RELIABILISM AND THE GENERALITY PROBLEM

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

By

Jeffrey Tolly

Ted A. Warfield, Director

Graduate Program in Philosophy
Notre Dame, Indiana
July 2016
RELIABILISM AND THE GENERALITY PROBLEM

Abstract

by

Jeffrey Tolly

In the post-Gettier era, process reliabilism is one of the most influential epistemic theories on offer. According to process reliabilism, a belief counts as knowledge only if that belief was arrived at by use of a reliable belief-forming process. This reliable process condition is the central constituent in both reliabilist accounts of warrant and justification. Over the course of the past three decades, the generality problem has been seen as one of the most pressing objections to process reliabilism. Roughly, the generality problem begins with the observation that we currently understand very little about the relevant belief forming process types that are measured for reliability for any given belief forming process token. According to the generality problem objection, there is some normative burden for the reliabilist to supply some account of type relevance. This dissertation is an extended investigation into the generality problem—what it is, and what the reliabilist can say in response.

For all of the attention given to the generality problem, there’s been no attempt to systematically formulate how the objection to reliabilism is supposed to go. In chapter one, I present what I take to be the four most reasonable approaches to formulating the generality problem objection, and find all of them subject to compelling objections. In chapter two, I argue that if the reliabilist has some special burden to produce a theory of type relevance, then
so do theorists of almost every other competitor epistemic theory. Hence, the generality problem fails to constitute some *unique* objection to reliabilism.

In chapter three, I criticize recent attempts to ground type relevance in subjective factors like a subject’s higher order beliefs, or a subject’s practical interests. In chapter four, I argue that relevance theories that merely take into account causal features of the tokens aren’t extensionally correct. Rather, extensionally correct relevance theories must make relevant types a function of the token’s modal properties. Finally, in chapter five, I present relevance principles for arithmetical intuition belief formation while highlighting two neglected aspects for crafting a correct relevance theory.
For Tim and Lori, my parents

If love is what you’re looking for

Then all roads lead to an open door

And you’ll find your way

You’ll find your way back home.

—Andrew Peterson
CONTENTS

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

CHAPTER 1: WHAT IS THE GENERALITY PROBLEM? ............................................. 1
  1.1 Introduction ................................................................................................................... 1
  1.2 Type Relevance—Misunderstood .............................................................................. 4
  1.3 The Theory Incompleteness Approach ................................................................. 8
  1.4 The Non-existence of Relevant Types Approach .................................................. 12
  1.5 The No-conceptual Grasp Approach ...................................................................... 19
  1.6 The Theoretical Virtue Approach .......................................................................... 28
  1.7 Conclusion .................................................................................................................. 33
    1.7.1 The Future Prospects for the GPO ............................................................... 33
    1.7.2 The GD Dilemma and the Current State of the GPO ............................ 35

CHAPTER 2: A DEFENSE OF PARRYING RESPONSES TO THE GENERALITY
PROBLEM .......................................................................................................................... 38
  2.1 Introduction ................................................................................................................ 38
  2.2 Reflective Justification and Bishop’s Parrying Response ................................... 43
    2.2.1 Bishop’s Parrying Response ......................................................................... 43
    2.2.2 Conee’s Rejoinder ......................................................................................... 47
    2.2.3 Bishop’s Parrying Response Defended ....................................................... 48
  2.3 Comesaña’s Response to the Generality Problem .................................................. 52
    2.3.1 Comesaña’s Head-On Response, and a Simple Parrying Response ....... 52
    2.3.2 Refuting Comesaña’s Head-on and Simple Parrying Responses .......... 54
    2.3.3 Evidence Possession and a Parrying Response ........................................ 57
    2.3.4 Competent Basing and a Warrant Parrying Response .......................... 64
  2.4 Conclusion .................................................................................................................. 68

CHAPTER 3: AGAINST SUBJECT-BASED THEORIES OF RELEVANCE ...... 69
  3.1 Introduction ................................................................................................................ 69
  3.2 Belief-relevance Theories ....................................................................................... 73
  3.3 Against Belief-relevance Theories ....................................................................... 78
  3.4 Practical Interest-relevance Theories ..................................................................... 85
  3.5 Against Practical Interest-relevance Theories ...................................................... 89
CHAPTER 4: CAUSAL RELEVANCE PRINCIPLES WORTH REJECTING  
4.1 Introduction ................................................................. 97
4.2 Causal Relevance Principles and CN Features ...................... 101
   4.2.1 Causal Relevance Principles Motivated and Characterized .... 101
   4.2.2 Contingent Causal Non-mental Features (CNs) .................. 106
   4.2.3 Two Sorts of Causal Relevance Theories ......................... 107
4.3 Against CN Relevance Principles ..................................... 112
   4.3.1 Step One ............................................................... 113
   4.3.2 Step Two ............................................................. 120
4.4 Conclusion ................................................................. 122

CHAPTER 5: ARITHMETICAL INTUITION AND TWO NEGLECTED PARTS OF A RELEVANCE THEORY  
5.1 Introduction ................................................................. 124
5.2 Type Relevance and Relevance Principles ............................ 128
   5.2.1 The Metaphysics of Type Relevance .............................. 128
   5.2.2 Multiple Relevant Types, or Only One? ......................... 131
5.3 The Structure of a Relevance Theory .................................. 132
5.4 Arithmetical Intuition ..................................................... 139
   5.4.1 Negative Reliability and Multiple Relevant Types .............. 139
   5.4.2 Positive Reliability and a Key Difference from Perception .... 143
5.5 Conclusion ................................................................. 152

BIBLIOGRAPHY ................................................................. 154
ACKNOWLEDGMENTS

Courage, dear heart.

— C.S. Lewis, The Voyage of the Dawn Treader

Although the next journey is about to begin, I must say something about the one coming to a close. Pursuing a graduate degree in philosophy has been without a doubt the most difficult thing I’ve done in my life. But amidst the difficulty and frustrating days, there has been much joy and excitement. While I don’t know what the future holds, I do know that these five years at Notre Dame were right for me. In addition to all of the skills and concepts I’ve learned, I can honestly say that my love of philosophy has grown. There are few things I would trade for these years of learning philosophy from such amazing mentors and colleagues.

First, I must thank my advisor, Ted Warfield, who has supervised not only this dissertation but my oral exam as well. The quality of this dissertation is a large product of his helpful and incisive comments on numerous drafts of all these chapters—from start to finish. In addition, I am grateful for the support and feedback from my committee members, Mike DePaul, Blake Roeber, and Tom Senor. These next professors graciously mentored me in my time here at various points: Jeff Speaks, Sean Kelsey, Meghan Sullivan, and Sam Newlands. Without them, I wouldn’t have had the skills nor confidence to complete this project.

