of view' to avoid the controversies surrounding what is necessary or sufficient for knowledge (Machery 2017: 97). What I wish to get across is that it does not appear that this imprecise phrase allows Machery to avoid the thrust of such controversies since premise 1 is a generic statement that must be interpreted with universal quantification. That is, even though premise 1, as stated, avoids commitments to how reliability relates to more familiar epistemic notions like 'justification' and 'knowledge,' it does this by being opaque and uninformative. And it is

unclear how this imprecise language is helpful to Machery's project since it is still committed to some forbidden modal connection between 'unreliability' and being 'severely deficient from an epistemic point of view.' Hence, Machery's own argument very likely defeats its first premise.

I think it is natural for Machery to respond by steadfastly claiming that premise 1 is obviously true and supported by everyday (i.e. non-philosophical) cases, like the coin flip example quoted above. But, for what it is worth, I think that the coin flip case to support premise 1 likely has the disturbing characteristics that Machery abhors. It is definitely unusual and it seems to pull things apart that typically go together. How often does one decide what to believe on the basis of a coin flip? However, it is difficult to know if this is the case since Machery's disturbing characteristics are too imprecisely defined to make this determination.

Regardless, appealing to such (alleged) everyday cases will not help Machery support the premises of his argument for a variety of reasons. To illustrate, let's turn to the rotten eggs example Machery uses to support premise 4 of his Unreliability Argument: "We ought to refrain from making a judgement of a particular kind K (i.e., we ought to suspend judgement of kind K) when most judgements of this kind are plausibly severely deficient from an epistemic point a view, except when this judgement is known to be an exception" (Machery 2017: 102-103). Firstly, such examples cannot provide adequate support for this premise since there are other (supposedly) everyday examples that contradict this premise. For example, imagine that I will certainly die unless I receive a lifesaving surgery. Unfortunately, there are only two options: Surgery 1 which has a 25% success rate or Surgery 2 which has a 45% success rate. Even more unfortunately, I

have unknowingly been given misinformation about the success rates of these surgeries. I am told that Surgery 1 has a 40% success rate and Surgery 2 has a 30% success rate. However, contra premise 4, in such a case I will not suspend judgment about which surgery I will request. Even though both are severely deficient in saving lives and I cannot reliable judge which surgery is best, I will pick Surgery 1 over certain death.

This counter-case illustrates the fickle justificatory nature of everyday cases in supporting imprecise premises, like premise 4 (and 1). This surgery example is only able to contradict premise 4 because this premise uses "severely deficient from an epistemic point of view" instead of more precise characterizations like: 'insufficient for justification' or 'insufficient for knowledge.' It is clear that I would neither know nor be justified in believing I would survive the surgery. But, again, Machery explicitly avoids putting his argument's premises in these terms to avoid the controversies surrounding justification and knowledge. And, again, to do so would seemingly commit himself to the kind of modal knowledge his argument blocks. But it seems that Machery's premises need these notions (i.e. justification, knowledge, etc.) since without them it is not clear how preferring Surgery 1 (or Surgery 2) is still "severely deficient from an epistemic point of view." So, this provides more reason to believe that Machery's unreliability argument is self-defeating since it seems it either relies on forbidden modal knowledge or needs to appeal to forbidden philosophical cases (instead of everyday cases) to support its premises.

The fickle justificatory nature of everyday examples also blocks another way Machery might attempt to defend his argument from self-defeat. Machery (2017: 15) claims that his attack on philosophers' use of the method of cases does not impugn the

use of cases in all contexts. For instance, he argues that using cases to *illustrate* abstract philosophical notions and distinctions is innocuous and exempt from his argument which attempts to undermine the alleged *justificatory* nature of philosophical cases. Machery could appeal to this distinction and argue that his above cases are merely meant to illustrate the content of his premises but do not attempt to justify them since that would lead to the self-defeat worries that I have raised.

But this is a potentially problematic move for Machery since appealing to this distinction would seem to severely limit the scope of his Unreliability Argument. If philosophical cases can serve a merely illustrative function, then this offers a way for proponents of philosophers' use of the method of cases to avoid Machery's argument entirely. For instance, on this construal the experience machine thought-experiment would be merely illustrating the claim that having real world experiences are often considered to be more valuable than virtual experiences. It seems that a similar move can be made for all of the philosophical cases examined in the x-phi literature.[23] Additionally, and more worrisome for Machery, is that it is not clear that Machery can appeal to this distinction between illustrative versus justificatory cases since this distinction will seemingly have to be justified by cases which will likely have the disturbing characteristics he forbids. Indeed, it seems that appealing to everyday cases often serves to *both* illustrate and justify the claims we are making, and this seems to be true of Machery's examples above (e.g. coin flip, fake bill detector, etc.). So, this is not a way Machery can avoid the challenge of self-defeat while retaining his pervasive philosophical skepticism.

[23] Cf. Strohminger (2018).

Additionally, there is another even more pernicious way that Machery's Unreliability Argument is self-defeating. In surveying the vast x-phi literature to support his second premise (i.e. "Judgements elicited by most of the philosophical cases that have been examined by experimental philosophers are unreliable" (Machery 2017: 102)), Machery overlooks a study by John Turri (2016). In this paper, Turri (2016) presents nine experiments that he argues show that unreliably produced knowledge is compatible with our everyday folk intuitions about knowledge. In other words, Turri argues that non-philosophers are likely to judge that unreliable judgments are *not* always severely deficient from an epistemic point of view, but are often conducive to acquiring *knowledge*. This empirical study directly contradicts premise 1 of Machery's argument and if cogent, shows an additional way that Machery's argument is self-defeating. The goal of Machery's argument is to show how the x-phi literature undermines "philosophers' flights of fancy" by showing that "resolving many traditional and contemporary philosophical issues is beyond our epistemic reach" (Machery 2017: 1). But if Turri's study is correct, then Machery's argument was presupposing some claims about reliability that are also (allegedly) beyond our epistemic reach.[24]

To be clear, I am not convinced by Turri's (2016) study since in my estimation (see Chapter 1) Turri's study is flawed in a number of ways. What is important for our present purposes is to show that it will be extremely difficult *for Machery* to deny the cogency of Turri's (2016) study without also thereby undermining some of the empirical

[24] I should note that Machery (2017: 97) mentions and briefly discusses another paper by Turri (2015) that argues for unreliable judgments being able to produce knowledge. Machery argues that he and Turri "concur on the importance of reliability in the present case" since they both agree that achievement of knowledge "requires doing substantially better than luck." But this is incorrect since Machery wishes to deny that unreliable judgments can produce knowledge while Turri's (2015 and 2016) arguments accept this.

support for his own Unreliability Argument. But unless Machery can deny the cogency of this study, his first premise is defeated by his own conclusion.

To show why this is the case, let's consider one of Turri's (2016) experiments that seems to support the claim that non-philosophers are willing to ascribe knowledge to agents who have unreliable abilities. A portion of Turri's 136 participants read the following case involving Alvin who has a true but unreliably produced belief:

> Alvin is very unreliable at remembering driving directions. Today he is visiting a friend in an unfamiliar town. Alvin needs to pick up a prescription while his is there, so his friend gives him directions to the pharmacy. On the way, Alvin needs to make a right turn at the intersection. Alvin gets to the intersection and turns right. (Turri 2016: 205)

After reading this case, participants were asked whether Alvin "knew/only thought" that he should turn right to get to the pharmacy. Participants who read this case overwhelming attributed knowledge to Alvin (77%) despite his unreliable memory.

This empirical support is particularly troubling for Machery since this case does not have any of the disturbing characteristics that he claims philosophical cases typically have: an unusual hypothetical case that pulls apart what typically goes together. Indeed, being unreliable at remembering driving directions is an everyday occurrence that many can directly relate to through their own experience. And if not, it is well within any competent adult's capacity to imagine such an everyday case. As such, Machery cannot write-off this case as suspicious because it is unusual or philosophically extravagant.

It seems then that Machery must challenge the nuts-and-bolts of this study. While I do not want to pontificate on how Machery would challenge Turri's study, what I will

say is that this is a dicey proposition for Machery. Since Turri's study is done with the same methods and statistical rigor as many other x-phi studies, challenging the nuts-and-bolts of this study is likely to impugn other x-phi studies that Machery relies on in making his Unreliability Argument. In effect, by challenging Turri's study to avoid the possibility of Machery's argument being empirically self-defeating, Machery will likely be challenging much of the empirical support for his own argument.[25]

In summary, Machery's Unreliability argument seems to be unavoidably self-defeating. Premise 1 and 4 seem to either rely on forbidden modal knowledge or banned philosophical cases. And, neither of these premises can be supported by everyday non-philosophical cases since such cases have a fickle justificatory nature in supporting imprecise premises. Lastly, Machery overlooks some empirical evidence that undermines a key premise in his argument, which he cannot challenge without likely impugning the empirical support for his Unreliability Argument and philosophical skepticism.[26]

5. Conclusion

Historically, considerations of self-defeat have often been used to defend philosophical knowledge from various arguments for philosophical skepticism. For example, against Hume's famous passage at the end of the *Enquiry* which decrees that

[25] Of course, Machery could argue that this kind of problem is only temporary since more testing should be done to see if Turri's findings are replicated. If they are not, then Machery can avoid the problems raised by Turri's study and potentially involve the exception clause in premise 4 of his Unreliability Argument. While I admit this would help Machery's argument against Turri's study, until such testing is done, Machery has to follow his own advice and suspend judgment about the cogency of his own argument.

[26] I should note that Machery responds to a similar but different kind of self-defeat challenge. In his review of Machery (2017), Hughes (2019) argues that another one of Machery's arguments is self-defeating and Machery's (2019) response is relevant to our discussion of his Unreliability Argument. In essence, Machery argues that he can accept theses and premises that all interlocutors agree on without further argument. And since his interlocutors are likely to accept premises 1 and 4, he has good inductive reason to accept these premises. But as we have already seen, this response will not do since his interlocutors likely would accept these premises on the basis of intuitions about philosophical cases.

large portions of philosophy (esp. metaphysics) should be "consigned to the flames," some commentators have argued that much of the philosophy he condemns is used to support his philosophical skepticism. Similarly, against verificationism's claim that many philosophical words are meaningless because they cannot be verified by observation, commentators have argued that verificationism itself would be meaningless by this standard.

In this chapter I have argued that this general self-defeat challenge also applies to Unreliability Arguments for philosophical skepticism. I take this to be a significant result that casts some doubt on the possibility of using x-phi studies to argue for a broad philosophical skepticism. I believe that the lesson to be drawn from my discussion is that philosophical knowledge cannot be substantially threatened by such arguments.

CHAPTER 3

SAVING PHILOSOPHICAL KNOWELDGE FROM DISAGREEMENT SKEPTICISM

Imagine you and a friend are out to dinner and you both decide to evenly split the bill with a 20% tip.[1] The bill comes and you both independently calculate the total amount you each owe and come to different numbers; you calculated $43 and your friend calculated $45. You and your friend know that you are both equally competent at this elementary math, neither of you is impaired in any way in making this calculation, and neither of you is intentionally attempting to deceive the other. In short, you are *epistemic peers* when it comes to this kind of math. Intuitively, it seems that at the moment of recognized peer disagreement you and your friend should suspend belief in your answers, and that neither of you knows the correct split-bill amount. This split-bill case and others like it are often used to support Conciliationism. Roughly, Conciliationism is the family of views that says in cases of recognized epistemic peer disagreement agents should significantly reduce their confidence or suspend belief in their contested beliefs. According to Conciliationists, rationality requires that agents revise their confidence or beliefs in this way because peer disagreements provide defeating evidence against one's beliefs.[2]

[1] This example is adapted from Christensen (2007).

[2] Conciliationism is defended by Bogardus (2009), Christensen (2007, 2009, 2010, 2011, 2013), Cohen (2013), Elga (2010), Feldman (2005, 2006a,b, 2009), Littlejohn (2013), Matheson (2009, 2015), and Vavova (2014a,b), among many others.

While such everyday cases provide *prima facie* support to Conciliationism, the skepticism that results from such cases is limited and usually temporary.[3] In the split-bill case, only you and your friend's beliefs about the correct amount are undermined, and you and your friend will very likely come to a consensus about the right split-bill amount if given enough time to compare each other's calculations. But, Conciliationism can lead to a widespread and permanent skepticism in any domain where there is a considerable amount of intractable peer disagreement. In the domain of philosophy, there are a plethora of examples of philosophical peers or philosophical schools of thought that are perennially unable to come to a consensus. Indeed, after surveying a large portion of contemporary philosophers, Bourget and Chalmers (2014) found that contemporary philosophers' opinions on many current and traditional philosophical problems are fairly evenly distributed between incompatible philosophical views.[4] Conciliatory reasoning dictates that because a great many (or nearly all) philosophical propositions are disagreed upon by philosophical peers that philosophers are rationally required to reduce confidence or suspend belief in their contested philosophical beliefs. And, since undefeated belief is a necessary condition for knowledge, a pervasive philosophical

[3] I say "usually" because there are cases where peers have the ability to resolve their disagreement but do not because something prevents them (e.g. time constraints, laziness, death, etc.).

[4] These facts about *actual* peer disagreement in philosophy allow me to set aside the issue about whether *potential* peer disagreements also have skeptical consequences (see Barnett and Han (2016), Cary (2011), Cary and Matheson (2013), Christensen (2007), Kelly (2005), and Kornblith (2010)); and these facts about the distribution of disagreement between philosophers also allow me to set aside issues regarding the difficulties surrounding the number of disagree-ers (see Cary and Matheson (2013), and Lackey (2013)).

skepticism follows.[5] If we couple this argument with the fact that philosophers have a poor track-record of ever resolving disagreements and reaching a consensus on philosophical claims, then philosophical knowledge threatens to be perennially unattainable or impossible.

In this chapter I aim to save philosophy from this kind of skepticism by arguing that all plausible versions of Conciliationism are false.[6] In short, I argue that any version of Conciliationism that leads to philosophical skepticism for some philosophical proposition will also be epistemically self-undermining, veridically self-undermining (both of which are defined below), and lead to a hitherto unrecognized *reductio ad absurdum.*

In §1 I explain how a version of Conciliationism known as the Equal Weight View (EWV) leads to philosophical skepticism. In §2 I argue that the EWV is epistemically self-undermining by responding to the large literature that attempts to defend the EWV (and Conciliationism) from this self-undermining challenge. In §3 I further argue, contra defenders of the EWV (and Conciliationism), that *if* the EWV is epistemically self-undermining then it is also false because it is veridically self-undermining. While I believe that my arguments in §2 and §3 are sufficient to refute the EWV, in §4 I argue that *even if* there is a way for the EWV to avoid both the epistemic and veridic self-undermining challenges, the EWV also leads to an additional *reductio ad*

[5] See Barnett (2019), Brennan (2010), Christensen (2007, 2009), Frances (2005, 2013) Feldman (2006b), Fumerton (2010), Goldberg (2013a,b), Kornblith (2010, 2013), and Vavova (2014a) for discussions of philosophical skepticism following from Conciliationism.

[6] While I will focus on philosophical disagreement skepticism in this chapter since it is the most discussed kind of skepticism that follows from Conciliationism, my refutation of Conciliationism refutes disagreement skepticism in all domains where there is widespread and intractable peer disagreement (e.g. religion and politics).

absurdum. This reductio argument side-steps the issues surrounding self-incrimination and provides an independent argument that decisively refutes the EWV. In §5 I extend these results by showing that all plausible versions of Conciliationism are also false because they lead to this additional reductio argument. In §6 I further extend these results by showing that other competing views to Conciliationism (i.e. the Justificationist View and the Total Evidence View) are false because they also lead to this reductio argument. And in §7 I conclude by reflecting on the counterintuitive consequences of these results.

1. The EWV and Philosophical Skepticism

The split-bill case and others like it are often used to support a version of Conciliationism called the Equal Weight View (EWV).[7] The belief version of the EWV roughly says that when peers recognize that they disagree on p,[8] these peers should suspend belief on p. (For the time being I will focus on the simpler "suspending belief" versions of Conciliationism and return to the more complicated "reducing confidence" versions in §5). Defenders of the EWV provide two general reasons to support their view. Firstly, they claim that rationality requires agents to give equal weight to the opinions of their epistemic peers because epistemic peers are equally likely to be in a position to know p. And secondly, suspension of belief is rationally required in such cases because peer disagreement about p is strong reason to believe that at least one peer is mistaken, but, at the moment of disagreement, neither peer has good reason to discount the other's

[7] The EWV is defended by Bogardus (2009), Cary and Matheson (2013), Christensen (2007), Cohen (2013), Elga (2007), Feldman (2005, 2006a,b), and Matheson (2009).

[8] Henceforth, "peer disagreement" will always refer to *recognized* peer disagreement.

belief in favor of their own. In other words, peer disagreement counts as *defeating* evidence against one's belief in p.[9]

For clarity, it will be useful to provide a formal representation of how the EWV leads to philosophical skepticism. To this end, if we understand 'p' to be the variable for any proposition and 'q' to be the variable for any philosophical proposition (where set q is a subset of set p), then the argument that the EWV leads to philosophical skepticism can be formalized as:

The EWV Leads to Philosophical Skepticism

1. If peers disagree on p, then peers should suspend belief on p. (EWV)
2. If peers should suspend belief on p, then peers don't know that p. (EWV→~K)
3. Peers disagree on q. (Empirically true premise)
4. Thus, peers should suspend belief on q. (1, 3, MP)
5. Thus, peers don't know that q. (2, 4, MP)

Premise 1 is a simplified statement of the belief version of the EWV;[10] premise 2 embodies the claim that undefeated belief is a necessary condition for knowledge, which

[9] EWV holders, and Conciliationists more broadly, also adhere to the principle of Independence: "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 of the reasoning behind my initial belief about p" (Christensen 2011: 1-2). I will set this complication aside since my arguments below are not affected by whether or not this principle is true (See Arsenault and Irving (2013) and Lord (2014) for arguments against Independence).

[10] This is a simplified version of the EWV for two reasons. First, the antecedent is usually filled with many more conditions to make the EWV apply only to a restricted set of cases. For my purposes in this chapter I can safely set aside these additional conditions (e.g. (approx.) peers are competent, honest, unimpaired, etc.) since any such conditions will presumably be satisfied in (nearly) all, philosophical peer disagreements. Second, this statement of the EWV assumes that cases of peer disagreement rationality require agents to suspend belief on *any p*; while the EWV (and Conciliationism) is sometimes limited to quantify over only *most p* (e.g. Christensen 2007: 189; also see Decker (2014)). But, with the possible exception of the EWV itself (see §2), I can also safely set aside this complication since presumably (nearly) all cases of philosophical peer

in this case amounts to the claim that if one should suspend belief in p (because of defeating evidence against p (e.g. peer disagreement on p)), then one cannot know p; premise 3 is the empirical claim that philosophical peers disagree over various philosophical propositions; and 4-5 are logical consequences of premises 1-3.

To block this argument's conclusion at least one of the first three premises must be denied. While there are some reasons to doubt premises 2 and 3,[11] I will set these issues aside and assume for the sake of argument that they are true. Anyone who denies either of these premises should read the rest of the arguments in this chapter conditionally: *if* these premises are true then my arguments in §2-5 follow. This, I claim, is still a significant result.

Many have also denied premise 1 by denying the EWV (and Conciliationism) in favor of Steadfastness. Steadfastness is roughly the family of views that says it is rationally permissible in at least some (if not all) cases of peer disagreement to retain one's level of confidence or belief.[12] I will set aside the arguments in favor of Steadfastness and against the EWV (and Conciliationism) because my arguments in §2-5

---

disagreement are not about propositions that can plausibly thought to be outside the scope of the EWV (and Conciliationism) (e.g. logical truths). See Jehle and Fitelson (2009) for additional difficulties in providing a precise definition of the EWV.

[11] Myers-Schulz and Schwitzgebel (2013) and Tebben (2019) argue that, contra premise 2, belief is not necessary for knowledge (see Rose and Schaffer (2013) for a response). Elgin (2010) denies 2 by arguing that because beliefs are largely out of our control peer disagreement is not apt to motivate or rationally obligate suspension of belief. Cary and Matheson (2013), Frances (2014), King (2012), and Rotondo (2015) deny 3 by arguing that disagreement skepticism is of limited significance because it is difficult, if not impossible, to determine whether someone is an epistemic peer. As such, disagreement skepticism is rarely justified since epistemic peers are non-existent or hard to find.

[12] Proponents of Steadfastness include Enoch (2010), Fumerton (2010), Kelly (2005, 2010, 2013), Lackey (2010a,b), Pasnau (2015), Schafer (2015), Titelbaum (2015), Weatherson (2013), and Wedgewood (2010).

provide novel arguments for Steadfastness and against all plausible versions of Conciliationism.

2. The EWV is Epistemically Self-Undermining

Another popular way to deny 1 is to argue that the EWV is epistemically self-undermining.[13] A view or proposition p is epistemically self-undermining iff p applies to itself and defeats one's justification for believing p, i.e. p epistemically defeats itself. For example, the belief that "no one is justified in believing a normative epistemic claim" (Decker 2014: 1102) is epistemically self-undermining because if this claim were true then it would apply to itself and no one would be justified in believing it. Similarly, the EWV seems to be epistemically self-undermining because there is disagreement about whether it is true and so the EWV applies to and epistemically defeats itself.[14] This challenge can be formalized using the previous argument for philosophical skepticism as:

The EWV is Epistemically Self-Undermining

1. If peers disagree on p, then peers should suspend belief on p. (EWV)
2. If peers should suspend belief on p, then peers don't know that p. (EWV→~K)
3. Peers disagree on EWV. (Empirically true premise)
4. Thus, peers should suspend belief on EWV. (1, 3, MP)
5. Thus, peers don't know that EWV. (2, 4, MP)

[13] This is discussed by Bogardus (2009), Blessenhol (2015), Christensen (2009, 2013), Decker (2014), Elga (2010), Littlejohn (2013, 2014), Matheson (2015a,b), Pittard (2015), and Weatherson (2013).

[14] Of course, the proposition that "no one is justified in believing a normative epistemic claim" is *necessarily* epistemically self-undermining because its being epistemically self-undermining does not depend on anyone believing it. In contrast, the EWV is only *contingently* epistemically self-undermining because its being epistemically self-undermining depends on their being epistemic peers who disagree on it. It is a virtue of the epistemic self-undermining definition that it allows for both of these.

This argument is identical to The EWV Leads to Philosophical Skepticism argument in §1 except philosophical proposition q in premises 3-5 is filled-in with 'EWV.'

While I believe that this self-undermining argument is sound and provides strong reason to deny the EWV, many defenders of the EWV (and Conciliationism more broadly) have provided many different responses to it. These responses can be put into two groups: the *Solution Responses* which attempt to provide an argument that shows the above epistemic self-undermining argument is flawed; and the *Skeptical Responses* which accepts that this argument is sound but attempts to mitigate the impact of its skeptical conclusion (i.e. this argument does not provide reason to believe the EWV is *false*, but only provides reason to believe that no epistemic peer knows the EWV). While a detailed survey of these responses is outside the scope of this chapter,[15] I will briefly address each kind of response in turn. I argue that both kinds of responses (almost certainly) fail to save the EWV from this self-undermining challenge. In this section I address the Solution Responses, and in the next section I address the Skeptical Responses.

To begin, I argue that all potential Solution Responses are likely to be defeated by the very same Conciliatory reasoning they are attempting to defend. More specifically, I argue that because Solution Response *arguments* almost certainly contain reasoning or premises that are or will be disagreed upon by epistemic peers, that they will likely fail to provide an adequate (i.e. epistemically undefeated) defense of the EWV from the epistemic self-undermining challenge. My argument can be formalized as:

Solution Responses (almost certainly) Fail

[15] See Decker (2014) for such a survey.

I. If argument A adequately defends the EWV from the epistemic self-undermining argument, then epistemic peers do not disagree with the reasoning or auxiliary premises of A.

II. Epistemic Peers (almost certainly) disagree with the reasoning or auxiliary premises of A.

III. Thus, A (almost certainly) fails to adequately defend the EWV from the epistemic self-undermining argument.

Premise I essentially requires defenders of the EWV to consistently apply the EWV (and Conciliationism more broadly) to any reasoning or auxiliary premises in their Solution Response arguments, such that: if epistemic peers disagree on these aspects of their Solution Response arguments then they are committed to suspending belief on these aspects. Premise II is an empirical claim that is indirectly supported by the general fact mentioned above that philosophical peers and schools of thought tend to disagree about any substantive philosophical proposition (Bourget and Chalmers (2014)).

But premise II is also directly supported by the large literature on Solution Responses to the epistemically self-undermining challenge. For instance, the most common Solution Response from defenders of the EWV (and Conciliationism more broadly)[16] is to argue that when Conciliatory views are properly understood they are self-exempting and so they are not epistemically self-undermining. The most discussed version of this kind of response comes from Elga (2010) who argues that in order for Conciliatory views to be consistent (i.e. they do not call for their own rejection) and not be epistemically self-undermining they must be dogmatic about their own correctness.

[16] See Christensen (2013) and Frances (2010) for why this self-undermining challenge implicates all plausible versions of Conciliationism.

While this may seem like an ad-hoc solution, Elga contends that this constraint is not ad-hoc because "this is a completely general constraint that applies to any fundamental policy, rule, or method" (Elga 2010: 185).

The details of Elga's argument are unimportant because they fail, by Conciliatory reasoning, to save the EWV from being epistemically self-undermining. This is because the *literature* on Elga's argument is full of epistemic peers, including many Conciliationists, who contest the soundness of his argument, per premise II. For example, Christensen (2013) argues that Elga's self-exempting constraint begs the question and undermines the motivations for Conciliationism; Decker (2014) argues that Elga's argument fails to show that Conciliationism leads to inconsistency because it makes unstated false assumptions and commits a deontic fallacy, and without this inconsistency the self-exempting constraint remains ad-hoc; and Littlejohn (2013) also argues that Elga's argument fails to show that Conciliationism leads to inconsistency and contends that there is no need for it to be self-exempting since the epistemic self-undermining challenge is not very worrisome. Indeed, there are many more examples in this literature of epistemic peers denying some aspect of Elga's argument, per premise II.[17] So, by the Conciliatory reasoning in premise I, Elga's self-exempting Solution Response to the

[17] Additionally, Pittard (2015) argues that Elga's argument fails to consider alternative non-self-exempting ways to respond to the self-undermining challenge and that his arguments fail to adequately address the ad-hoc (or arbitrariness) worry of making Conciliatory views self-exempting; and Matheson (2015a) argues that to avoid Elga's inconsistency and the epistemic self-undermining challenge, instead of making Conciliationism self-exempting, Conciliationism should be revised into a weaker form that also adopts an Evidentialist meta-rule that both allows it to be consistent and prevent it from being epistemically self-defeating.

epistemic self-undermining challenge fails to adequately defend the EWV from The EWV is Epistemic Self-Undermining argument.[18]

Furthermore, this same Conciliatory reasoning is also likely to epistemically defeat other proposed Solution Responses to the self-undermining challenge. To support this claim, I will now provide a brief survey of other Solution Responses that have been proposed in this literature and show that the soundness of these arguments has been disagreed upon by epistemic peers, per premise II. For example, Bogardus (2009) argued that the EWV is not epistemically self-undermining because we can just see, via direct acquaintance, that the EWV is obviously true. This solution fails because many Conciliationists and Steadfast holders deny that the EWV is obviously true (e.g. Christensen (2013), Fumerton (2010), Kelly (2005, 2010, 2013), Weatherson (2013)).[19] Similarly, Frances (2010) argues that we should accept Conciliationism because it is closer to the truth than Steadfastness. This solution fails because, as Christensen (2013) argues, this closeness to the truth claim is denied by those who adhere to Steadfastness. Christensen (2009) argues that this self-undermining challenge is not very worrisome because it implausibly overgeneralizes to indict many other plausible Conciliatory principles (Matheson (2015a) argues similarly). This solution fails because Christensen (2013) himself later denies this response by arguing it does not adequately answer this challenge since epistemic self-undermining seems to unavoidably issue from the

[18] Of course, someone like Elga could deny premise I by arguing that the reasoning and auxiliary premises of his self-exempting argument are, like the EWV, also exempted from the EWV. But this response is extremely ad hoc and any argument for this will very likely also have reasoning or auxiliary premises that is disagreed upon by epistemic peers.

[19] Maybe it is more charitable to understand Bogardus as not providing a Solution Response but merely stipulating the EWV is true. If this is the case then this is of no help to those who wish to provide an adequate defense of the EWV from the epistemic self-undermining challenge.

commitments of Conciliationism. And, again, there are many more examples in this literature of epistemic peers denying some aspect of a given Solution Response to the epistemic self-undermining challenge.[20]

Therefore, I conclude that all potential Solution Responses to the epistemic self-undermining challenge are very likely to be epistemically defeated by the very view they are attempting to defend. To be clear, my Solution Responses (almost certainly) Fail argument is not meant to decisively refute all possible arguments against the epistemic self-undermining challenge. Instead, my argument places a heavy burden of proof on those who wish to defend the EWV from The EWV is Epistemic Self-Undermining argument. This is a burden of proof that I do not believe can be met. But, to foreshadow my argument's below, if one denies that all Solution Responses to this self-undermining challenge do or are likely to fail, §4 shows that issue is of secondary importance because as long as the EWV has skeptical consequences for some philosophical proposition q it will lead to its own falsity.

3. The EWV is Veridically Self-Undermining

Others have responded to this self-undermining challenge by providing a Skeptical Response which argues that while this The EWV is Epistemically Self-Undermining argument is sound, this does *not* show that the EWV is *false*. In other words, at best what this self-undermining challenge shows is that the EWV cannot

[20] Additionally, Pittard (2015) argues that self-exempting Conciliationism is not arbitrary because as long as Conciliationism is motivated by "epistemic deference" (449) to one's peers, then disagreement over Conciliationism gives the Conciliationist no reason to reduce confidence or suspend belief in Conciliationism itself. This solution fails according to Blessenhol (2015) because epistemic deference does not prevent self-exempting Conciliationism from being ad-hoc since epistemic deference cannot be justified by Conciliatory motivations; additionally, such epistemic deference leads Conciliationism to inconsistency in certain circumstances.

currently be justifiably believed or known to be true. The problem here is not that the EWV is false but that too many epistemic peers disagree that the EWV is true. For example, Christensen (2009: 763) says,

> Indeed, it seems to me those of us who find ourselves strongly drawn toward Conciliationism in these contentious times should not be confident that Conciliationism is correct. (Of course, we may still work hard in producing and disseminating arguments for the view, hoping to hasten thereby the day when epistemic conditions will brighten, consensus will blossom, and all will rationally and whole-heartedly embrace Conciliationism.)[21]

Similarly, Matheson (2015a:149) argues,

> If [EWV] is true, then in some possible circumstances individuals are not epistemically justified in believing some truth about epistemic justification. However, it is hard to see why this consequence should be thought to be a problem for [EWV]. It seems that there are plenty of truths (including truths about epistemic justification) such that given the way the world is, numerous individuals are not epistemically justified in believing them. This fact does not count against these truths and it certainly does not affect their status as truths.