Much of the joy in graduate school has been shared joy, learning together with my colleagues and classmates in the department—whether it was in the lecture hall, the lounge, the library, or the pub. I single out for special mention Alex Arnold, Brad Rettler, Daniel
Immerman, Justin Christy, Ben Rossi, Peter Finocchiaro, Will Smith, Sarah Schramm, Amy Seymour, Meg Schmitt, Nevin Climenhaga, Paul Blaschko, Liz Jackson, Ting Lau, Rachel Jonker, Andrew Helms, Kate Finley, Mike Rauschenbach, and Father Raphael.

I can’t say I’ve been particularly proficient at keeping a clear headspace for the past five years. Often times, I “got lost in philosophy land” and needed to be pulled back down to reality. My community from outside the philosophy department kept me grounded, somewhat sane, and laughing. At times they carried me, when the mental, physical, and spiritual burdens felt too much to bear. I’m thankful for the friendship and care of the Luke Potter, Del Fehsenfeld, Kevin Estep, Brian Hedges, the Seachrises, the Smiths (Brette and Adam), the Smiths (Billy and Kate), the Roebers, the Newlands, the Messers, the Colemans, the Lodges, the Hendricksons, Ben Wetzel, Jonathan Riddle, Karie Cross, Lucy Dunderdale, Emily Dubie, Clair Mesick, Andrew Carreon, Alex Vawter, Laura McMillion, Jake Long, Nate Rupprecht, and Brocks Burnick and Hedlund. I could not have done this without you all.

Undoubtedly, the sweetest surprise of graduate school was meeting my beautiful wife and best friend, Caroline May. Interestingly enough, she and I started dating the same week that I started writing this dissertation. She has been there every step of the way with unending support, encouragement, and jokes (so I don’t take myself too seriously!) I am also grateful for her help in editing large portions of this manuscript. Due to her steady and peaceful presence, writing this dissertation never felt like a labor of toil. Instead, it was a labor of love, from first to last. I am thrilled to be taking my future journeys with her.

Finally, I dedicate this dissertation to my parents, Tim and Lori. From a young age, they never squelched my questions or curiosity. By their discipline, grace, and example, they gave me everything I needed to grow physically, spiritually, and intellectually. Much of who I am today is due to their sacrifices and love for me. This is for them.
 CHAPTER 1

WHAT IS THE GENERALITY PROBLEM?

1.1 Introduction

The generality problem is commonly thought of as one of the most pressing issues for process reliabilism.¹ Process reliabilism about justification and process reliabilism about warrant share this common feature: they both claim that the agent’s use of a reliable belief-forming process to arrive at her belief is the key ingredient that grounds the belief’s possession of the relevant epistemic property (whether warrant or justification).² In this chapter, I will focus on process reliabilism about justification (henceforth PRJ), but surely many of the upshots of this chapter could apply to process reliabilism about warrant.³

The generality problem for PRJ begins with a puzzle. Processes can be thought of as general repeatable types or as precise one-off tokens. Plausibly, only the former can be

¹ See Goldman and Beddor (2015). In their overview of the work done on reliabilist epistemology, they include the generality problem as one of the top six “problems,” or, “objections” to reliabilism.

² I use the term “key ingredient” here so as to include reliabilist theories which hold that process reliability is a crucial part of justification (or warrant), but that there is some interesting no-defeater condition that must also be satisfied in order for an agent to have justification (or warrant). The relevant kind of defeater I have in mind is not one that can be analyzed in terms of unreliability.

³ For an early formulation of PRJ, see. Goldman (1979:13-14) for an account prima facie justification and (pg. 20) for the additional no defeater condition (for ultima facie justification). Both of these analyses involve reliable processes as the central feature that grounds an agent’s possession of justification. In addition, Goldman (1986:44-5) presents a reliabilist theory of warrant, on which warrant requires that an agent’s belief-forming process has both “local” and “global” reliability. Also, I use the term “warrant” in the functional sense much like Plantinga does, where warrant is the state the possession of which grounds one’s having knowledge that p so long as she also possesses belief that p and p is true. (Plantinga (1993: v) In other words, functionally, warrant is the state that “takes” true belief to knowledge.
evaluated for reliability or unreliability. As both defenders and critics of PRJ have noticed, a given process token exemplifies many types. A given case of seeing a red ball at close distance, and then coming to believe that a red ball is right there, exemplifies the types [belief formation], [visual belief formation], [vision in good light at a close distance with a medium sized object], etc. Indeed, it exemplifies an infinite number of types. Reasonably, not all of these types are evaluated for reliability in determining whether the target belief has justification. But now we can ask, which of these types is relevant for determining reliability and justification? In other words, which set of details from the token is “held fixed” in the relevant type description? Call this question the generality problem question. It more or less constitutes the starting point for the generality problem objection.

According to the literature, the generality problem question turns into the generality problem objection (henceforth GPO) to PRJ upon realizing that humans—including the best philosophers—do (and have done) a poor job at answering the generality problem question.

---

4 Richard Feldman (1985) and Goldman (1979) are key figures who highlighted this important distinction for making sense of reliabilism. Most philosophers agree that types, rather than tokens, are the entities that can be measured for reliability. Although recently, Juan Comesaña (2006) articulates a way in which tokens could be evaluated for reliability with respect to some space of possible worlds as a reference class. Even if this notion of measuring a token for reliability makes sense, Comesaña, correctly, recognizes that framing reliabilism like this doesn’t get the reliabilist out of the generality problem. The reliabilist still would have to provide an account of which possible worlds were contained within the reference class used to evaluate the token’s reliability. See §5 for a discussion regarding the close connection between the reference class for measuring reliability and the relevant type.

5 For example, presumably, [belief-formation] isn’t the type that gets evaluated for reliability. Consider someone who, sadly, is a brain in a vat, such that her perceptual beliefs are all systematically false. Such a person, presumably, can still have reliable and justified intuition beliefs in a priori claims. But, if the relevant type for these intuition belief tokens was just [belief-formation], then the reliability measurement would be significantly decreased from all of the false perceptual beliefs formed by the agent (since [perceptual belief-formation] is contained within the broader category of [belief-formation]).

6 The specific language of relevant and irrelevant types was introduced by Feldman. The following quote is illustrative:

   Obviously, then, one of these types must be the one whose reliability is relevant to the assessment of my belief…Let us say, then, that for each belief-forming process token there is some “relevant” type such that it is the reliability of that type which determines the justifiability of the belief produced by that token…In order to evaluate [process reliabilism], we need some account of what the relevant types of belief-forming processes are. (Feldman 1985:160, emphasis mine)
It is the nature of our lack of understanding into the notion of type relevance that constitutes the key premise in the GPO. Earl Conee and Richard Feldman are the arch defenders of the GPO, and they conclude that “process reliability theories of justification and knowledge look hopeless.”