I believe that this Skeptical Response only offers cold-comfort to defenders of the EWV (and Conciliationism).[22] This is because, as explained above, philosophical consensuses are rarely (if ever) reached, and so it is likely that the EWV will perennially epistemically

[21] Christensen (2013: 82) later denies this response by saying of this passage that "this degree of sanguinity underestimates the difficulty of the [epistemic self-undermining] problem considerably."

[22] In addition to Christensen and Matheson, see Decker (2014: 1132), Kornblith (2013: 274), and Littlejohn (2013: 178) for more examples of this Skeptical Response.

defeat itself and remain unjustified.[23] However, in this section I will argue that this defense of the EWV is mistaken because it *does* follow that *if* the EWV is epistemically self-undermining then the EWV is also false.

Specifically, I argue that *if* the EWV is epistemically self-undermining, then it is also *veridically* self-undermining. A view or proposition is veridically self-undermining iff this view or proposition entails its own falsity. For example, the proposition that "all universally quantified sentences are false" (Decker 2014: 1101) is veridically self-undermining because the truth of this proposition coupled with the fact that this proposition is a universally quantified sentence entails its own falsity. While the argument that the EWV is veridically self-undermining involves more steps than this argument, it nevertheless shows that the EWV is veridically self-undermining if it is epistemically self-undermining.[24]

To begin, the first five premises of this argument are the same as premises 1-5 in The EWV is Epistemically Self-Undermining argument above, which I have reproduced for convenience:

The EWV is Epistemically Self-Undermining

1. If peers disagree on p, then peers should suspend belief on p. (EWV)
2. If peers should suspend belief on p, then peers don't know that p. (EWV→~K)

[23] See Decker (2014: esp. 1127 and 1132) for an explanation of why epistemic self-defeat (his terminology) "puts the Conciliationist in a dialectically untenable situation" because such epistemic self-defeat is a "devastating problem" for Conciliationists.

[24] Of course, the proposition that "all universally quantified sentences are false" is *necessarily* veridically self-undermining because its being veridically self-undermining does not depend on anyone believing it. In contrast, the EWV is only *contingently* veridically self-undermining because its being veridically self-undermining depends on their being epistemic peers who disagree on it. It is a virtue of the veridic self-undermining definition that it allows for both of these.

3. Peers disagree on EWV. (Empirically true premise)
4. Thus, peers should suspend belief on EWV. (1, 3, MP)
5. Thus, peers don't know that EWV. (2, 4, MP)

To reiterate, this argument establishes, in premise 5, that peers do not know the EWV because there is peer disagreement on the EWV, per premise 3. The first step in my argument that the EWV is veridically self-undermining is to assume, for the time being, that it's empirically true that there is also peer disagreement about premise 5, i.e. there is peer disagreement on that peers don't know the EWV. Next, *if* there is such peer disagreement on premise 5, then it follows from premise 1 that peers should suspend belief on premise 5; and from premise 2 it further follows that peers do not know premise 5. In other words, peer disagreement on premise 5 results in the following meta-skepticism: peers don't know that peers don't know the EWV. However, I next argue that this meta-skepticism is inconsistent with the EWV because it seems that the EWV provides one with the resources to *know* premise 5. That is, I claim that if the EWV is right that peer disagreement provides one defeating evidence against one's beliefs, then when one suspends belief on premise 5 in accordance with the EWV, they should *know* that they don't know premise 5. But if this is correct, then the EWV leads to the following contradiction: peers know premise 5 and peers don't know premise 5.

For clarity, it will be useful to provide a formal representation of this argument. To this end, if we use parentheses to indicate the content of premise 5, this argument that the EWV is veridically self-undermining can be formalized as:

The EWV is Veridically Self-Undermining

6. Peers disagree on (peers don't know that EWV). (Empirically true premise)

7. Thus, peers should suspend belief on (peers don't know that EWV). (1, 6, MP)
8. Thus, peers don't know that (peers don't know that EWV). (2, 7, MP)
9. If peers should suspend belief on p, then when peers suspend belief on p *because of* peer disagreement on p, peers *know* that peers don't know that p. (Premise)
10. Thus, peers know that (peers don't know that EWV). (4, 9, MP)
11. Thus ~EWV. (8, 10, Explosion)

If sound, I claim this argument shows that *if* the EWV is epistemically self-undermining (via premises 1-5) then it also entails its own *falsity* (via premises 6-11) *contra* the Skeptical Responses from Christensen (2009), Matheson (2015a), and others. To establish this entailment I will have to show that premises 6 and 9 (i.e. the only two premises that are not logical consequences of previous premises) are either empirically true (6) or follow from the commitments of the EWV (9).

Beginning with 9, I claim that this premise is not something that the EWV holder can deny because it follows from their commitment to the EWV and their views on defeating evidence. To see why, it is important to recall (from §1) that, according to the EWV, peer disagreement on p provides *defeating* evidence against one's belief in p. Together with the claim in premise 9 that "when peers suspend beliefs on p *because of* peer disagreement on p," it follows that their recognized peer disagreement gives them sufficient justification to believe and know that believing p is defeated.[25] In other words, *if* the EWV is true and peer disagreement on p is defeating evidence against p, and peers do suspend belief on p *because of* peer disagreement on p in accordance with the EWV,

[25] More precisely, one is doxastically justified if one bases their belief on justifying reasons *in the right way*. For the EWV, this success condition is satisfied when agents base their belief that belief in p is defeated *because of* peer disagreement on p. See Turri (2010) for more on the basing relation and Kornblith (2017) for more on the nature of doxastic justification.

then it follows *a fortiori* that one knows that p is defeated. And furthermore, if you know your belief in p is defeated, then you know that you don't know p. For example, in the split-bill case, if the EWV is true and you suspend your belief that $43 is the correct split-bill amount in accordance with the EWV, they *a fortiori* you know that you don't know the correct split-bill amount.

If the EWV holder were to deny the conditional embodied in premise 9, this would mean that one's defeating evidence would defeat belief in p but would not support or justify the belief that belief in p is defeated. This is untenable because defeaters of a belief also support that this belief is defeated. In general terms, when you recognize and form the new belief that your original belief in p is defeated by some evidence E, if one were to ask you what supports this new belief, intuitively you would have to say that that E supports that your original belief is defeated. For example, to refer back to the split-bill case again, imagine you believe in accordance with the EWV that your original belief that $43 is the correct split-bill amount is defeated by peer disagreement. If I were to ask you what supports this new belief, intuitively you would have to say that peer disagreement on correct bill amount is what supports your new belief that your original belief is defeated. Therefore, premise 9 is not something that the EWV holder can deny since it follows *a fortiori* from their commitment to the EWV and peer disagreement providing *defeating* evidence.[26]

[26] Furthermore, EWV holders often appeal (if only implicitly) to something like premise 9. For instance, recall The EWV Leads to Philosophical Skepticism argument in §1. It stands to reason that *if* this argument is sound and one learns of this argument, then *a fortiori* they learn its conclusion that peers don't know q. And the same is true of The EWV is Epistemically Self-Undermining argument: *if* this argument is sound, as the Skeptical Responses to this argument assume, then *a fortiori* this argument provides one sufficient justification to know its conclusions, per premise 9. Additionally, the Skeptical Responses to the epistemic self-undermining argument quoted above from Christensen and Matheson above seem to appeal to something like premise 9

Premise 6 is something no one, including the EWV holder, can deny. This premise says that there is peer disagreement on premise 5, i.e. there is peer disagreement on whether peers don't know the EWV. This empirical premise is supported by the fact that there are many philosophers in the literature on the EWV who claim to know the EWV despite being aware that some of their philosophical peers believe ~EWV. Recall that §2 surveyed the many Solution Responses to the epistemic self-undermining challenge. This survey shows that there are philosophers who believe the EWV is false (and so believe that their peers don't know the EWV), and other philosophical peers who disagree about this. Therefore, as premise 6 says, peers disagree on peers don't know the EWV.

However, while one cannot object to premise 6 on empirical grounds, one might object that the disagreements embodied in this premise are epistemically inappropriate or irrational. One way to spell this objection out is to argue that: since The EWV is Epistemically Self-Undermining establishes that peers do not know the EWV (i.e. premise 5), such peers are not justified in continuing to believe the EWV; and, as such, the continued disagreement embodied in premise 6 carries no epistemic weight. In response, I argue that this denial of premise 6 is implausible because it amounts to the claim that such disagreements (i.e. peer disagreements on premise 5) are somehow *exempt* from the skeptical consequences of the EWV. But it seems that the EWV holders

---

to explain why they and others do not (currently) know the EWV. That is, their commitment to the EWV plus the large amount of peer disagreement on the EWV defeats their belief in the EWV. And given that they suspend belief in the EWV *because of* such peer disagreement it follows *a fortiori* that they know they don't know the EWV, per premise 9. Indeed, it is difficult to understand these above quotes from Christensen and Matheson if they are doing something other than explaining why they currently know that they don't know the EWV.

themselves cannot adequately justify this exemption because, as argued in §2, any argument to support this exemption will likely be disagreed upon by epistemic peers.

A more pressing objection to premise 6 is to argue that The EWV is Veridically Self-Undermining argument above assumes that the disagreements in premise 3 and 6 take place at the same time. But one could argue that the EWV is a *diachronic* rule rather than a *synchronic* rule for updating one's beliefs in light of peer disagreement. On this interpretation of the EWV, first philosophical peers discover that they disagree on the EWV in premise 3, and then they later disagree over whether they don't know the EWV in premise 6. If this were the right interpretation of the EWV, then no contradiction results in The EWV is Veridically Self-Undermining argument because the time-stamps on these premises would block the contradiction between premise 8 and 10.[27]

In response I believe, following Gardiner (2014) and Wilson (2010), that the diachronic interpretation of the EWV or Conciliationism is implausible because this interpretation commits it to denying the commutativity of evidence (i.e. the order in which evidence is acquired should not rationally influence what is reasonable to believe based on that evidence). Since I believe the commutativity of evidence is true, I consider its denial a fatal blow to the EWV and Conciliationism.

However, I additionally argue that *even if* we grant the diachronic interpretation of the EWV, there are many cases in which the disagreement in premise 3 takes place at the *same time* as the disagreement in premise 6. As such, as long as premise 3 is true, there is no plausible way to deny premise 6. Indeed, in many cases of peer disagreement premise 6 is a *logical consequence* of premise 3. This logical consequence occurs when

[27] I would like to thank Hilary Kornblith and Christopher Meacham for independently raising this objection.

those who disagree about the EWV also each, respectively, believe they know or don't know the EWV. To illustrate, consider the following understanding of what philosophical peer disagreement amounts to: *before* they recognize their disagreement on the EWV, peer1 believes they know the EWV and peer2 believes they know ~EWV. These beliefs entail that: peer1 believes they know ~EWV is false and peer2 believes they know the EWV is false. And if we assume knowledge is *factive* (i.e. you cannot know something that is false), these beliefs entail that: peer1 believes no one knows ~EWV and peer2 believes that no one knows the EWV. So when these peers recognize their disagreement on the EWV, peer1 believes peer2 doesn't know ~EWV and peer2 believes that peer1 doesn't know the EWV. Thus, when such peers disagree on the EWV, there is one peer (peer2) who believes the other doesn't know the EWV and another (peer1) who disagrees about this. Thus, when such peers disagree on the EWV (premise 3), they also disagree on peers not knowing the EWV (premise 6). So, contra the above objections, there is no plausible way to deny premise 6.

Therefore, contra Christensen (2009), Matheson (2015a), and others, Skeptical Responses to the epistemic self-undermining challenge fail because *if* the EWV is epistemically self-undermining then the EWV is also false because it is veridically self-undermining.

4. Reductio of the EWV

In the previous two sections I have argued that both Solution and Skeptical Responses to the epistemic self-undermining challenge fail. For clarity, my overall argument thus far can be summarized as:

A. The EWV is (almost certainly) epistemically self-undermining.

a. (via the Solution Responses (almost certainly) Fail argument in §2)

B. If the EWV is epistemically self-undermining, then it is false.

a. (via The EWV is Veridically Self-Undermining argument in §3)

C. Thus, the EWV is (almost certainly) false.

In other words, neither Solution nor Skeptical Responses save the EWV from being either epistemically or veridically self-undermining.

While I believe that the above argument decisively refutes the EWV, one might also object that The EWV is Veridically Self-Undermining argument *depends* on The EWV is Epistemically Self-Undermining argument. That is, one could maintain that, despite my arguments against providing a Solution Response to the epistemic self-undermining challenge in §2, the EWV avoids these skeptical consequences (e.g. because Elga (2010) is correct that the EWV is self-exempting) and thereby also prevents it from being veridically self-undermining.

While I believe my Solution Responses (almost certainly) Fail argument in §2 refutes this kind of response, I grant that *if* there is such a successful Solution Response to the epistemic self-undermining challenge that the EWV would also not be veridically *self*-undermining (since now the EWV would not lead to its *own* falsity). But I argue that this issue is of secondary importance. *Even if* there is such a solution, the EWV is nevertheless still refuted by the same type of reasoning in The EWV is Veridically Self-Undermining argument. In this section I will show that the EWV leads to a hitherto unrecognized *reductio ad absurdum* which side-steps the issues surrounding self-incrimination and provides an independent refutation of the EWV.

To see this, recall the original argument that The EWV Leads to Philosophical Skepticism in §1 above, which I have reproduced for convenience:

The EWV leads to Philosophical Skepticism

1. If peers disagree on p, then peers should suspend belief on p. (EWV)
2. If peers should suspend belief on p, then peers don't know that p. (EWV→~K)
3. Peers disagree on q. (Empirically true premise)
4. Thus, peers should suspend belief on q. (1, 3, MP)
5. Thus, peers don't know that q. (2, 4, MP)

This argument uncontroversially demonstrates how the EWV leads to philosophical skepticism for some philosophical propositions q. The controversy surrounding whether the EWV is epistemically self-undermining arises because the EWV is a philosophical proposition and seemingly applies to and epistemically defeats itself. All of the potential Solution Responses to the epistemic self-undermining challenge mentioned in §2 attempt to explain how the EWV can avoid skeptical consequences *for itself*, e.g. by arguing that it is inappropriate to plug-in 'EWV' into 'q' in the argument above.

However, I argue that this entire debate can be side-stepped by using a similar but more general *reductio ad absurdum* argument. This reductio argument uses the same type of reasoning as The EWV is Veridically Self-Undermining argument in §3 but does not rely on The EWV is Epistemically Self-Undermining argument. This reductio arguments begins by assuming, like The EWV is Veridically Self-Undermining argument, that there is peer disagreement on *this* premise 5, i.e. there is peer disagreement on that peers don't know that *q*. Next, it argues that *if* there is such peer disagreement, then if follows via premises 1 and 2 that peers don't know premise 5. In other words, such peer

disagreement results in the following meta-skepticism: peers don't know that peers don't know that q. But, just like before, I argue that this meta-skepticism is inconsistent with the EWV because the EWV provides one with the resources to know premise 5. And if this is correct, then the EWV leads to the following contradiction: peers know premise 5 and peers don't know premise 5.

For clarity, it will also be useful to provide a formal representation of this reductio argument. To this end, if we again use parentheses to indicate the content of premise 5, this general reductio argument can be formalized as:

Reductio of the EWV

6*. Peers disagree on (peers don't know that q). (Empirically true premise)

7*. Thus, peers should suspend belief on (peers don't know that q). (1, 4, MP)

8*. Thus, peers don't know that (peers don't know that q). (2, 7*, MP)

9*. If peers should suspend belief on p, then when peers suspend belief on p *because of* peer disagreement on p, peers *know* that peers don't know that p. (Premise)

10*. Thus, peers know that (peers don't know that q). (4, 9*, MP)

11*. Thus, ~EWV. (8*, 10*, Explosion)

It is evident that this reductio argument uses the same type of reasoning as The EWV is Veridically Self-Undermining argument in §3. Indeed, the Reductio of the EWV argument above is *identical* to The EWV is Veridically Self-Undermining argument except that 'EWV' has been replaced with any philosophical proposition 'q'. This is a crucial and significant difference between these two arguments for two reasons. Firstly, this difference shows that the Reductio of the EWV argument above uses the same type of reasoning as The EWV is Veridically Self-Undermining argument. As such, premises

6* and 9* in the Reductio of the EWV argument above are supported *mutatis mutandis* by the same considerations that support premises 6 and 9 in The EWV is Veridically Self-Undermining argument. To briefly reiterate, like premise 9, premise 9* is not something the EWV holder can deny since it also follows from their commitment to the EWV and their views on defeating evidence. And like premise 6, 6* is an empirically true premise that philosophical peers disagree about whether they know certain philosophical propositions q (e.g. see the illustrative example in the following paragraph); the disagreements embodied in premise 6* cannot plausibly be *exempted* from the skeptical consequences of the EWV, and that in many cases of peer disagreement premise 6* a logical consequence of premise 3 (peers disagree on q).

To illustrate this last point again, consider the case of two philosophical peers regarding the free will debate. Peer1 believes they know Libertarianism (L) and peer2 believes they know ~L, e.g. because peer2 is a Compatibilist. *Before* they recognize their disagreement on L, their beliefs entail that: peer1 believes they know ~L is false and peer2 believes they know L is false. Assuming knowledge is *factive*, these beliefs entail that: peer1 believes no one knows ~L and peer2 believes that no one knows L. So when these peers recognize their disagreement on L, peer1 believes peer2 doesn't know ~L and peer2 believes that peer1 doesn't know L. Thus, when such peers disagree on L, there is one peer (peer2) who believes the other doesn't know L and another (peer1) who disagrees about this. Thus, when such peers disagree on L (as in premise 3), they also disagree on peers not knowing L (as in premise 6*). As such, as long as premise 3 is true, there is no plausible way to deny premise 6*. Therefore, this Reductio of the EWV

argument uses the same type of reasoning as The EWV is Veridically Self-Undermining argument.

Secondly, because the Reductio of the EWV argument is about any philosophical proposition 'q' and not the 'EWV', this reductio argument concerns the skeptical consequences that the EWV has for *any* peer disagreed upon philosophical propositions q and *not* just the skeptical consequences the *EWV* has *for itself*. In other words, the above reductio argument does not depend on the argument that The EWV is Epistemically Self-Undermining argument to reach its conclusion. As such, the Reductio of the EWV argument side-steps the entire debate regarding whether the EWV is epistemically and veridically self-undermining. Indeed, this is illustrated by the example of peer disagreement over Libertarianism above. As long as there is peer disagreement on L, 'q' can be replaced with 'L' in the Reductio of the EWV argument above. And while this L version of my reductio argument does not depend on The EWV is Epistemically Self-Undermining argument, it nevertheless shows that the EWV leads to a contradiction. In other words, this reductio argument shows that *if* the EWV leads to philosophical skepticism for some philosophical proposition q (via premises 1-5), then the EWV also entails its own *falsity* (via premises 6*-11*).

Therefore, the Reductio of the EWV argument side-steps the entire debate about whether the EWV is epistemically and veridically self-undermining. Consequently, *even if* the *EWV* avoids being epistemically self-undermining and veridically self-undermining, the EWV is still refuted by the same type of reasoning, via the Reductio of the EWV argument above. Therefore, the EWV is false because the above reductio argument shows that it leads to a contradiction in many cases of peer disagreement on some philosophical

propositions q. Consequently, philosophical knowledge is saved from the disagreement skepticism posed by The EWV Leads to Philosophical Skepticism argument in §1.

5. Reductio of Conciliationism

One might object that my arguments against the EWV above only apply to versions of Conciliationism that involve "suspending belief" but not to "degrees of confidence" versions of Conciliationism. In this section I will show how my Reductio of the EWV argument can be modified to apply to all plausible degree of confidence versions of Conciliationism. As such, I argue that all plausible versions of Conciliationism are false because they also lead to a version of my reductio argument.

My general argument in this section poses the following dilemma: either Conciliationism has the skeptical consequences for some philosophical proposition q or it doesn't. If it does not, then philosophical disagreement skepticism does not follow from Conciliationism and philosophy is saved from Conciliationism disagreement skepticism. However, this horn of the dilemma is untenable because, as I will argue below, all plausible versions of Conciliationism (i.e. versions that require a *significant* or large reduction of confidence in cases of peer disagreement) have skeptical consequences for some q. But like my Reductio of the EWV in §4, I argue that a version of this argument follows for Conciliationism as long as Conciliationism has skeptical consequences for some q. Either way, philosophy is saved from Conciliationism disagreement skepticism.

To accomplish this, I will first explain some general features of Conciliationism and show how to represent the general argument that Conciliationism leads to philosophical skepticism. Second, I will show how an *im*plausible version of Conciliationism (i.e. one that only requires an *in*significant or minuscule reduction of

confidence in cases of peer disagreement) uses this general Conciliatory argument to argue for philosophical skepticism for some q. Third, I will show how this implausible version of Conciliationism succumbs to my reductio argument. From this example it will be clear that the same reductio argument can be made for any plausible version of Conciliationism.

To begin, there are two salient features of Conciliationism: an original degree of confidence in some proposition p and subsequent reductions of confidence in p in cases of peer disagreement. Confidence values in p are on a scale of 0-1 where 1 is complete confidence and 0 is no confidence. As such, we can represent of all degree of confidence versions of Conciliationism as having the following form: In cases of peer disagreement on p, agents are rationally required to reduce their original confidence in p by some degree d – where d is some *non-zero* amount between 0 and 1. Furthermore, it is natural to assume that even in degree of confidence versions of Conciliationism that there will be a threshold where the following will be true: If one's degree of confidence in p falls below some point between 0 and 1, then rationality requires one to suspend belief in p. For convenience, let's just arbitrarily stipulate that having a confidence of .7 or greater is necessary for agents to retain their belief in p.

With these stipulations, imagine you have .7 confidence in some philosophical proposition q. Now further imagine that in some case C, you come across a philosophical peer who disagrees with you about q, i.e. they are confident .7 or greater in ~q. In case C, any Conciliatory view is going to dictate that you should reduce confidence in your belief to be below .7. This is because the degree d by which you have to reduce your confidence

by is a *non-zero* amount. And given our stipulations that .7 or greater is required to retain one's belief, in case C you can no longer retain your belief in q.

We can represent this general type of Conciliatory skeptical argument by using similar reasoning to The EWV Leads to Philosophical Skepticism argument in §1 and §4 above:

Conciliationism Leads to Philosophical Skepticism

i. There is some case C, if peers disagree on p, then peers should suspend belief on p. (Conciliationism)[28]
ii. If peers should suspend belief on p, then peers don't know that p. (Conciliationism→~K)
iii. In C, peers disagree on q. (Empirically true premise)
iv. Thus, peers should suspend belief on q. (i, iii, MP)
v. Thus, peers don't know that q. (ii, iv, MP)

I claim that any plausible version of Conciliationism uses this argument to establish philosophical skepticism. To see why this is the case, consider the following *im*plausible version of Conciliationism called *.0001-Conciliationism*: In cases of peer disagreement on p, agents are rationally required to reduce their confidence in p by .0001. This version of Conciliationism is implausible because it seems that no good rationale can be given to justify why one should only reduce their confidence in such cases by such a minuscule

[28] This is also a simplified statement of Conciliationism for the same two reasons mention in fn. 10 about the EWV. For my purposes I can also safely set aside these complications (i.e. the antecedent usually has more conditions and the scope of Conciliationism sometimes only quantifies over *most p*) since presumably these complications are not applicable to (nearly) all cases of philosophical peer disagreement. Indeed, recalling my argument in §2, it is not clear that philosophical disagreement or any kind can be exempted from the skeptical consequences of Conciliationism since any argument for such an exemption will likely be disagreed upon by epistemic peers.

amount. However, *even if* .0001-Conciliationism arbitrarily and implausibly puts in .0001 for degree d, in the *case C* where I have confidence of .7 in q and then a peer disagrees by believing ~q to a confidence level of .7 or greater, I can no longer retain my belief in q. So even .0001-Conciliationism can lead to philosophical skepticism for some q using the above argument. Likewise, it is clear that plausible versions of Conciliationism (e.g. ones that have a much larger number for degree d) will also lead to philosophical skepticism in case C by using the above argument.

Like the previous two sections of this chapter, I claim that a version of my reductio argument arises for this Conciliationism Leads to Philosophical Skepticism argument:

Reductio of Conciliationism

vi. In C, peers disagree on (peers don't know that q). (Empirically true premise)

vii. Thus, peers should suspend belief on (peers don't know that q). (i, vi, MP)

viii. Thus, peers don't know that (peers don't know that q). (ii, vii, MP)

ix. If peers should suspend belief on p, then when peers suspend belief on p *because of* peer disagreement on p, peers *know* that peers don't know that p. (Premise)

x. Thus, peers know that (peers don't know that q). (iv, ix, MP)

xi. Thus, ~Conciliationism. (viii, x, Explosion)

This argument is just like the Reductio of the EWV argument above except that it uses degree of confidence versions of Conciliationism instead of the EWV. As such, the reasoning used in the above reductio argument is the same type of reasoning as in the Reductio of the EWV argument; and premises vi and ix in the above reductio argument are supported *mutatis mutandis* by the same considerations that support the previous

iterations of these premises. To briefly reiterate, like premise 9* in the Reductio of the EWV argument, premise ix is not something that the Conciliationist can deny since it follows from their commitment to Conciliationism and their views on defeating evidence. And like premise 6* in the Reductio of the EWV argument, premise vi above is an empirically true premise that philosophical peers disagree about whether they know certain philosophical propositions q; the disagreements embodied in premise vi cannot plausible be *exempted* from the skeptical consequences of the EWV; and in many cases of peer disagreement premise vi is just a logical consequence of premise iii (i.e. recall the Libertarianism example). As such, the Reductio of Conciliationism argument above is just as cogent as the Reductio of the EWV argument.

Furthermore, it is easy to see that .0001-Conciliationism is refuted by the above reductio argument. In the case C where I have .7 confidence in q and then a peer disagrees by believing ~q to a confidence level of .7 or greater, .0001-Concilationism dictates that I reduce my confidence in q below the .7 threshold that is required to retain my belief (via premises i-v). And, as explained in the previous paragraph, since premises vi is empirically true and premise ix follows from of Conciliationism, a version of my reductio argument arises and refutes .0001-Concilationaism (via premises vi-xi).

Additionally, it is also clear that this same reductio argument can be made using any stipulated belief confidence threshold to refute any plausible version of Conciliationism (e.g. ones that have a much larger number for degree d). As long as Conciliationism allows for there to be some case C where peer disagreement on q rationally requires a reduction of confidence sufficient to warrant skepticism in q (via premises i-v), then this skepticism will lead to a *reductio ad absurdum* (via premises vi-

xi). Therefore, all plausible versions of Conciliationism are false because any (even implausible) values that can be plugged into d will lead to skepticism for some q and for some stipulated belief confidence threshold.

6. Extending the Reductio Argument

To summarize this chapter thus far, §1 explained how the EWV leads to philosophical skepticism; §2 argued that the EWV is (almost certainly) epistemically self-undermining; §3 argued that if the EWV is epistemically self-undermining then it is also veridically self-undermining; §4 argued that the EWV is subject to a previously unrecognized reductio argument; and §5 argued that this reductio argument applies to all plausible versions of Conciliationism. I take the all of the above arguments to decisively refute the EWV and, more broadly, all plausible version of Conciliationism. Consequently, I conclude that the threat that these view pose for philosophical knowledge has been neutralized, and that philosophical knowledge is saved from disagreement skepticism.

In this section I will extend the results of the previous section and argue that other competing views to Conciliationism are false because they also lead to a version of my reductio argument. To accomplish this, it is important to note that while the above reductio argument against Conciliationism is cast in terms of peer disagreement over philosophical propositions q, the argument also works just as well for other non-philosophical propositions p. For example, the proposition that $43 is the correct split-bill amount can be plugged-into the reductio of Conciliationism argument in the previous section and a contradiction would follow that also refutes Conciliationism. This is very counterintuitive because this means the intuitive response to conciliate in the split-bill

case, as explained in the introduction, is wrong. As Christensen (2011: 2) rightly notes "even opponents of Conciliationism typically concede that I should become much less confident that my share is $43, and indeed should not be significantly more confident in $43 than in $45." I will now show that any rival views to Conciliationism which dictate that suspension of belief is required in cases like the split-bill case both affirm premise i and consequently leads to a version of my reductio argument.

One opponent to Conciliationism is Lackey (2010a,b) who articulates what she calls the Justificationist View (JV) of disagreement. According to Lackey, the JV is a rival theory to Conciliationism because this theory denies that we should always be conciliatory in the face of peer disagreement. Lackey argues that it is permissible to remain Steadfast in one's beliefs if one has a high degree of antecedent justification or justified confidence in one's beliefs. For example, imagine a version of the split-bill case were you did not just quickly and unreliably do the calculations in your head but you used a calculator and/or a pen and paper to slowly and reliably arrive at $43 as the split-bill amount. Lackey argues that in this version of the split-bill case it is permissible to remain Steadfast in your belief because you have a high degree of antecedent justification and confidence that you have arrived at the correct amount. In short, the JV holds that remaining Steadfast in one's beliefs can be permissible as long as one has a high-degree of antecedent justification and confidence for their beliefs.[29]

I should note that the JV has a *prima facie* plausible way to prevent philosophical disagreement skepticism. Because philosophical peer disagreement is often more like the

[29] Christensen (2011) argues that Conciliationism can accommodate these intuitions about the careful checking split-bill case, while Vavova (2014b) argues that the JV is false because it is subject to counterexamples.

careful checking split-bill case, it can be permissible for philosophers to Steadfastly retain their beliefs as long as they have a high degree of antecedent justification for their beliefs. But, however intuitively plausible the JV is, it still leads to my reductio argument. This is in large part because it affirms premise i that there are some cases of peer disagreement in which one should conciliate enough to suspend belief on some p. As Lackey (2010b: 319) writes,

> If, however, my belief that *p* does not have a high degree of justified confidence prior to the disagreement, then the force of the disagreement can do the defeating work in question, thereby requiring substantial doxastic revision on my part.

And Lackey (2010a,b) also says that one should suspend belief in the original split-bill case because in this case one does not have a high degree of antecedent justification and confidence prior to the peer disagreement. For my purposes, the split-bill case provides an example of a non-philosophical proposition p that can be plugged into premise iii. I further claim, like the arguments in the previous two sections of this chapter, that premise vi will follow as a logical consequence of premise iii and that premise ix follows from commitment to premise i and plausible assumptions about defeating evidence. Thus, a version of my reductio argument arises and refutes the JV.