I think it is high time to take a step back and re-examine how this GPO argument is supposed to work. In particular, one might wonder how it is that our lack of understanding and competence with the notion of type relevance generates a defeater for PRJ. Despite the popularity of the generality problem, there has been no substantive attempt to rigorously formulate the GPO or explain how the unanswered generality problem question should lead us to conclude that PRJ is hopeless. Investigating how such an objection might be formulated is one endeavor of this chapter.

In §1.2, I present Conee and Feldman’s rendering of our current lack of understanding into the nature of type relevance, and formulate the sorts of premises that might be used in the GPO. Over the course of the next four sections, I discuss what I take to be the four most plausible candidate approaches—one for each section—to formulating the GPO argument, giving each approach a critical appraisal. Lastly, in §1.7, I close with a discussion of what we learn from these four attempts to formulate the GPO, and how the structure of these formulations suggests two main ways the reliabilist could respond to GPO. Ultimately, I conclude that none of these four approaches delivers a clear and compelling defeater for reliabilism. This is because each of these four reasonable approaches to formulating the GPO falls prey to a rather problematic dilemma. I show that the reasonable formulations for each

---

7 Conee and Feldman (1998: 24). It turns out that others are inclined to agree with them. For example, see Plantinga (1993: 28–29), Matheson (2015).

8 Each of these candidate GPO formulations has as its conclusion either that reliabilism is false, or that we’re unjustified in believing reliabilism.
approach either end up invoking principles that clearly prove far too much, or end up invoking claims about PRJ that we have little or no reason to believe. I take this result to show that there is currently no compelling generality problem objection to PRJ.

1.2 Type Relevance—Misunderstood

Those formulating the GPO invoke, as premises in the GPO argument, particular details about our lack of understanding of type relevance. Importantly, there are two ways to think about this lack of understanding: descriptively, and explanatorily. Descriptively, defenders of the GPO are keen to invoke the following thesis:

The degree to which humans can describe the relevant types for belief-forming process tokens, in general, is rather low.

No GPO defender would want to commit to the thesis that we literally know nothing about relevant type descriptions. Presumably, for tokens of visual belief formation, they’d grant that humans know (at least implicitly) that the use of vision is a detail from the token that is relevant, i.e., a detail from the token that is held fixed in the relevant type description. But it is also clear that we can’t (at least for most cases) give anything close to a complete description of the relevant type. According to some formulations of the GPO, this relatively low degree of descriptive knowledge generates a problem for PRJ. But how low is “low”? It is not obvious how to quantify this value. The good news is that this un-clarity doesn’t halt our examination of the candidate GPO argument formulations that invoke our low degree of descriptive knowledge. In the formulations I’ll examine below, the other premises in the arguments shed enough light on what “low” must amount to in order for the argument to have any plausibility.

---

9 For example, consider the fake barn county case (Goldman: 1976). For that case of barn-identifying perceptual belief-formation, we can ask the following question: which details regarding the ratio of barns to barn-facades around the agent at the time of belief-formation are relevant (i.e., held fixed in the type description)? The ratio of barns to facades within 500 meters of the agent? Within 1000 meters? Within 5 miles? It is not immediately clear what to say here.
Conee and Feldman are more keen to point out the problems associated with our lack of explanatory understanding of type relevance. In other words, they think there is some problem associated with our lacking a theory of type relevance. Such a theory would posit principles (call them relevance principles) that determine which features of tokens get held fixed in their relevant type descriptions. According to Conee and Feldman, our lack of progress in finding a theory of type relevance is particularly problematic. To begin, Feldman writes,

Of course, the arguments of this paper do not show that no acceptable version of the reliability theory can be constructed. However, it is fair to say that The Problem of Generality is a serious problem for the theory… To make the reliability theory plausible, then, some other way must be found to specify processes, some way that assures that only reliable processes operate in cases in which one's evidence does support a belief adequately and only unreliable processes operate when one's evidence fails to support a belief. While it may be possible to come up with a general account of processes that satisfies this requirement, I believe that the prospects for doing so are not good. (Feldman 1985: 172, emphasis mine)

By a “general account” of type relevance, Conee and Feldman have in mind a relevance theory that will always “specify the relevant type wherever there are definite facts about justification.” According to Conee and Feldman, all of the extant relevance theory proposals fail, either by failing to specify a particular relevant type for each token, or by specifying particular types that are intuitively non-relevant, or, incorrect for measuring reliability. This leads them to a sort of pessimism about the project of finding a relevance theory.

13 Also, for a criticism of Becker (2008), see Brueckner and Buford (2013). For a criticism of Beebe (2004), see Dutant and Olsson (2013).
At the beginning of their overview and criticism of many relevance theories currently on offer, they write,

Our thesis is that the prospects for a solution to the generality problem for process reliabilism are worse than bleak. We will investigate the merits of approaches exemplified by several recent proposals. There is no significant progress in any of these approaches, singly or in combination. The basic process reliabilist idea just does not pan out (Conee and Feldman 1998: 5, emphasis mine)

In addition, they claim,

That exhausts the reasonable philosophical approaches to the generality problem. If they all fail, then so does process reliabilism. (6, emphasis mine)

They conclude with the following sobering remark:

That is the full variety of existing approaches to disposing of the generality problem. In the absence of a brand new idea about relevant types, the problem looks insoluble. Consequently, process reliability theories of justification and knowledge look hopeless. (24, emphasis mine)

In essence, Conee and Feldman think the attempts to answer the generality problem are utter failures, and (crucially) that there is no hope in sight. These comments communicate that things are looking quite grim for PRJ—i.e., that the generality problem poses something of a devastating blow to PRJ.

From these ideas, we can discern at least three different kinds of theses, about our explanatory lack of understanding that one might use as premises in a GPO argument. First, GPO defenders could invoke the fact that our progress on a relevance theory, after all the effort that philosophers have put in, has fallen short of a complete relevance theory. Here, a complete relevance theory posits necessary and jointly sufficient conditions on type relevance that can determine the relevant type for all possible tokens. While we clearly lack such a complete relevance theory, GPO arguments might only need a more scaled-back thesis regarding the theoretical progress on type relevance. Such a thesis could simply state that, up
through this current point in time, the progress made has been sufficiently low. Here, let theoretical progress on a relevance theory refer to the discovery of either a correct necessary condition or a correct sufficient condition on type relevance. Much like the “low” degree of descriptive understanding discussed above, the surrounding premises of the GPO argument would suffice to specify the correct sense of “low” regarding the progress on finding relevance principles. Thirdly, one might invoke a certain sort of hopelessness, or, unlikelihood of making more significant progress on a relevance theory, much like the last conclusion drawn by Conee and Feldman in the quotes above.