Another opponent to Conciliationism is Kelly (2010, 2013) who argues for what he calls the Total Evidence View (TEV) of disagreement. According to Kelly, the TEV is also a rival theory to Conciliationism because this theory denies that we should always be conciliatory in the face of peer disagreement. Kelly argued that it is permissible to remain Steadfast in one's beliefs if one's total evidence supports their beliefs. For example, imagine I have been considering some p for some time and have amassed many

arguments and reasons in its favor when I find out that an epistemic peer disagrees and believes ~p. At the moment of recognized disagreement I can retain my belief in this case since this is only one piece of evidence against my belief in p. In this case, this peer disagreement does not defeat my belief in p because it is only one piece of evidence against my well-supported belief in p.[30]

Like the JV, I should also note that the TEV has a *prima facie* plausible way to prevent philosophical disagreement skepticism: in many cases of philosophical peer disagreement one's total evidence will likely favor their philosophical beliefs because they will have many arguments and evidence in favor of these beliefs that outweigh the peer disagreement evidence against their beliefs. But, again, however intuitively plausible the TEV is, it also leads to a version of my reductio argument in large part because it affirms premise i that there are some cases of peer disagreement in which one should conciliate enough to suspend belief on some p. As Kelly (2010:167) writes,

> As we have seen, the Total Evidence View and the Equal Weight View yield the same verdict when applied to some cases of disagreement.

Kelly (2010) then cites the original split-bill case (fn. 39) and similar cases involving mathematical calculation (167, 137) to illustrate cases in which the TEV dictates that one should suspend belief. Again, for my purposes the split-bill case provides an example of a non-philosophical proposition p that can be plugged into premise iii. I again further claim, like the arguments in the previous two sections of this chapter, that premise vi will follow as a logical consequence of premise iii and that premise ix follows from

[30] Of course, as Kelly (2010: 154) notes, one often should investigate the reasons behind such peer disagreement. But as Kelly also notes, it often can be reasonable to retain one's belief even after investigating the reasons behind their disagreement because in many cases one's total evidence will still favor one's belief.

commitment to premise i and plausible assumptions about defeating evidence. Thus, like the JV, a version of my reductio argument arises and refutes the TEV.

7. Conclusion

While my reductio argument saves philosophy from disagreement skepticism, it also refutes all plausible versions of Conciliationism and refutes rival views (even prima facie plausible ones) that allow for suspension of belief in some cases of peer disagreement. This is a very counterintuitive result. To make matters worse, the reductio argument also seems to imply that the intuitive response to suspend belief in cases where *many* peers disagree with you is also wrong.

To make this point clear, imagine that one holds the following view on peer disagreement: one should suspend belief on p if n peers disagree and believe ~p, where $n \geq 17$. In other words, agents are rationally required to suspend belief if they discover that 17 or more peers hold a contradictory belief. Let's call this 17-1 Conciliationism. For instance, let's change the split-bill case to be that you are out to dinner with seventeen other friends who are your mathematical peers, and they all come to believe that $45 is correct amount while you alone think $43 is the correct amount. Even in this case it seems that my reductio argument will also arise and refute 17-1 Conciliationism. In this version of the argument, case C would be 17-1 peer disagreement in premises i, iii, vi, and ix. And by the same reasoning in the previous reductio arguments, once a Conciliatory view has skeptical consequences for believing some p, this view will lead to a contradiction. But since 17 is just an arbitrarily chosen number; any numbers for n-to-1 Conciliationism will be subject to my reductio argument. In short, it appears that my reductio argument has the following counterintuitive consequence: no matter how many

peers disagree, peer disagreement *never* rationally requires that agents suspend belief or reduce confidence in *any* of their contested beliefs.

I acknowledge that this is an extremely counterintuitive consequence of my reductio argument. To my knowledge, there is only one theory of peer disagreement that explicitly accepts this implication, its counterintuitive consequences, and, consequently, avoids my reductio argument: the Right Reasons View (RRV) (e.g. Kelly (2005), Titelbaum (2015)). The RRV says that peer disagreement itself does not rationally require peers to revise their beliefs in any way. Rather, peers should revise their beliefs only in accordance with what the non-disagreement reasons and evidence actually support. For example, if you have correctly calculated the split-bill amount, then since your belief is supported by this reasoning, you should retain your belief and your peer should revise their belief since their reasoning does not support their belief. This is true even in one vs. many peer disagreement cases. Kelly (2005: 192-193) says that no matter how many epistemic peers disagree with you about some p, it is rational to stick to one's guns as long as your beliefs are supported by what the reasons and evidence actually support. According to the RRV peer disagreement at best only serves as "proxy" (Kelly 2005: 187) for one's underlying reasons and evidence for their beliefs. Agents should always look to the underlying non-disagreement evidence and reasons before revising their beliefs.

It is easy to see that the RRV rejects premise i of my reductio argument and as such avoids this argument. For this reason, my reductio argument then can be thought to be an indirect argument for this theory. However, because the RRV rejects premise i it is extremely counterintuitive and is consequently not accepted by nearly all epistemologists.

Kelly (2010, 2013) himself abandons the RRV in favor of the TEV; while Titelbaum (2015) acknowledges that the RRV is extremely counterintuitive for the reasons given above regarding one vs. many peer disagreements. Indeed, I agree that retaining one's belief in such cases is hopelessly epistemically immodest and dogmatic because it fails to appreciate one's own fallibility and seeming evidence that one is mistaken. Furthermore, because Conciliationist intuitions about epistemic modesty are very robust and shared by a great many (nearly all) philosophers who write on the epistemic significance of peer disagreement,[31] it is incumbent that these intuitions should be explained and accounted for in light of my reductio argument. However, I do not (currently) see any way to accomplish this. Usually, this would be taken by many philosophers to be good grounds to believe that there is something wrong with my reductio argument. But since I see no fault in its reasoning and its premises seem to be true, it is also incumbent on those who wish to defend Conciliationist intuitions and views that allow for Conciliatory reasoning to find a flaw in the argument. Until a flaw is discovered or a way to accommodate these intuitions is found, I find these results deeply counterintuitive.

[31] See fn. 2, 5, and 7.

CHAPTER 4

MORAL DISAGREEMENT SKEPTICISM LEVELED

One of the oldest and most discussed challenges to moral knowledge is the argument from moral disagreement to moral skepticism.[1] In its most general form, this argument reasons that moral knowledge is unattainable as long as there is (some kind of) disagreement about moral matters. But, as many authors have acknowledged, "the" argument from moral disagreement to moral skepticism is a misnomer since there are many distinct arguments in this literature that attempt to use moral disagreements to establish moral skepticism.[2] Some versions argue that moral disagreements undermine the truth of moral claims. For instance, Mackie (1977) argues that the best explanation for the widespread moral disagreements between different cultures is that there are no moral truths; and since knowledge requires truth, moral knowledge is impossible.[3] Other versions argue that moral disagreements challenge the accessibility of knowing moral facts. For instance, Risberg and Tersman (2019) argue that the possibility of moral disagreement between ideal agents in ideal conditions shows that, even in the most favorable epistemic conditions, we lack the ability to know moral facts.[4]

---

[1] See Gowans' (2000a) anthology for a historical overview and Campbell (2015).

[2] Cf. Tersman (2006: xiii) and Enoch (2009: 16; 2011: 185).

[3] See Joyce (2018) for a discussion of Mackie's argument.

[4] See Tolhurst (1987) and Rowland (2017a) for similar moral disagreement arguments that involve ideal conditions.

Relatedly, other versions use moral disagreement to show that moral beliefs are not justified. For instance, many have argued that moral disagreements between epistemic peers (i.e. those who are about as equally well-positioned to acquire moral knowledge) defeats one's justification for their moral beliefs; and since knowledge requires undefeated justification, peer disagreement blocks moral knowledge.[5] While there is also a large literature on the threat that peer disagreement poses to knowledge in other (non-moral) domains, this chapter focuses on the threat that moral disagreements pose for *moral* knowledge. And while this chapter does discuss the general peer disagreement literature (see §2), focusing on moral disagreement arguments is dialectically well-founded because, as just explained, there are many additional and varied kinds of moral disagreement to moral skepticism arguments that do not rely on peer disagreement. So ironically, focusing on moral disagreements arguments allows this chapter to have a wider scope and cover more argumentative ground than if it just focused on the threat posed by peer disagreement to moral (and non-moral) knowledge.[6]

But regardless of the precise way in which moral disagreements establish moral skepticism,[7] it appears that there are two necessary conditions for any moral disagreement argument. Firstly, such arguments claim that there is moral disagreement

[5] See Vavova (2014), Sampson (forthcoming), Rowland (2017b), McGrath (2008), Wedgwood (2010), Setiya (2012), Fritz and McPherson (2019), and Fritz (2019) for discussions of this version of the moral disagreement argument.

[6] Indeed, the general disagreement literature has focused almost exclusively on the skeptical implications of Conciliationism – the view that in the face of peer disagreement one should suspend belief or reduce confidence in one's contested beliefs (e.g. see Christensen (2009), Kornblith (2013), Elga (2010), and Palmira (2019)). See Palmira and Stroud (2019) for a general discussion of the many additional kinds of disagreement arguments.

[7] See Loeb (1998), Bennigson (1996), Leiter (2014), Mogensen (2017), and Wedgwood (2019) for additional versions of moral disagreement arguments.

(of some kind) over moral proposition(s) p; call this the *Moral Disagreement Thesis* (MDT). Secondly, such arguments claim that moral disagreement (of some kind) on p is (in some way) sufficient to establish moral skepticism on p is true; call this the *Sufficiency Thesis* (ST).[8] In this chapter I argue that any version of the moral disagreement argument that satisfies both MDT and ST cannot establish moral skepticism because together these claims lead to a previously unrecognized *reductio ad absurdum*. Furthermore, I argue that because it is very plausible that all versions of the moral disagreement argument satisfy both MDT and ST that this reductio argument (very likely) refutes all versions of this argument in one fell swoop.

§1 defines key terms and then presents a compelling version of the moral disagreement argument; §2 shows how this compelling version leads to a reductio; §3 defends this reductio argument from possible objections; §4 shows how any version of the moral disagreement argument that satisfies MDT and ST leads to a reductio; and §5 explains why this general reductio argument (very likely) refutes all moral disagreement arguments.

§1

In this chapter 'moral disagreement(s)' refers to any disagreement over moral propositions. And 'moral proposition(s)' refers to propositions that have moral content. These definitions allow first-order propositions like "abortion is immoral" and "you should not kick puppies" as well as second-order propositions like "I know abortion is immoral" and "moral relativism is false" to count as *moral* propositions. As such, these definitions also allow for disagreements within applied, normative, and metaethics to

[8] While I believe that these conditions are also jointly sufficient for any moral disagreement argument, my arguments in this chapter only rely on the necessity claims.

count as *moral* disagreements. One might object to this terminology on the grounds that metaethical propositions and disagreements are really *about* morality, and therefore should not be considered *moral* disagreements. However, in this chapter nothing of consequence hangs on choosing this terminology over mine since it is widely accepted that disagreements over metaethical and non-moral propositions can also lead to skepticism.[9] My choice of terminology just allows for a convenient and parsimonious way to refer to disagreement skepticism that can result for both first-order moral and metaethical propositions.

Additionally, throughout this chapter 'sufficient to establish' is used to refer to any kind of reasoning that a moral disagreement argument might use to reach its (allegedly) *true* conclusion (e.g. *sound* deduction, cogent induction or inference to the best explanation with *true* conclusions, etc.), and to refer to any way in which moral disagreements might lead to moral skepticism (e.g. such disagreements undermine the truth of moral claims (i.e. Mackie (1977)), challenge the accessibility of moral facts (i.e. Risberg and Tersman (2019)), undermine one's justification (i.e. via peer disagreement), etc.). The purpose of this terminology is to allow many different versions of "the" moral disagreement to moral skepticism argument to be referred to at once without having to specify "the" type of reasoning or the precise way in which these arguments reach their conclusions. As such, rather than just focusing on one kind of moral disagreement to moral skepticism argument (e.g. via peer disagreement), this terminology allows my analysis of moral disagreement arguments to cover more and varied argumentative ground within the larger category of moral disagreement arguments.

[9] This is further discussed in §2, but also see fn. 6, fn. 12, and fn. 13.

And while there are many different versions of "the" moral disagreement argument, not all versions are equally compelling. Consider the following formulation of this kind of argument, where 'C1' and 'C2' are different cultures and 'p' is a particular moral proposition:

Moral Disagreement between Cultures

A. Moral disagreement between C1 and C2 about p is sufficient to establish that no one in C1 or C2 knows that p.

B. There is moral disagreement between cultures C1 and C2 about p.

C. Thus, no one in C1 or C2 knows that p. (A, B)

Many have argued that this and similar versions of the moral disagreement argument are unconvincing since, without knowing more about these cultures and the kind of disagreement between them, there are many convincing ways to resist this argument's conclusion. Firstly, if C1 is full of people who exhibit some sort of cognitive shortcoming (i.e. they are irrational, uninformed, conceptually confused, etc.), then moral disagreement with C1 is intuitively not sufficient to establish moral skepticism about p. Secondly, if the disagreement between these cultures is one that will be resolved in short order (e.g. because these cultures are amenable), then at best this moral disagreement only provides a short-lived moral skepticism. Lastly, if these cultures only disagree about a few moral propositions, then their moral disagreements do not pose an extensive challenge to moral knowledge. For these (and other) reasons, it is generally agreed that this version of the moral disagreement argument (and others like it) does not pose a serious challenge to moral knowledge.[10]

[10] See Tersman (2006: chap. 2), Enoch (2009: 21-29; 2011: 187-196), and Joyce (2018) for the difficulties in determining when moral disagreements are a genuine threat to moral knowledge.

A more compelling version of "the" moral disagreement argument takes something like the following form:

Moral Disagreement among Ethicists

1. Intractable moral disagreement among ethicists about p is sufficient to establish that ethicists don't know that p.
2. There is intractable moral disagreement among ethicists about p. (Empirically true premise)
3. Thus, ethicists don't know that p. (1, 2)

Prima facie, this version of the argument is more plausible because it seems to avoid the problems of the previous version. Firstly, ethicists are (presumably) cognitively competent moral agents (i.e. they are not irrational, uninformed, conceptually confused, etc.). Indeed, they are (presumably) in the best position to acquire moral knowledge since they spend their professional lives attempting to acquire it. But secondly, despite this prolonged study, ethicists seem to be perennially unable to resolve their moral disagreements. Indeed, Bourget and Chalmers (2014) found that there is no clear consensus among contemporary ethicists about the following perennial moral issues: which kind of normative ethical theory is correct (i.e. Deontology, Consequentialism, or Virtue Ethics) and whether metaethical realism or antirealism is true. And, lastly, I take it as obviously true that ethicists disagree about a great many other moral propositions (e.g. abortion, euthanasia, affirmative action, etc.).[11] In short, this version of the moral disagreement argument poses a more serious challenge to moral knowledge.

[11] Specifically, it is widely accepted that ethicists disagree about a great many *philosophical* moral propositions (cf. Vavova 2014: 323; Leiter 2014: 141-142). Furthermore, these facts about the *actual distribution* of moral disagreements among ethicists allow me to set aside the issue

§2

But despite the Moral Disagreement among Ethicists argument's *prima facie* plausibility, in this section I will argue that this argument leads to a *reductio ad absurdum.* This reductio builds upon one common challenge to moral disagreement arguments: they are *epistemically self-defeating*.

An argument is epistemically self-defeating iff at least one of this argument's components (i.e. premises, conclusions, or reasoning) applies to and epistemically defeats at least one such component. In the moral disagreement literature, discussions of the self-defeat challenge usually take the following form: if there is also disagreement about the *premises* of a moral disagreement argument, the skeptical implications of this argument will likely apply to these premises and epistemically defeat this argument. For example, Shafer-Landau (2006: 218-221; 2012: 328), Decker and Groll (2013: 157), and Sampson (forthcoming) argue that because the premises of moral disagreement arguments are also disagreed upon, such arguments are epistemically self-defeating.[12] Likewise, the Moral Disagreement among Ethicists argument is vulnerable to this kind of self-defeat challenge: if premise 1 is a metaethical claim about which there is intractable moral

about whether *potential* moral disagreements also have skeptical consequences (see Tersman (2013)).