With these theses—regarding our lack of type-relevance understanding—on the table, we’re now in a position to present and evaluate some reasonable approaches to how these theses might, in conjunction with other premises, generate a defeater for PRJ. By my lights, there are four different reasonable approaches one might take to formulating the GPO. I address all four of these options in turn, and find them all to be rather problematic. Ultimately, each of these approaches falls prey to the following dilemma.

The Generality Problem Dilemma (GD)

An approach to formulating the GPO succumbs to the GD if and only if the reasonable argument formulations of that approach either

a. invoke principles that prove too much. These principles generate defeaters for all kinds of plausible theses that, intuitively, don’t seem threatened at all by epistemic principles like these. Call this the overgeneralization horn of GD.

or,

b. invoke premises about PRJ that we have little or no reason to believe. Call this the no-motivation horn of GD.

Arguments that prove far too much or contain unmotivated premises are clearly unsuccessful. If all the reasonable approaches to formulating the GPO fall into GD, then we have no reason to think the GPO successfully generates a defeater for PRJ.
1.3 The Theory Incompleteness Approach

The first GPO approach I examine is what I call the \textit{theory incompleteness approach}. Interpreting the GPO like this appears to most closely capture what Conee and Feldman have in mind given what they have put in print. Specifically, Conee and Feldman claim that “Without a specification of the relevant type, process reliabilism is radically \textsl{incomplete}. \textit{Only when a bearer of reliability has been identified does the theory have any implications about the justification of beliefs in particular cases.}”\textsuperscript{14} They continue, “[Reliabilism] must first be elaborated at least enough to imply \textsl{exactly} what process type has to be reliable in the case in question.”\textsuperscript{15} So, perhaps Conee and Feldman think that PRJ, as a theory—in its current state of development—has \textit{no} implications (i.e., makes no verdicts) about whether particular cases of belief formation exemplify justification.\textsuperscript{16}

Clearly, it would be particularly bad for PRJ if it failed to entail justification verdicts for even the most obvious intuitive cases of justified belief formation. Another way to phrase this objection is in terms of the theoretical virtue of \textit{explanatory power}. A theory of some property F has high explanatory power only to the extent that it correctly determines whether F is exemplified in particular cases. Whether we’re justified in accepting some theory significantly depends on its degree of explanatory power. One might read Conee and Feldman as claiming that, without a supplementary complete theory of relevance, PRJ has no explanatory power because it offers \textit{no verdicts} on whether particular cases exemplify justification.

We can formulate this argument against PRJ as follows.

---

\textsuperscript{14} Conee and Feldman (1998: 3, emphasis mine).

\textsuperscript{15} Ibid., 3 emphasis mine.

\textsuperscript{16} Also see Matheson (2015: 459): “Without a solution to the generality problem, reliabilism fails to give concrete verdicts about cases.”
A1 PRJ doesn't have a complete relevance theory.

A2 PRJ needs to identify the bearer of reliability (the relevant type) in order for PRJ to yield any justification verdicts about particular cases.

A3 PRJ can identify bearers of reliability only if there is a complete relevance theory in place that specifies the specific relevant type for each possible token case.

A4 PRJ cannot identify the bearers of reliability for particular cases (A3, A1).

A5 PRJ currently fails to deliver any verdicts about which cases have or fail to have justification (A4, A2).

A6 Therefore, PRJ fails as a theory of justification due to inadequate explanatory power.

While A1-A6 might have a ring of plausibility, it exemplifies the overgeneralization horn of the GD dilemma. To begin, in the above quotation, Conee and Feldman seem to be concerned with the features of the theory, PRJ, as opposed to something amiss with agents who subscribe to and use PRJ. This distinction is of critical importance for understanding A1-A6. There are two possible readings of what it might mean for PRJ, the theory, to “identify the bearer of reliability” in A2 and A3.

(I) PRJ’s identifying the bearer of reliability amounts to PRJ giving a definite description that, when applied to each particular belief-forming case, has as its referent the specific entity that gets evaluated for reliability (which is the relevant type).

(II) PRJ’s identifying the bearer of reliability amounts to PRJ giving some theory or set of principles that determine, for each particular belief-forming case, what the specific entity is that gets evaluated for reliability (which is the relevant type).

On both of these readings, we can see A1-A6 as invoking a principle that generates defeat far too easily and pervasively. On reading (I), A2 is clearly true. PRJ claims that justification is a function of a token’s relevant type being sufficiently reliable. That description, “the token’s
relevant type,” does pick out, when applied to each particular token, the specific entity being evaluated for reliability—namely, that token’s relevant type.

But on reading (I), A3 is rather implausible. PRJ is a theory of *justification*. It is *not* a theory of *type relevance*. PRJ, as a theory of justification, is *complete* with respect to the property of justification, because it offers necessary *and* sufficient conditions on justification. Here’s a rough characterization of PRJ: b is prima facie justified iff the process used to produce b is sufficiently reliable, and b is ultima facie justified iff it is prima facie justified and S lacks a defeater for b.17 Concerning prima facie justification, as we’ve seen, PRJ offers the further analysis: the process used to produce b is sufficiently reliable iff the relevant type—of the process token that produced b—is sufficiently reliable. For each possible belief-forming case, PRJ gives a verdict on the identity of the entity that gets measured for reliability: it is the token’s *relevant process type*. So long as there is *in fact* a referent for this definite description for each token case that determinately confers (or fails to confer) justification, then PRJ—the theory—will make/entail justification verdicts for these tokens.

It is true that PRJ invokes the concept *relevant type*, at least indirectly, in its account of justification. But why think that PRJ needs to posit an additional supplementary theory (or analysis) of type relevance, along with its analysis of justification, in order to entail any justification verdicts on particular cases? It seems as if A3 depends on (in addition to the (I) reading) the following false principle:

---

17 I take this analysis here to constitute the core commitment of any specific process reliabilist theory of justification. Different precise reliabilist theories of justification might disagree on what the no-defeater condition consists in, or on what process reliability amounts to (on type relevance, or on the threshold for justification). Also, technically speaking, PRJ as formulated here is incomplete, because it is only framed in terms of justification at the moment of belief-formation. It says nothing about what justifies our belief as we hold it in storage after the moment of belief-formation. In other words, it doesn’t say anything about justified belief *maintenance*. But the need to supply such an account isn’t unique to PRJ. Any epistemic theory will need to account for how beliefs kept in storage *keep* their justified status, not just how beliefs acquire justification at the moment of *formation*. 
Higher Order Explanation (HE)

Necessarily, for any philosophical theory T, property F, and property G, if philosophical theory T of F—in which T invokes G as playing a central role in determining whether something is F—is to entail any verdicts of whether F is exemplified in particular cases, then T has to be supplemented with a theory of G.

We can see rather clearly that HE is too strong. Consider the following analogy. Suppose one subscribes to a rough consequentialist theory of morally permissible action: action A is morally permissible iff performing A maximizes the good in the world. In order for this theory to generate verdicts of moral permissibility, does it need to be supplemented with a theory of the good? That seems wrongheaded. The theory identifies the entity (the good) that gets summed up in the determination of whether A is permissible. Intuitively, that is good enough for a theory of permissible action.

If HE were true, it seems that almost every philosophical theory we have makes no verdicts at all. Here’s why: almost every philosophical theory invokes concepts that are non-primitive (i.e., they admit of further analysis). I also contend that virtually no philosophical theories offer supplemental accounts of these non-primitive notions that eventually bottom out in analyses that only make reference to primitive concepts. But if this is the case, then given HE, the vast majority of philosophical theories make no verdicts on particular cases. But this is clearly the wrong result. We should reject HE. If HE is false, then why require PRJ to possess a supplementary account of type relevance? Now, maybe PRJ is a special case, and there is some unique reason why PRJ needs a supplemental account of type relevance to entail justification verdicts. But Conee and Feldman haven’t given us some reason to think it is unique in this way.

On the (II) reading of “identifying the bearer of reliability,” identifying such a bearer requires that PRJ provide a theory that gives some analysis or general account of type relevance. On this reading, A3 is true. Unfortunately, on the (II) reading, A2 is unreasonable,
given that on this reading, A2 seems to rely on HE. Most reasonably, PRJ only needs to “identify” relevant types in sense (I) in order to deliver justification verdicts. We can see that A1-A6 exemplifies the overgeneralization horn because it seems to rely on principle HE. But HE would analogously generate a defeater for so many other reasonable philosophical positions in a rather implausible fashion.

Conee and Feldman might be intending to invoke an analogous principle to HE that is more limited—only applying to a particular sub-class of theories O—and then claim that PRJ is in O. But such an argument has yet to be supplied, and we’ve yet to be given any account of such a revised HE principle. Hence, this “repaired” version of A1-A6 currently stands rather undeveloped and unmotivated—thus exemplifying the no motivation horn.

1.4 The Non-existence of Relevant Types Approach

The second approach is what I call the non-existence of relevant types approach. One might think that our current lousy state of progress on a relevance theory reveals something ontologically suspect about PRJ. Perhaps our failures to clearly grasp the notion of type relevance is an indication that relevant belief-forming process types don’t actually exist. Perhaps all of our theoretical failures on this front show us that reliabilists are just conceptually confused, and that there is nothing that exemplifies the concept relevant process type for justification. Indeed, many of those who think justification has nothing to do with reliability would be happy with this result!

Here’s a first pass formulation of how this GPO argument might go:

B1 After all of our efforts to find the complete set of relevance principles, we’ve made very little theoretical progress.

B2 If B1, then it is probably the case that there is no complete set of relevance principles.

B3 If there are relevant types, then there is a complete set of relevance principles.
B4  There probably are no relevant types (B1-B3).

B5  If PRJ is true, there are relevant types.

B6  PRJ is probably false. (B4-B5)

The relevant sense of probability here is *epistemic* probability. Clearly, the crucial premise of this argument is B2.\(^\text{18}\) Importantly, B2 expresses an inference pattern that satisfies the description of a no-see-um inference. No-see-um inferences progress to conclusions that assert the probable non-existence of \(x\) from premises of the form \(S\) cannot see, find, or discover \(x\). As many have pointed out, no-see-um inferences transmit or confer justification only in certain circumstances.\(^\text{19}\) For instance, a no-see-um inference from the premise *I cannot find a polar bear in my 2-man tent after 1 minute of looking* to the conclusion *there probably isn’t a polar bear in the tent* is a justification-conferring inference. But if one replaced the term “polar bear” with “gnat” in this inference scheme, then the resultant inference wouldn’t confer justification on the conclusion.

For our purposes, it is important to note that B2 relies on a suppressed premise.

B2*  The circumstances that allow no-see-um inferences to confer justification obtain in the case of B2.

Why should we believe B2*? There seem to be two ways one might argue for B2*. First, one can do so by appeal to a general principle that PRJ satisfies, given its current state of theoretical development. Secondly, one could argue that features *unique* to PRJ make it the case that B2* is true. I’ll argue that, on the general principle approach to supporting B2*, B1-

\(^{18}\) One might also question premise B3. For instance, one might think that process type relevance is *primitive*, and doesn’t admit of any further analysis of explanation. See Kappel (2006) for a defense of this view. Ultimately I agree with most who have written on the generality problem that type relevance is not primitive and can in fact be explained or analyzed in terms of other conditions or principles.

\(^{19}\) Perhaps the best discussions of no-see-um inferences comes from the evidential problem of evil literature in the philosophy of religion. For specific discussions of no-see-um inferences, see Wykstra (1984), Bergmann (2002), Dougherty (2014).
B6 exemplifies the overgeneralization horn. On the latter approach, B1-B6 exemplifies the no motivation horn.

First, consider the following general principle:

**NG**  
For every possible concept and theory, if after much effort and time, we still haven’t made much significant progress towards a complete theory of some concept, then it is epistemically probable that there are no informative principles or jointly necessary and sufficient conditions for that concept.

One might think that the current state of theoretical progress on PRJ satisfies NG, thus entailing B2*. But what exactly does NG mean? First, let “we” refer to the philosophical community, seeing as how it is the philosophical community that commits itself to reflecting on and finding analyses of more or less abstract notions like type relevance. Also, let “much effort and time” refer to a quantity of time and effort that is greater than or equal to the amount of time and effort that the philosophical community has put in to trying to develop a relevance theory for PRJ. I think that rather straightforwardly, we can see a principle like NG to overgeneralize and generate defeaters for claims that, intuitively, don’t seem to be threatened at all by claims like NG. Hence we should reject NG. I show this, by analogy, below.