[12] To be clear, while some of these authors do not *explicitly* explicate their self-defeat challenges as undermining the premises of moral disagreement arguments, a close reading of these passages shows they are describing this kind of self-defeat challenge. Additionally, Sampson (forthcoming), Shafer-Landau (2003: 220), and Leiter (2014: 146-148) also discuss an indirect version of the self-defeat challenge to moral disagreement arguments where such arguments overgeneralize to a general philosophical skepticism.

disagreement among ethicists, then this premise applies to and epistemically defeats itself.[13]

However, there is a different and often-overlooked version of the epistemic self-defeat challenge to moral disagreement arguments: if there is disagreement about the *conclusion* of a moral disagreement argument, then the skeptical consequences of this argument will apply to this conclusion and epistemically defeat this argument.[14] It appears that the Moral Disagreement among Ethicists argument is also vulnerable to this overlooked kind of epistemic self-defeat: *if* there is also intractable moral disagreement among ethicists about whether the conclusion of this argument (i.e. ethicists don't know that p) is true, it follows from premise 1 that ethicists (including moral disagreement skeptics) also don't know the conclusion of this argument.

At first glance, this distinction between moral disagreement arguments epistemically self-defeating their own premises verses their own conclusion may not seem novel or significant. While I do not claim that this distinction is very novel, I do claim that this later kind of epistemic self-defeat is significant for two reasons. Firstly, it avoids the general and frequently discussed *self-exempting* response to the epistemic self-defeat challenge in the general peer disagreement literature. Following Elga (2010) and

[13] This kind of self-defeat challenge is also discussed in the general peer disagreement literature regarding Conciliationism – the view that in the face of peer disagreement one should suspend belief or reduce confidence in one's contested beliefs. That is, because there is peer disagreement about Conciliationism, Conciliationism itself dictates that one should suspend belief or reduce confidence about it, and so any argument with Conciliationism as a *premise* will likewise be epistemically self-defeating. See Christensen (2009, 2013), Decker (2014), Kornblith (2013), Elga (2010), Pittard (2015), and Weatherson (2013).

[14] To my knowledge, Enoch (2009: 47-48; 2011: 216) is the only commentator to explicitly discuss this kind of self-defeat for moral disagreement arguments. And to my knowledge, Dixon (chapter 3) is the only commentator to explicitly discuss this kind of self-defeat challenge in the general peer disagreement literature.

Pittard (2015), the moral skeptic might argue that all fundamental principles must be self-exempting for them to be consistent (i.e. they don't call for their own rejection), and since premise 1 plausibly embodies a fundamental epistemic principle, it is exempt from applying to and defeating itself.

Many find this kind of self-exempting response unconvincing and *ad-hoc* (e.g. Decker (2014), Christensen (2013), and Sampson (forthcoming)). What is important for my purposes is to notice that this self-exempting response cannot work against the often-overlooked version of the epistemic self-defeat challenge. The moral skeptic cannot plausibly argue that the *conclusion* of the Moral Disagreement among Ethicists argument (i.e. ethicists don't know that p) is self-exempting because it embodies a fundamental principle without thereby *begging the question* against the moral non-skeptic. This conclusion is something the moral skeptic has to argue for. To illustrate, imagine that the Moral Disagreement among Ethicists argument was used to argue for the conclusion that "ethicists don't know that abortion is permissible." This is obviously not the kind of claim that the moral skeptic can say embodies a fundamental principle that is exempt from the skeptic implications of moral disagreements. And to make this exemption would beg the question against those ethicists who believe they know abortion is permissible. So, *even if* the self-exempting response has some merit against the former kind of self-defeat challenge (against the premises of a moral disagreement argument), it has no merit against the often-overlooked kind of self-defeat challenge (against the conclusion of a moral disagreement argument).[15]

[15] This is also an important point for the general peer disagreement literature where the focus of the self-defeat challenge has been on whether Conciliationism itself is self-defeating (see fn. 13). That is, even if Conciliationism is self-exempting and avoids the former kind of self-defeat (a la Elga (2010) and Pittard (2015)), this self-exemption cannot help in avoiding the later kind of self-

Secondly, the latter type of self-defeat challenge is significant because it allows for my reductio argument. To see why this is the case, it will be helpful to more fully explain this self-defeat challenge by providing a formalized representation of it. To this end, if we use parentheses to indicate the conclusion of the Moral Disagreement among Ethicists argument (i.e. ethicists don't know that p), this overlooked version of the self-defeat challenge can be formalized as:

The Self-Defeat Challenge to "Moral Disagreement among Ethicists"

4. There is intractable moral disagreement among ethicists about (ethicists don't know that p). (Empirically true premise)
5. Thus, ethicists don't know that (ethicists don't know that p). (1, 4)

Prima facie, premise 4 is empirically supported by the facts explained in §1 that there are many ethicists who believe some moral proposition p is true despite being aware that other ethicists believe ~p (e.g. see Bourget and Chalmers (2014)) and that ethicists are perennially unable to resolve their moral disagreements about a great many moral issues (e.g. abortion, euthanasia, affirmative action, etc.). And while this premise can be challenged in several ways (see §3), for ease of exposition let's assume for the time being that it is true.[16]

---

defeat. Thanks to an anonymous reviewer and Hilary Kornblith for raising and pressing the importance of clarifying these points.

[16] One might object to premise 4 on the grounds that the disagreement it embodied is not a *moral* disagreement but is rather an *epistemic* disagreement about whether ethicists know a certain moral proposition. In response, recall from §1 that 'moral disagreement' and 'moral proposition' are defined in this chapter to allow such second-order or meta-ethical disagreements to count as moral disagreements. Again, this terminology provides a convenient way to account for the widely accepted claim that disagreements over metaethical and non-moral propositions can also lead to skepticism.

However, following Christensen (2009: 763), Matheson (2015:149), and Barnett (2019) in the general peer disagreement literature, one could argue that *even if* premise 4 is true, this self-defeat challenge is not very worrisome since it does *not* refute the Moral Disagreement among Ethicists argument. At best, this challenge only establishes that, by the moral disagreement skeptic's own reasoning, the conclusion of the Moral Disagreement among Ethicists argument cannot be known. Indeed, the moral disagreement skeptic might welcome this result by accepting that the conclusion of the Moral Disagreement among Ethicists argument is just another moral proposition that they and other ethicists don't know.

I agree that this self-defeat challenge does not *by itself* refute the Moral Disagreement among Ethicists argument. However, I argue that unless this self-defeat challenge can be blocked, the Moral Disagreement among Ethicists argument is refuted by an additional argument. This additional argument assumes (for reductio) that the Moral Disagreement among Ethicists argument is indeed sufficient to establish its conclusion – ethicists don't know that p – is true. As such, this argument allows one, including ethicists, to know its conclusion. But this result contradicts what is established by the overlooked version of the self-defeat challenge – ethicists don't know this conclusion. And from this contradiction it follows that intractable moral disagreement among ethicists is not sufficient to establish moral skepticism (i.e. premise 1 is false). For clarity, this reductio argument can be formalized as:

Reductio of "Moral Disagreement among Ethicists"

6. If an ethicist E (a) learns of the Moral Disagreement among Ethicists argument (premises 1-3), and (b) accepts its conclusion (ethicists don't know that p)

because they believe this argument is sufficient to establish that (ethicists don't know that p), and (c) this argument is in fact sufficient to establish that (ethicists don't know that p) is true, then they *know* that (ethicists don't know that p). (Premise)

7. An ethicist E satisfies (a) and (b), and (c) is true. (Assumption for Reductio)
8. Thus, E *knows* that (ethicists don't know that p). (6, 7)
9. Thus, premise 1 (i.e. intractable moral disagreement among ethicists about p is sufficient to establish that ethicists don't know that p) is false. (5, 8 Explosion)

In short, contra Christensen, Matheson, and Barnett, I claim that this reductio shows that *if* the Moral Disagreement among Ethicists argument is epistemically self-defeating (via premises 4-5) then this argument fails to establish moral skepticism because its first premise is false (via premises 6-9).

§3

To establish this entailment, I will defend premises 6 and 7 (i.e. the only premises that are not logical consequences of previous premises) from possible objections. Additionally, since this entailment depends on the above self-defeat challenge, I will also defend premise 4 of this challenge from possible objections.

Beginning with premise 6, I claim that this premise is not something that anyone, including the moral disagreement skeptic (e.g. a defender of the Moral Disagreement among Ethicists argument), can deny. Premise 6 essentially says that *if* the premises and reasoning of the Moral Disagreement among Ethicists argument are *sufficient to establish* its conclusion, the conclusion is true,[17] and one believes these claims, then *a fortiori* they

[17] I should note that the claim that 'the conclusion is true' is redundant because, as explained in §1, 'sufficient to establish' is used to refer to any kind of reasoning that a moral disagreement

know its conclusion. Denying this premise is untenable because it follows from the following true principle: if an argument A is sufficient to establish its conclusion, its conclusion is true, and one believes these claims, then *a fortiori* they know A's conclusion. This principle can be understood in terms of 'sufficient justification': if some evidence E sufficiently justifies some true proposition q and one believes this, then *a fortiori* they know q. For instance, if I acquire sufficient justification and believe that the Battle of Britain was fought in 1940 (e.g. from reading a reliable encyclopedia), then *a fortiori* I know this proposition. Likewise, *if* moral disagreement among ethicists is sufficient justification for the truth of moral skepticism and I believe this, then *a fortiori* I know moral skepticism is true.[18]

This principle can also be understood in terms of 'epistemic defeat': if a defeater d is sufficient to undermine belief in q, one's belief in q is defeated by d, and one believes these claims, then *a fortiori* one knows that their belief in some proposition q is defeated. For instance, if consulting an unreliable encyclopedia is sufficient to defeat one's belief on when the Battle of Britain occurred, I did this, and I believe these claims, then *a*

---

argument might use to reach its (allegedly) *true* conclusion (e.g. *sound* deduction, cogent induction or inference to the best explanation with *true* conclusions, etc.). However, I will use this redundancy throughout the rest of the chapter for added clarity and emphasis.

[18] The 'a fortiori' qualification is important because it acknowledges that there are cases where this inference does not follow. For instance, suppose that I competently believe that p on the basis of a flawless argument, but I've been credibly (but falsely) told that I've been slipped a reasoning-distorting pill. Many hold that this high-order evidence blocks one's knowledge of p (e.g. Christensen (2010)). I accept that such higher-order evidence would defeat one's knowledge, but this is not a problem for my use of this principle because this is something everyone, including the moral skeptic, should accept. Such high-order defeating evidence affects all arguments, including both my reductio argument and any version of a moral disagreement argument. So, such higher-order defeat cannot be used by the moral skeptic to deny this principle and block my reductio since this would also defeat their own moral disagreement argument. And the same is true of other similar reasoning distorting kinds of higher-order defeaters (e.g. hypoxia cases).

fortiori I know that I don't know when it occurred. Likewise, if moral disagreement on p is sufficient to defeat knowledge of moral proposition p, one's knowledge of p is defeated by such moral disagreement, and I believe these claims, then *a fortiori* I know that no one knows that p. So, as premise 6 says, *if* the Moral Disagreement among Ethicists argument is sufficient to establish its conclusion as true, then provided one believes this, it follows that this argument provides sufficient grounds (via justification or defeat) to know its conclusion.

Denying premise 7 is also untenable because it just assumes (for reductio) that there is at least one ethicist who is convinced by the Moral Disagreement among Ethicists argument and that this argument is sufficient to establish its conclusion. Indeed, moral disagreement skeptics *themselves* satisfy this premise because they are convinced by and espouse this argument.[19] Thus, there is no plausible way for the moral disagreement skeptic to deny premises 6 or 7.

But since my reductio argument depends on the overlooked version of the self-defeat challenge, one could also deny premise 4 to undermine my reductio argument. As mentioned in §2, denying premise 4 on empirical grounds is *prima facie* implausible since such disagreements do occur (again, see Bourget and Chalmers (2014)) and occur often (e.g. on abortion, euthanasia, affirmative action, etc.). Additionally, denying this premise on empirical grounds is precarious at best for the moral disagreement skeptic.

[19] One might object that the Pyrrhonian moral skeptic can deny this premise since she does not assent to or believe any moral propositions, including the claim that moral skepticism is true (Sinnott-Armstrong 2006: 10). While I believe this objection and Pyrrhonianism in general are implausible (see Gowans 2000b: 10), I will set these issues aside and direct my arguments in this chapter only to the Academic moral skeptic who assents to or believes that moral disagreement arguments are sufficient to establish moral skepticism (Sinnott-Armstrong 2006: 11). But, if appealing to Pyrrhonian moral skepticism is the only way for the moral disagreement skeptic to avoid my reductio arguments, then this, I claim, is still a significant result.

Because premise 2 of the Moral Disagreement among Ethicists argument is also an empirical premise, to avoid the possibility of undermining the empirical support for her own argument the moral disagreement skeptic will need some way to *differentiate* the kind of moral disagreement in premise 2 from the kind in premise 4. And it is not obvious how this can be accomplished non-arbitrarily.

However, to this end, another way to deny premise 4 is on normative grounds: while the moral disagreements in premise 4 do occur, they are epistemically inappropriate or irrational. One way to spell this out is to argue that: even if a large number of ethicists continue to believe p after learning of the disagreement among ethicists on p, since the Moral Disagreement among Ethicists argument establishes that ethicists don't know p, such ethicists are not justified in continuing to believe p. And, as such, this continued moral disagreement in premise 4 carries no epistemic weight. But, this normative denial of premise 4 is also implausible. It amounts to claiming that certain moral propositions (i.e. premise 3) are somehow *exempt* from the skeptical consequences of moral disagreement on them (per the self-defeat challenge). But there seems to be no good justification for this exemption because, as mentioned above, moral disagreements between ethicists are presumably not the result of some cognitive shortcoming and to make this exemption would beg the question against the moral non-skeptic.

Indeed, denying premise 4 on either empirical or normative grounds is not a viable way to block my reductio argument because in many cases of moral disagreement, premise 4 is a *logical consequence* of premise 2. In other words, there is no plausible way to differentiate or exempt the moral disagreements in premise 4 from the moral disagreements in premise 2 because in many cases of moral disagreement, if premise 2 is

empirically true (as the moral skeptic accepts), then so is premise 4. This logical consequence occurs when those ethicists who disagree about p also each, respectively, believe they know or don't know that p *before* they recognize their disagreement on p. To illustrate, in the case where p is Utilitarianism (U), *before* they recognize their disagreement on U, Ethicist-1 (E1) believes they know that U and Ethicist-2 (E2) believes they know that ~U (e.g. because E2 is a Kantian). These beliefs entail that: E1 believes they know ~U is false and E2 believes they know that U is false. And if we assume knowledge is *factive* (i.e. you cannot know something that is false), these beliefs entail that: E1 believes no one knows that ~U and E2 believes that no one knows that U. So, *when* these ethicists recognize their disagreement on U, E1 believes E2 doesn't know that ~U and E2 believes that E1 doesn't know that U. Thus, when such ethicists disagree on U, there is one ethicist (E2) who believes the other doesn't know that U and another (E1) who disagrees about this. Thus, when such ethicists disagree on U (premise 2), they also disagree on ethicists not knowing that U (premise 4). Therefore, there is no plausible way for the moral disagreement skeptic to deny premise 4 on either empirical or normative grounds.

This entailment between premise 2 and 4 is additionally significant and novel to the moral disagreement literature because it shows that the overlooked version of the epistemic self-defeat challenge (i.e. when moral disagreement arguments undermine their own conclusion) is *unavoidable* for the Moral Disagreement among Ethicists argument. As long as there is *at least one* case where ethicists believe they know a moral proposition *before* they learn another ethicist intractably disagrees, then premise 2 will entail premise 4 in these cases and the Moral Disagreement among Ethicists argument

will be epistemically self-defeating. And this unavoidable entailment paves the way for my reductio argument to refute the Moral Disagreement among Ethicists argument.

Thus, since the moral disagreement skeptic cannot deny the only three premises (i.e. premises 6, 7, and 4) that the above reductio depends on, this reductio argument refutes the Moral Disagreement among Ethicists argument.

§4

In this section I will demonstrate how the above reductio argument can be generalized to any moral disagreement argument that satisfies the *Moral Disagreement Thesis* (MDT) – there is moral disagreement (of some kind) over moral proposition(s) p; and the *Sufficiency Thesis* (ST) – moral disagreement (of some kind) on p is (in some way) sufficient to establish moral skepticism on p is true. I will first explain this reductio argument and then briefly defend it from possible objections.

To begin, if we assume for the time being that all moral disagreement arguments satisfy both MDT and ST, then all moral disagreement arguments can be formalized as having *something like* the following argumentative structure:

All Moral Disagreement Arguments

I. Moral disagreement of some kind K about p among or between the members of group(s) G is sufficient to establish that no one in G knows that p. (ST)

II. There is moral disagreement of kind K about p in G. (MDT)

III. Thus, no one in G knows that p. (I, II)

Variable K is meant to stand for any kind of moral disagreement (e.g. mere disagreement; wide-ranging disagreement; deep disagreement; radical disagreement; persistent and pervasive disagreement; irresolvable disagreement; undecidable disagreement; intractable

disagreement; peer disagreement; etc.),[20] and group G can be filled in with any moral agents (e.g. different cultures, ethicists, etc.). These variables allow for many moral disagreement arguments to be captured by this general argumentative structure. For example, it is easy to see that both the Moral Disagreement between Cultures and the Moral Disagreement among Ethicists arguments above are instantiations of this structure.[21]

My reductio of this structure, like the previous reductio argument, also builds upon the overlooked version of the epistemic self-defeat challenge. This argument structure is also vulnerable to this challenge because it seems that premise I applies to and epistemically defeats this argument's conclusion. If we use parentheses to indicate the conclusion of this argument (i.e. no one in G knows that p), this self-defeat challenge can be formalized as:

Self-Defeat Challenge to "All Moral Disagreement Arguments"

IV. There is moral disagreement of kind K in G about (no one in G knows that p). (Empirical Premise)

V. Thus, no one in G knows that (no one in G knows that p). (I, IV)

And, again, while this self-defeat challenge does not *by itself* refute All Moral Disagreement Arguments, I argue that unless this self-defeat challenge can be blocked, the Moral Disagreement among Ethicists argument is refuted by an additional reductio argument.

---

[20] These are examples of the different words that commentators have used to describe moral disagreements. While some of these might be equivalent, I want to allow for the possibility that there are many different kinds of moral disagreement that pose a challenge to moral knowledge.

[21] See Enoch (2009, 2011: chap. 8) for examples of other formalized moral disagreement arguments that are captured by this argumentative structure.

Like my previous reductio argument, this version also assumes (for reductio) that All Moral Disagreement Arguments is indeed sufficient to establish that its conclusion is true. As such, this argument allows one to know its conclusion. But this result contradicts what is established by the self-defeat challenge – that no one knows this conclusion. And from this contradiction it follows that moral disagreement, of any kind, is not sufficient to establish moral skepticism (i.e. premise I is false). For clarity, this reductio argument can be formalized as:

Reductio of "All Moral Disagreement Arguments"

VI. If at least one member in G (a) learns of a moral disagreement to moral skepticism argument (premises I-III), and (b) accepts its conclusion (no one in G knows that p) because they believe this argument is sufficient to establish that (no one in G knows that p), and (c) this argument is in fact sufficient to establish that (no one in G knows that p) is true, then they *know* that (no one in G knows that p)). (Premise)

VII. At least one member in G satisfies (a) and (b), and (c) is true. (Assumption for Reductio)

VIII. Thus, at least one member in G *knows* that (no one in G knows that p). (VI, VII)

IX. Thus, ~ST. (V, VIII Explosion)

Furthermore, like my previous reductio argument, I claim that this reductio argument shows that *if* All Moral Disagreement Arguments is epistemically self-defeating (via premises IV-V) then this argument fails because its first premise is false (via premises VI-IX). And like my previous reductio argument, to establish this entailment I will

defend premises VI, VII, and IV (i.e. the only premises that are not logical consequences of previous premises) from possible objections.

Fortunately, my defense of the previous reductio argument applies *mutatis mutandis* to the above reductio argument. To briefly reiterate, like premise 6, premise VI is not something anyone, including the moral disagreement skeptic (e.g. a defender of All Moral Disagreement Arguments), can deny. This premise essentially says that *if* the premises and reasoning of All Moral Disagreement Arguments are *sufficient to establish* its conclusion, its conclusion is true, and one believes these claims, then *a fortiori* they know its conclusion. Again, denying this claim is untenable because it follows from the following true principle: if an argument A is sufficient to establish its conclusion, its conclusion is true, and one believes these claims, then *a fortiori* they know A's conclusion (e.g. the Battle of Britain examples). And like premise 7, premise VII is an assumption (for reductio) that the moral disagreement skeptic also cannot deny because they *themselves* satisfy this premise.[22]

And, like premise 4, premise IV is also not something the moral skeptic can plausibly deny on empirical or normative grounds since there seems to be no good justification to exempt certain moral propositions (i.e. premise III) from the skeptical consequences of moral disagreement on them. Additionally, in many cases of moral disagreement, premise IV is a logical consequence of premise II. As explained in the previous section, this entailment occurs when those who disagree about p also each, respectively, believe they know or don't know that p *before* they recognize their moral disagreement. In brief, if Group-1 (G1) believes they know p and Group-2 (G2) believes

[22] But, again, see fn. 18 and fn. 19.

they know ~p, then when G1 and G2 discover their moral disagreement of kind K on p (premise II) they are also disagreeing on whether each other knows that p (premise IV). This entailment is additionally significant and novel to the moral disagreement literature because it shows that the overlooked version of the epistemic self-defeat challenge is *unavoidable* for All Moral Disagreement Arguments. There only needs to *one* case where the disagreement in premise II entails premise IV, in the way described above, for All Moral Disagreement Arguments to be epistemically self-defeating. And this unavoidable entailment paves the way for my reductio argument to refute All Moral Disagreement Arguments.

Thus, since the moral disagreement skeptic cannot deny the only three premises that the above reductio depends on (i.e. premises VI, VII, and IV), this reductio refutes All Moral Disagreement Arguments.

§5

My overall argument in the last section was a conditional: *if* all moral disagreement arguments satisfy MDT and ST, then the above reductio follows. And it seems obviously true that MDT and ST are necessary conditions for any moral disagreement argument. Furthermore, because it seems obviously true that all moral disagreement arguments satisfy MDT and ST, the burden of proof is on the moral disagreement skeptic to provide a moral disagreement argument that does not satisfy these (or similar) theses. Either way, my reductio argument refutes those moral disagreement arguments that do satisfy MDT and ST which, I claim, is still a significant result.

In conclusion, despite being one of the oldest and most discussed arguments for moral skepticism, “the” argument from moral disagreement to moral skepticism poses no problem for moral knowledge.

BIBLIOGRAPHY

Alston, William, P., (1995), "How to Think about Reliability", *Philosophical Topics,* 23(1): 1-29.

Arsenault, Michael and Irving, Zachary, (2013), "Aha! Trick Questions, Independence, and the Epistemology of Disagreement," *Thought*, 1(3): 185-194.

Barnett, Zack, (2019), "Philosophy Without Belief," *Mind,* 128 (509): 109-138.

Barnett, Nathan and Han, Li, (2016), "Conciliationism and Merely Possible Disagreement," *Synthese,* 193(9): 2973-2985.

Bealer, George, (1992), "The Incoherence of Empiricism," *Proceedings of the Aristotelian Society* (Supplemental Volume) 66: 99–138.

Bealer, George, (1998), "Intuition and the Autonomy of Philosophy," in DePaul, M. and W. Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, Lanham, MD: Rowman and Littlefield.

Beddor, Bob and Pavese, Carlotta, (forthcoming), "Modal Virtue Epistemology", *Philosophy and Phenomenological Research.*

Bennigson, Thomas, (1996), "Irresolvable Disagreement and the Case Against Moral Realism", *The Southern Journal of Philosophy,* 34(4): 411-437.

Bishop, Michael, (2010), "Why the generality problem is everybody's problem", *Philosophical Studies,* 151: 285-298.

Blessenohl, Simon, (2015), "Self-Exempting Conciliationism is Arbitrary," *Kriterion,* 29(3): 1-22.

Bogardus, Tomas, (2009), "A Vindication of the Equal Weight View," *Episteme,* 6 (3): 324–335.

BonJour, Laurence, (1998), *In Defense of Pure Reason,* Cambridge: Cambridge University Press.

Bourget, David, and Chalmers, David, (2014), "What Do Philosophers Believe?" *Philosophical Studies,* 170(3): 465-500.

Bradford, Gwen, (2015), *Achievement,* Oxford University Press.

Brennan, Jason, (2010), "Skepticism about Philosophy," *Ratio,* 23(1): 1-16.

Campbell, Richmond, "Moral Epistemology", *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/moral-epistemology/>.

Cary, Brandon, (2011), "Possible Disagreements and Defeat," *Philosophical Studies,* 155(3): 371-381.

Cary, Brandon, and Matheson, Jonathan, (2013), "How Skeptical is the Equal Weight View," in Diego Machuca (ed.), *Disagreement and Skepticism*, New York: Routledge, pp. 131-149.

Cappelen, Herman, (2012), *Philosophy without Intuitions,* Oxford: Oxford University Press.

Carter, J. Adam, (2016), "Robust Virtue Epistemology as Anti-Luck Epistemology: A New Solution", *Pacific Philosophical Quarterly,* 97(1): 140-155.

Christensen, David, (2007), "Epistemology of Disagreement: The Good News," *Philosophical Review,* 116: 187–218.

Christensen, David, (2009), "Disagreement as Evidence: The Epistemology of Controversy," *Philosophy Compass*, 4(5): 756–767.

Christensen, David, (2010), "Higher-Order Evidence," *Philosophy and Phenomenological Research*, 81 (1): 185–215.

Christensen, David, (2011), "Disagreement, Question-Begging and Epistemic Self-Criticism," *Philosophers Imprint*, 11(6): 1–22.

Christensen, David, (2013), "Epistemic Modesty Defended," in Jennifer Lackey and David Christensen (eds.), *The Epistemology of Disagreement: New Essays*, Oxford: Oxford University Press, pp. 77-97.

Cohen, Stewart, (2013), "A Defense of the (Almost) Equal Weight View," in David Christensen and Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays*, Oxford: Oxford University Press, pp. 98- 117.

Comesana, Juan, (2006), "A Well-Founded Solution to the Generality Problem", *Philosophical Studies,* 129: 27-47.

de Brigard, Felipe, (2010), "If you like it, does it matter if it's real?" *Philosophical Psychology*, 23: 43-57.

Decker, Jason, (2014), "Conciliationism and Self-incrimination," *Erkenntnis,* 79: 1099-1134.

Decker, Jason, and Groll, Daniel, (2013), "On the (In)Significance of Moral Disagreement for Moral Knowledge", in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics, Vol. 8*, Oxford University Press: 140-167.

Dellsen, Finnur, (2017), "Reactionary responses to the Bad Lot Objection", *Studies in History and Philosophy of Science*, 61: 32-40.

Dellsen, Finnur, (forthcoming), "The Heuristic Conception of Inference to the Best Explanation", *Philosophical Studies*.

Deutsch, Max, (2010), "Intuitions, Counter-Examples, and Experimental Philosophy," *Review of Philosophy and Psychology*, 1: 447–60.

Devitt, Michael, (2015), "Relying on Intuitions: Where Cappelen and Deutsch Go Wrong," *Inquiry*, 58: 669-699.

Earlenbaugh, Joshua, and Molyneux, Bernard, (2009), "Intuitions are inclinations to believe", *Philosophical Studies*, 145(1): 89-109.

Elga, Adam, (2007), "Reflection and Disagreement," *Noûs*, 41: 478–502.

Elga, Adam, (2010), "How to Disagree About How to Disagree," in Richard Feldman and Ted Warfield (eds.), *Disagreement*, New York: Oxford University Press, pp.175-186.

Elgin, Catherine, (2010), "Persistent Disagreement," in Richard Feldman and Ted Warfield (eds.), *Disagreement*, New York: Oxford University Press, pp. 51-68.

Enoch, David, (2009), "How is Moral Disagreement a Problem for Realism?", *The Journal of Ethics*, 13: 15-50.

Enoch, David, (2010), "Not Just a Truthometer: Taking Oneself Seriously (but not Too Seriously) in Cases of Peer Disagreement," *Mind*, 119: 953–997.

Enoch, David, (2011), *Taking Morality Seriously: A Defense of Robust Realism*, Oxford: Oxford University Press.

Feldman, Richard, (2005), "Respecting the Evidence," *Philosophical Perspectives*, 19: 95-119.

Feldman, Richard, (2006a), "Epistemological Puzzles about Disagreement," in Steve Hetherington (ed.), *Epistemic Futures*, New York: Oxford University Press, pp. 216–236.

Feldman, Richard, (2006b), “Reasonable Religious Disagreements,” in L. Antony (ed.), *Philosophers without Gods: Meditations on Atheism and the Secular Life*, New York: Oxford University Press, pp. 194-214.

Feldman, Richard, (2009), “Evidentialism, Higher-Order Evidence, and Disagreement,” *Episteme*, 6 (3): 294–312.

Fitz, James, (2018), “Conciliationism and Moral Spinelessness”, *Episteme*, 15(1): 101-118.

Fitz, James, and McPherson, Tristram, (2019), “Moral Steadfastness and Meta-Ethics”, *American Philosophical Quarterly*, 56(1): 43-55.

Frances, Bryan, (2005), “When a Skeptical Hypothesis is Live,” *Noûs*, 39: 559–95.

Frances, Bryan, (2010), “The Reflective Epistemic Renegade,” *Philosophy and Phenomenological Research*, 81(2): 419-463.

Frances, Bryan, (2013), “Philosophical Renegades,” in Jennifer Lackey and David Christensen (eds.), *The Epistemology of Disagreement: New Essays*, Oxford: Oxford University Press, pp. 121–166.

Frances, Bryan, (2014), *Disagreement*, Cambridge, UK: Polity Press.

Frances, Bryan, (2018), “Scepticism and Disagreement,” in Diego Machuca and Baron Reed (eds.), *Skepticism: From Antiquity to the Present*, New York: Bloomsbury, pp. 581-591.

Frances, Bryan and Matheson, Jonathan, "Disagreement", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/disagreement/>.

Frise, Matthew, (2018), “The reliability problem for reliabilism”, *Philosophical Studies*, 175: 923-945.

Fumerton, Richard, (2010), “You Can’t Trust a Philosopher,” in Richard Feldman and Ted Warfield (eds.), *Disagreement*, New York: Oxford University Press, pp. 91-110.