To begin, consider the concept of *table*. Presumably, we see tables all the time. The existence of tables, to many, will probably seem as obvious as any Moorean claim one can think of. But, believe it or not, we’ve yet to come up with a good informative analysis of *being a table*. We might know some necessary conditions for *being a table*, e.g., that tables possess an approximately flat surface. But the necessary conditions we do know don’t seem anywhere close to jointly sufficient. While this fact is indeed surprising, given everything else

---

20 Here, for compositional nihilists (who don’t believe that there are any composite material objects) or compositional organicists (who believe the only composite material objects are living organisms), we can get the same point across by instead thinking of instances of the property *being arranged table-wise*, which these two camps will often grant is instantiated by sets/classes of simples.
philosophers have settled, it is clear that this fact doesn’t seem to generate anything like a defeater for our belief that tables exist. But if NG is true, a parallel argument to B1-B6 could generate a defeater for the claim tables exist.\textsuperscript{21} But that is absurd. Already, it seems we should be skeptical of NG.

A defender of GPO might respond that we can repair NG by claiming that it only applies to philosophical concepts. While this is a rather vague notion, a grasp of some paradigm examples will suffice for understanding the subsequent criticisms of this modified version of NG. Examples of philosophical concepts include freedom, existence, causation, knowledge, rationality, etc.

There are a handful of examples of concepts that show NG to overgeneralize to generate a defeater for a great many philosophical theories. First, consider the notion of something as commonplace as truth itself. For quite some time, especially throughout the twentieth century, philosophers have attempted to offer a theory of truth. Looking at the philosophical literature on truth, there doesn’t seem to be a clear frontrunner theory of truth out there. Broadly speaking, the “deflationist camp” has levied powerful criticisms against the “correspondence/truth-maker camp” and vice versa.\textsuperscript{22} It appears as if the concept truth has seen a similar sort of theoretical stagnation as that seen by type relevance. If this is the case, then NG would allow us to generate a parallel argument to B1-B6 to justify the claim that there is no truth. But this can’t be right. In addition, this result would generate a defeater for any theory that invokes truth. Given that most theories of knowledge have a truth condition, this

\textsuperscript{21} Presumably, there is an informative set of necessary and sufficient conditions for being a table, even though we haven’t discovered them.

\textsuperscript{22} See Ramsey (1927) and Strawson (1949) for deflationary theories of truth. See Armstrong (1997) for a realist truth-maker correspondence theory of truth.
result would undermine all of those. But this seems like a patently bad way to object to all of these theories of knowledge.

Secondly, consider the notion of moral responsibility. Presumably, actions for which people are morally responsible happen quite often. Much philosophical writing and thinking has gone in to providing a theory of moral responsibility. Some view moral responsibility as requiring libertarian free will. But these views have the challenge of explaining what free will is. But understanding free will has proven quite difficult as well. A complete account of free will needs to say something in response to the famed mind argument. The mind argument claims that any action for which there is a legitimate alternative possibility open to the agent at the moment of decision is more like a twitch, rather than a willing or an action. It is doubtful that a satisfying response has been given. On the other hand, there are the compatibilists who deny that moral responsibility requires free will. These philosophers have the burden of explaining how morally responsible action can occur even if it is the case that the agent has no say in whether a particular outcome/future occurs rather than some other outcome. All sorts of criticisms have been made of the various moves each side makes, and we’re left without anything close to a complete theory of moral responsibility. But, if NG is true, then one could generate a parallel argument to B1-B6 and derive the (probable) non-existence of morally responsible action. This would also generate a defeater for any philosophical theory that invokes moral responsibility. But, intuitively, this is an absurd way to argue against the existence of moral responsibility and theories that invoke this concept.

---

23 For example, see Van Inwagen (1983).

24 See Fischer (2006) for a good example of a defense and presentation of a compatibilist view of moral responsibility. See Van Inwagen (2000) for a discussion of both of these sorts of arguments against the compatibilist and libertarian positions.
To make matters worse for the GPO, type relevance is much better off, vis-a-vis NG, than truth or moral responsibility. Philosophers have only been thinking about type relevance for PRJ for the past three decades. In that time span, only about thirty papers or chapters that touch on this issue (and even fewer that focus on it) have been published in books or scholarly journals. Much more work has been done on truth and moral responsibility.25

Hence, it looks like a general principle like NG is false due to the fact that it overgeneralizes and proves too much. That way of arguing for B2* leads to the overgeneralization horn. But, setting more general principles to the side, perhaps paying attention to the unique details of how previous attempts to build a relevance theory have failed reveals some reason to adopt B2*.26 Something like this might be what Conee and Feldman have in mind. They write,

It is reasonable to look for a solution to the generality problem in three places: common sense, science, and context… We shall argue that none of these approaches works out… That exhausts the reasonable philosophical approaches to the generality problem. If they all fail, then so does process reliabilism (1998:5-6).

---

25 I don’t take my comments here to disconfirm or count against the following more modest general principle:

\[ \text{NG’} \quad \text{For every possible concept and theory of that concept, the fact that much time and effort hasn’t yielded much significant positive progress towards a complete theory of some concept increases the probability that there is no complete set of grounding principles for that concept.} \]

Such lack of progress might not increase the relevant probability much, but, reasonably, it might just a little. But the GPO defender needs a stronger principle than this to generate a defeater for PRJ. It has to be the case that a lousy track record of theory development makes it improbable that a correct theory exists to be found. It is this stronger claim that we’ve seen to be rather unmotivated.

26 This revised sort of GPO argument might proceed in a similar manner to Peter van Inwagen’s argument against the existence of non-organic composite material objects (1990: 56-71). Van Inwagen presents arbitrariness arguments to show this. His methodology roughly involves positing one candidate answer to the special composition question (Just when do material simples compose some composite object?) that would allow for non-organic composites, showing how it fails, and then doing the same thing for a number of other slightly altered answers to the special composition question. The rhetorical, and perhaps insightful, effect of this line of reasoning has created the sense in many of his readers that there probably is no good (principled, non-ad hoc) answer to the special composition question that allows for non-organic composites. The exercise more or less enables the reader to see that there is no such correct answer to the special composition question. Perhaps Conee and Feldman think that, rationally, we ought to have the same reaction once we consider all of the failed attempts to answer the generality problem, and in particular the way in which they’ve failed.
Here’s one plausible reading of this text: there are three key conceptual areas in which it is reasonable to think that we’d “find” or “develop” a correct relevance theory if there is such a correct theory to be found. According to Conee and Feldman, the proposals from those three areas are failures as general and complete relevance theories, and this gives us reason to think such a theory won’t be found. More formally,

i. It’s epistemically probable that, if there are correct necessary and sufficient principles of relevance, the correct principles would come from one of the three following conceptual domains: science, common sense, and context-based linguistics.

ii. The proposals given in the last thirty years in these three areas fail so drastically that it is probable (epistemically) that further attempts to refine versions of these proposals will all fail.

iii. The proposals given in each of these three areas are representative of the kinds of reasonable proposals one could possibly give in these three areas.

iv. Therefore, there probably isn’t a complete set of necessary and sufficient principles of relevance to be found.

Then from iv., along with a premise like B3, one can infer that there probably aren’t relevant types. It also seems as if premises i.-iii. could also support the more modest claim, made by Conee and Feldman, that finding a correct theory of relevance is hopeless (improbable).27

However, each of i.-iii. can reasonably be questioned and rejected. First, Conee and Feldman say nothing in support of i. Upon reflection, there doesn’t seem to be any good reason to accept i. Other than scientific, common-sense, and contextual-linguistic factors, why not also consider pragmatic factors, modal factors, or purely mental (rather than causal/scientific) factors as conceptual domains from which to find relevance principles?28

27 See my discussion at the end of §2.

28 Indeed, in chapter 5, my relevance principles for basic belief-formation are both mental and modal in nature, rather than causal/scientific in nature.
Also, I think we should be skeptical about ii. and iii. given the relatively small amount of literature that is been devoted to answering the generality problem question.

To review, one might reasonably abandon attempts to support B2* with some general principle like NG. But if one in turn attempts to support B2* by reference to unique features of the reliabilist’s track record of failure in finding a relevance theory, we’re left without any clear idea of what these features might be. Hence, we can see that the non-existence of relevant types approach falls in to the no-motivation horn as well.

1.5  The No-conceptual Grasp Approach

One might grant that our impoverished knowledge of type descriptions and minimal theoretical progress on type relevance fail to give us good reason to think that relevant types don’t exist. But, perhaps these observations do give us reason to think that there is something wrong with our apprehension of the concept reliable belief formation, and hence a problem with our grasp of PRJ. Here, I’m using conceptual grasp in a very minimal sense: the sort of conceptual grasp that it takes to competently use and deploy a concept in everyday thought and speech. According to this new GPO approach, perhaps we don’t even grasp the concepts reliable belief-forming process and unreliable belief-forming process. If this were the case, PRJ would be in trouble for two reasons. First, directly, it is reasonable that one cannot justifiably accept some theory of F, which explicitly invokes concept G in its analysis, if one doesn’t grasp G. Seeing as how PRJ explicitly invokes reliable belief-forming processes in its analysis of justification, if we fail to grasp the concept reliable process, then we wouldn’t be justified in accepting PRJ.29 Secondly,

29 Here, I’m putting aside cases in which one might have good, independent testimonial evidence in favor of some theory T while not being able to fully understand T. Consider you hear someone you trust say “The genius physicist at Oxford has come to believe the theory he’s been thinking about for a while.” Suppose you have no idea what this theory even is, but on the basis of hearing this, you come to believe, of the theory the physicist was considering, that it is true. In this situation, one, presumably, can have a justified de re belief (of T) that it (T) is true, without much insight into what T says. But this is not the sort of weak acquaintance epistemologists take
and indirectly, if one cannot even grasp the concept reliable belief-forming process, then one cannot apply this concept to particular cases of belief formation to test PRJ for explanatory power. This would close off one important means by which one could gain justification for believing PRJ. Call this approach to formulating the GPO the no-conceptual grasp approach.

Here’s a first pass formulation of this GPO argument.

C1 After all the work put in to considering the notion of type relevance for justification, we still have a low degree of knowledge regarding the complete descriptions of relevant types, and we’ve made only a low degree of progress on a relevance theory.

C2 If C1, then we don’t grasp the concept reliable belief formation.

C3 PRJ invokes the concept reliable belief formation in its analysis of justified belief.

C4 Necessarily, for all theories and concepts, if theory T of F explicitly invokes concept G as playing a central role, then one can justifiedly accept T only if she grasps G.

C5 Therefore, we’re not justified in accepting PRJ.

C2 is the crucial and controversial premise of this argument. Why believe C2? First, we might think that C2 is supported by a general principle about the necessary conditions for grasping reliability concepts. Examples of reliability concepts include reliable thermometer, reliable testifier about US history, reliable coffee maker, etc. Reliable belief-forming process (for justification) is just one kind of reliability concept.

Consider the following principle one might invoke in support of C2:

Reliability Grasp 1(RG1)

Necessarily, for all reliability concepts Y and subjects S, S grasps concept Y only if

a. S knows or has a disposition to know (upon reflection) a sufficiently high percentage of the principles of relevance for Y.

themselves to have with the competing theories of epistemic justification they’re examining (like reliabilism, coherentism, evidentialism, direct acquaintance, phenomenal conservatism, etc.)
b. For each of a sufficiently broad class M of possible tokens t, S knows or has a disposition to know (upon reflection of t and Y) a sufficiently high percentage of the description of t’s relevant type for concept Y.30

Let the “sufficiently high percentage” clauses in both RG1a and RG1b refer to percentage values that are at least higher than the percentage of relevance principles and relevant type descriptions—for the concept reliable belief-formation—known by philosophers working on the generality problem. This being the case, RG1 rather straightforwardly supports C2. But should we accept RG1? I think not. Consider the following case of reliability concept grasping:

CURRY

Stephen Curry is a professional basketball player who is excellent at free throw shooting (among other things). For the season, he’s shooting 90% from the line. Craig and Liz are watching Curry’s team, the Golden State Warriors, on television at a sports bar. Curry is about to step up to the free throw line. Here, Craig says, “Curry is an elite free throw shooter.” Liz agrees, adding, “Yes, absolutely. His free throw shot is so reliable these days.” Neither Liz nor Craig are professional philosophers. They’ve never heard of the type/token distinction. They are, however, competent adult basketball fans.

When Liz and Craig say things like “his free throw shot is so reliable,” intuitively, they say something meaningful and true. Presumably, in this context free throw shot refers to a process type. Given Liz and Craig’s competent understanding of NBA basketball, it is highly plausible that they grasp the concept reliable free throw shot in the context of NBA basketball standards.