Gardiner, Georgi, (2014), “The Commutativity of Evidence: A Problem for Conciliatory Views of Disagreement,” *Episteme*, 11 (1): 83–95.

Goldberg, Sanford, (2013a), “Disagreement, Defeat, and Assertion,” in Jennifer Lackey and David Christensen (eds.), *The Epistemology of Disagreement: New Essays*, Oxford: Oxford University Press, pp. 167-189.

Goldberg, Sanford, (2013b), "Defending Philosophy in the Face of Systematic Disagreement," in Diego Machuca (ed.), *Disagreement and Skepticism*, New York: Routledge, pp. 279-294.

Goldman, Alvin, (1979), "What Is Justified Belief?", in G.S. Pappas (ed.), *Justification and Knowledge*, Dordrecht: Reidel, pp. 1–25. Reprinted in *Epistemology: An Anthology* (Blackwell, 2000), ed. Ernest Sosa and Jaegwon Kim, pp. 340-353.

Goldman, Alvin, (2012), *Reliabilism in Contemporary Epistemology.* Oxford University Press.

Goldman, Alvin and Beddor, Bob, "Reliabilist Epistemology", *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/reliabilism/>.

Greco, John, (2010), *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, Cambridge: Cambridge University Press.

Gowans, Christopher, (ed.), (2000a), *Moral Disagreements: Classic and Contemporary Readings*, London: Routledge.

Gowans, Christopher, (2000b), "Introduction", in Gowans (ed.) *Moral Disagreements: Classic and Contemporary Readings*, London: Routledge: 1-43.

Hetherington, Stephen, (1998), "Actually Knowing", *The Philosophical Quarterly*, 48(193): 453-469.

Hetherington, Stephen, (1999), "Knowing Failably", *The Journal of Philosophy*, 96(11):565-587.

Hetherington, Stephen, (2016), *Knowledge and the Gettier Problem*, Cambridge University Press.

Hughes, Nick, (2019), "Disagreement, Dogmatism, and the Bounds of Philosophy," review of *Philosophy Within Its Proper Bounds* by Edouard Machery, in *International Journal of Philosophical Studies*, 1-6.

Hume, David, (1999), *An Enquiry Concerning Human Understanding* Ed. by Tom L. Beauchamp, Oxford: Oxford University Press.

Jehle, David and Brandon Fitelson, (2009), "What is the "Equal Weight View"?" *Episteme*, 6: 280–293.

Joyce, Richard, (2018), "Arguments from moral disagreement to moral skepticism" in Diego Machuca (ed.), *Moral Skepticism: New Essays*, Routledge: 141-162.

Kelly, Thomas, (2005), "The Epistemic Significance of Disagreement," in T. Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology*, vol. 1. Oxford: Oxford University Press, pp. 167-196.

Kelly, Thomas, (2010), "Peer Disagreement and Higher Order Evidence," in R. Feldman and T. Warfield (eds.), *Disagreement*, New York: Oxford University Press, pp. 111-174.

Kelly, Thomas, (2013), "Disagreement and the Burdens of Judgment," in David Christensen and Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays*, Oxford: Oxford University Press, pp. 31-53.

Kelp, Christoph, (2013), "Knowledge: The Safe-Apt View", *Australasian Journal of Philosophy*, 91(2): 265-278.

King, Nathan, (2012), "Disagreement: What's the Problem? Or A Good Peer is Hard to Find," *Philosophy and Phenomenological Research*, 85 (2): 249–272.

Kornblith, Hilary, (2008), "Knowledge Needs No Justification," in Q. Smith, ed., *Epistemology: New Essays*, Oxford University Press, pp. 5-23.

Kornblith, Hilary, (2010), "Belief in the Face of Controversy," in Richard Feldman and Ted Warfield (eds.), *Disagreement*, New York: Oxford University Press, pp. 29-52.

Kornblith, Hilary, (2013), "Is Philosophical Knowledge Possible?" in Diego Machuca (ed.) *Disagreement and Skepticism*, New York: Routledge, pp. 260-276.

Kornblith, Hilary, (2017), "Doxastic Justification is Fundamental," *Philosophical Topics*, 45(1): 63-80.

Lackey, Jennifer, (2007), "We don't deserve credit for everything we know", *Synthese*, 158(3): 345-361.

Lackey, Jennifer, (2009), "Knowledge and Credit", *Philosophical Studies*, 142(1): 27-43.

Lackey, Jennifer, (2010a), "What Should We Do When We Disagree?" in Tamar Szabo Gendler and John Hawthorne (eds.), *Oxford Studies in Epistemology*, Oxford: Oxford University Press.

Lackey, Jennifer, (2010b), "A Justificationalist View of Disagreement's Epistemic Significance," in Adrian Haddock, Alan Millar, and Duncan Pritchard (eds.), *Social Epistemology*, Oxford: Oxford University Press, pp. 298-325.

Lackey, Jennifer, (2013), "Disagreement and Belief Dependence: Why Numbers Matter," in David Christensen and Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays,* Oxford: Oxford University Press, pp. 243–68.

Leiter, Brian, (2014), "Moral Skepticism and Moral Disagreement in Nietzsche", in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics,* Vol. 9. Oxford University Press: 126-149.

Leslie, Sarah-Jane and Lerner, Adam, "Generic Generalizations", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/generics/>.

Littlejohn, Clayton, (2013), "Disagreement and Defeat," in Diego Machuca (ed.), *Disagreement and Skepticism,* New York: Routledge, pp. 169–192.

Littlejohn, Clayton, (2014), "A Note Concerning Conciliationism and Self-Defeat: A Reply to Matheson" *Social Epistemology Review and Reply Collective* 3, no. 12: 104-112.

Loeb, Don, (1998), "Moral Realism and the Argument from Disagreement", *Philosophical Studies,* 90(3): 281-303.

Lord, Errol, (2014), "From Independence to Conciliationism: An Obituary," *Australasian Journal of Philosophy,* 92: 365-77.

Machery, Edouard, (2017), *Philosophy Within Its Proper Bounds,* Oxford: Oxford University Press.

Machery, Edouard, (2019), "Response to Akagi, Hughes, and Springle," *International Journal of Philosophical Studies,* 1-15.

Mackie, James, (1977), *Ethics: Inventing Right and Wrong,* Penguin Books.

Matheson, Jonathan, (2009), "Conciliatory Views of Disagreement and Higher-Order Evidence," *Episteme,* 6(3): 269-279.

Matheson, Jonathan, (2015a), "Are Conciliatory Views of Disagreement Self-Defeating?" *Social Epistemology*, 29 (2): 145–159.

Matheson, Jonathan, (2015b), "Epistemic Norms and Self-Defeat: A Reply to Littlejohn." *Social Epistemology Review and Reply Collective* 4, no. 2: 26-32.

Mcauliffe, William, H.B., (2015), "How did Abduction Get Confused with Inference to the Best Explanation?", *Transactions of the Charles S. Peirce Society,* 51(3): 300-319.

McGrath, Sarah, (2008), “Moral Disagreement and Moral Expertise”, in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol. 3. Oxford University Press: 87-108.

Mogensen, Andreas, (2017), “Disagreement in Moral Intuitions as Defeaters”, *The Philosophical Quarterly*, 67(267): 282-302.

Myers-Schulz, Blake, and Schwitzgebel, Eric, (2013), “Knowing That P without Believing That P,” *Nous*, 47(2): 371-384.

Nozick, Robert, (1974), *Anarchy State, and Utopia*, New York: Basic Books.

Olivola, Christopher, et al., (ms), “Reality does not bite: Universal aversion to illusory experiences”.

Palmira, Michele, (2019), “How to Solve the Puzzle of Peer Disagreement”, *American Philosophical Quarterly*, 56(1): 83-95.

Palmira, Michele, and Stroud, Sarah, (2019), “The Nature and Implications of Disagreement”, *American Philosophical Quarterly*, 56(1): ii-iv.

Pasnau, Robert, (2015), “Disagreement and the Value of Self-Trust,” *Philosophical Studies*, 172 (9): 2315–2339.

Pittard, John, (2015), “Resolute Conciliationism,” *Philosophical Quarterly*, 65(260): 442-463.

Pritchard, Duncan, (2005), *Epistemic Luck*, New York: Oxford University Press.

Pritchard, Duncan, (2008), “Greco on Knowledge: Virtues, Contexts, and Achievements”, *Philosophical Quarterly*, 58: 232, 437-447.

Pritchard, Duncan, (2009), “Apt Performance and Epistemic Value”, *Philosophical Studies*, 143: 407–416.

Pritchard, Duncan, (2012), “Anti-Luck Virtue Epistemology”, *The Journal of Philosophy*, 109(3): 247-279.

Pust, Joel, (2000), *Intuitions as Evidence*, New York: Garland/Routledge.

Pust, Joel, (2001), “Against Explanationist Skepticism Regarding Philosophical Intuitions,” *Philosophical Studies*, 106: 227–258.

Pust, Joel, "Intuition", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/intuition/>.

Risberg, Olle, and Tersman, Folke, (2019), “A New Route from Moral Disagreement to Moral Skepticism”, *Journal of the American Philosophical Association,* 5(2): 189-207.

Rose, David, and Schaffer, Jonathan, (2013), “Knowledge entails dispositional belief,” *Philosophical Studies*, 166(1): S19-S50.

Rotondo, Andrew, (2013), “Undermining, Circularity, and Disagreement,” *Synthese*, 190 (3): 563–584.

Rotondo, Andrew, (2015), “Disagreement and Intellectual Skepticism,” *Australasian Journal of Philosophy,* 93 (2):251-271.

Rowland, Richard, (2017a), “The Significance of Significant Fundamental Moral Disagreement”, *Nous*, 51(4): 802-831.

Rowland, Richard, (2017b), “The Epistemology of Moral Disagreement”, *Philosophy Compass*, 12 (2), 1-16

Sampson, Eric, (forthcoming), “The Self-Undermining Arguments from Disagreement”, in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics,* Oxford University Press.

Sartwell, Crispin, (1991), “Knowledge is Merely True Belief”, *American Philosophical Quarterly,* 28(2): 157-165.

Sartwell, Crispin, (1992), “Why Knowledge is Merely True Belief”, *The Journal of Philosophy,* 89(4): 167-180.

Sayre-McCord, Geoff, "Moral Realism", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/moral-realism/>.

Schafer, Karl, (2015), “How Common is Peer Disagreement? On Self-Trust and Rational Symmetry,” *Philosophy and Phenomenological Research*, 91 (1): 25–46.

Schupbach, Jonah, (2014), “Is the Bad Lot Objection Just Misguided?”, *Erkenntnis*, 79:55-64.

Senor, Thomas. D., (1996), “The Prima/Ultima Facie Justification Distinction in Epistemology,” *Philosophy and Phenomenological Research*, 56: 551-566.

Setiya, Kieran, (2012), *Knowing Right from Wrong,* Oxford University Press.

Silva, Paul, (2013), "Epistemically Self-Defeating Arguments and Skepticism about Intuition," *Philosophical Studies*, 164: 579-589.

Sinnott-Armstrong, Walter, (2006), *Moral Skepticisms*, New York: Oxford University Press.

Shafer-Landau, Russ, (2003), *Moral Realism: A Defense*, New York: Oxford University Press.

Shafer-Landau, Russ, (2006), "Ethics as philosophy: A defense of ethical nonnaturalism", in *Metaethics After Moore*, T. Horgan and M. Timmons (eds.), Oxford and New York: Oxford University Press.

Shafer-Landau, Russ, (2012), *The Fundamentals of Ethics*, (second edition) New York: Oxford University Press.

Sosa, Ernest, (2007), *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, v.1. Oxford: Oxford University Press.

Sosa, Ernest, (2015), *Judgment and Agency*, Oxford University Press.

Sripada, Chandra Sekhar, and Stanley, Jason, (2012), "Empirical Tests of Interest-Relative Invariantism", *Episteme*, 9(1): 3-26.

Stanley, Jason, (2011), *Know How*, Oxford: Oxford University Press.

Stanley, Jason and Williamson, Timothy, (2001), "Knowing How", *The Journal of Philosophy*, 98(8): 411–44.

Strohminger, Margot, (2018), review of *Philosophy Within Its Proper Bounds*, by Edouard Machery, in *Norte Dame Philosophical Reviews.*

Tebben, Nicholas, (2019), "Knowledge requires commitment (instead of belief)," *Philosophical Studies*, 176 (2):321-338.

Tersman, Folke, (2006), *Moral Disagreement*, Cambridge: Cambridge University Press.

Tersman, Folke, (2013), "Moral Disagreement: Actual vs. Possible," in *Disagreement and Skepticism*, Machuca, Diego E., (ed.), New York: Routledge, pp. 90-108.

Titelbaum, Michael, (2015), "Rationality's Fixed Point (Or: In Defense of Right Reason)," in T. Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology*, vol. 5, Oxford: Oxford University Press, pp. 253–294.

Tolhurst, William, (1987), "The Argument from Moral Disagreement", *Ethics*, Vol. 97(3): 610-621.

Tolly, Jeffrey, (forthcoming), "Does reliabilism have a temporality problem", *Philosophical Studies.*

Turri, John, (2010), "On the Relationship between Propositional and Doxastic Justification," *Philosophy and Phenomenological Research,* 80(2): 312-326.

Turri, John, (2015), "Unreliable Knowledge", *Philosophy and Phenomenological Research,* 90: 529-545.

Turri, John, (2015a), "From virtue epistemology to abilism: theoretical and empirical evaluations", in C.B. Miller, M.R. Furr, A. Knobel, and W. Fleeson (Eds.), *Character: new directions from philosophy, psychology, and theology.* Oxford University Press. pp. 315-331.

Turri, John, (2015b), "Epistemic situationism and cognitive ability", in M. Alfano and A. Fairweather (Eds.), *Epistemic situationism.* Oxford: Oxford University Press. pp. 158-167.

Turri, John, (2016). "A new paradigm for epistemology: from reliabilism to abilism," *Ergo*, 3(8), 189-231.

Turri, John, (2016a), "Knowledge as Achievement, More or Less", In M. A. Fernandez (Ed.), *Performance Epistemology: Foundations and Applications*, Oxford: Oxford University Press. pp. 124-136.

Turri, John, (forthcoming), "Virtue epistemology and abilism on knowledge", in H. Battaly (Ed.), *Routledge Handbook or Virtue Epistemology*. Routledge.

van Fraassen, Bas, (1989), *Laws and Symmetry*, Oxford: Clarendon Press.

Vavova, Katia, (2014a), "Moral Disagreement and Moral Skepticism," *Philosophical Perspectives*, 28 (1): 302–333.

Vavova, Katia, (2014b), "Confidence, Evidence, and Disagreement," *Erkenntnis*, 79(1): 173-183.

Weatherson, Brian, (2013), "Disagreements, Philosophical and Otherwise," in Jennifer Lackey and David Christensen (eds.), *The Epistemology of Disagreement: New Essays*, Oxford University Press, pp. 54-73.

Wedgwood, Ralph, (2010), "The Moral Evil Demons," in R. Feldman and T. Warfield (eds.), *Disagreement*, Oxford: Oxford University Press, 216-246.

Wedgwood, Ralph, (2019), "Moral Disagreement and Inexcusable Irrationality", *American Philosophical Quarterly,* 56(1): 97-108.

Williamson, Timothy, (2000), *Knowledge and its Limits*, Oxford: Oxford University Press.

Wilson, Alastair, (2010), "Disagreement, Equal Weight and Commutativity," *Philosophical Studies*, 149 (3): 321 – 326.

Zagzebski, Linda, (2009), *On Epistemology*, Belmont, Calif.: Wadsworth.