30 Also, it is important to note that the tokens t (in M) that agents represent and reflect on have complete descriptions. This is important for understanding the psychology of how competent adult humans consider process tokens. For example, a human can perceptually be acquainted with and represent some specific token process x of a coffee maker brewing coffee. But it is clear that one’s perceptual apparatus doesn’t acquaint one (perceptually) with the complete description of token x. For instance, we don’t see details of x like it is not the case that invisible fairies are pushing the hot water through the coffee beans. But, plausibly, what we do when we consider whether x exemplifies the reliability concept reliable coffee maker is first represent x as exemplifying a complete token description t where we fill in (either correctly or incorrectly) the rest of the background information on the token x that we weren’t directly perceptually acquainted with. It is these complete token descriptions t that I’m invoking in RG1b.
The intuitive upshot of CURRY constitutes good reason to reject RG1. To begin, Liz and Craig rather clearly fail to satisfy RG1a. This is because on any straightforward reading of the case, they have no beliefs at all about the principles of relevance for the concept reliable free throw shot. First, out of the entire human population, it is only a handful of philosophers who actually grasp the concept of the process type/token distinction. Presumably, Liz and Craig fail to grasp the concept of this distinction, as well as the concept relevant type for free throw shooting. They’ve probably never heard or thought about these concepts. But grasping the type/token distinction and grasping concepts like relevant type for free throw shooting are necessary for having beliefs (either dispositional or occurrent) regarding the principles of type relevance for reliable free throw shooting. But clearly, people other than philosophers can grasp reliability concepts. The same sorts of considerations also undermine RG1b. One cannot have (occurrently or dispositionally) beliefs regarding the descriptions of process types if one doesn’t even grasp the type/token distinction. It is clear that any argument that relied on RG1 would prove far too much, showing that most humans in general fail to grasp almost every reliability concept we reasonably take ourselves to grasp. Hence, any argument that depends on RG1 falls into the over-generalization horn.

At this point, there is a reasonable straightforward way for the defender of this GPO approach to re-group and revise her reasoning. The GPO defender can grant that RG1 fails on the grounds that it makes the overly-demanding requirement that the agent possess knowledge of type-relevance truths in order to grasp reliability concepts. Instead, the GPO defender will claim that grasping a reliability concept Y only requires one to have a tacit sensitivity to type-relevance truths for Y. This tacit sensitivity to type-relevance truths for Y amounts to having stable dispositions to believe particular claims that are crucially related to relevant types for token instances of Y. Then, the GPO defender will claim, in a revised version of C1, that
our degree of understanding regarding type relevance is so low that we fail to satisfy even these tacit sensitivity dispositional conditions.

To further understand this revised version of C1-C5, it is instructive to investigate the nature of the tacit sensitivity required for grasping a reliability concept. To begin, I think that those who grasp reliability concepts understand that whether some particular case exemplifies a reliability concept depends on a certain sort of ratio: the ratio of a certain outcome $x$ occurring to outcome $y$ occurring in a particular relevant set of cases. This “set of cases” used for the reliability ratio measurement is what philosophers and statisticians call the reference class for a reliability measurement. Applied to the concept of reliable belief-formation, outcome $x$ is true belief-formation, and outcome $y$ is false belief-formation. The reference class is comprised of particular belief-forming events that all at least share some set of features in common with the token case being evaluated.

There is an important relationship between the reference class of a token, and a token’s relevant type: All the particular events that constitute the reference class satisfy the token’s relevant type description. It doesn’t make sense for a token’s degree of reliability to depend or partially depend on a truth/falsity ratio measured across a class of cases that don’t even satisfy the relevant type description. So, it follows that any principles that place constraints on the relevant type thereby place constraints on the reference class for the reliability measurement.

Importantly, providing the description of what all the reference class cases have in common suffices for answering the main kind of question that motivates the generality problem to begin with: Which subset of the token’s features are relevant for evaluating the token for
reliability and justification?\textsuperscript{31} Hence, one’s awareness of features held fixed across the reference class cases constitutes an awareness of the features of the relevant type description, even if this awareness is implicit in nature.\textsuperscript{32}

Reflection on everyday cases like CURRY reveals that grasping a reliability concept entails having a disposition to represent certain key features of the reference class (and hence, the relevant type as well). For instance, if someone were to, oddly, push Liz on what she means by “Curry’s free throw shot,” she’d be able to hum a few bars about how she means “when Curry walks up to the line and takes a free throw shot, in a normal game situation, with crowd noise, when he is uninjured, when he is not sick or nauseous, when he takes a deep breath and does his normal 2 dribble pre-shot routine, etc.” But since most adults (like Liz) lack a grasp of the concept reference class and relevant type, it is reasonable that Liz is only implicitly representing these features as features of the relevant type. Importantly, it is implausible that Liz’s grasp of the concept reliable free throw shot is constituted by a disposition to name (upon reflection) most of the features constitutive of relevant types for tokens she considers. That is too demanding. Moreover, it seems like Liz only needs a disposition to represent (implicitly) features of the

\textsuperscript{31} Given the characterization of the generality problem according to Feldman (1985) and Conee and Feldman (1998), the problem for the reliabilist is that, although we can know the features of a given token, we don’t know how to perform the reliability measurement for that token to determine whether that token confers justification on the produced belief. For the token case of seeing and coming to believe in the presence of a red ball (discussed in the introduction), we don’t know whether to perform the reliability measurement with respect to belief-formation in general, or with respect to perceptual belief-formation, or with respect to visual-belief formation, etc. But the more precise way to formulate this puzzle is in terms of the reference class corresponding to the token: we don’t know whether the token’s reference class is comprised of particular cases of all sorts of belief formation, or just with particular cases of perceptual belief-formation, or just with cases of visual belief-formation, etc. If we could completely determine the features of a token’s reference class, then we could effectively answer the questions Conee and Feldman raise to get the generality problem going.

\textsuperscript{32} Wallis (1994:251-262), argues that the relevant type and the reference class are distinct, and hence there are actually two questions raised by the generality problem: First, what is the relevant type for a token? Secondly, what is the reference class for measuring reliability for a token? He claims that making progress on the generality problem requires keeping these two questions (and their answers) distinct in our theorizing. Here, however, he doesn’t specifically argue why, methodologically, it is important for addressing these two questions separately. As I’ve articulated above (fn. 30), if one offers an account of the reference class, one thereby also succeeds in offering an account of the relevant type—given that all of the reference class cases share the relevant type description in common.
type that are particularly important for grounding the reliability measurement for that type’s token. For instance, Liz understands that Curry’s being healthy (rather than being ill) plays a significant role in explaining why his shot is currently very reliable rather than unreliable. It is only these important features of types that agents must be disposed to implicitly represent as being constitutive of the relevant type.

Secondly, the sensitivity necessary for grasping a reliability concept Y also requires an agent to be a competent judge of whether particular token cases exemplify Y. We wouldn’t say that Craig or Liz grasp the concept of reliable free throw shot if they regularly ascribed this property to free throw shots by players that they knew to be the worst shooters in the NBA. So, I think we have reason to accept the following set of necessary conditions for grasping a reliability concept:

Reliability Grasp 2 (RG2)

Necessarily, for all S and for all reliability concepts Y, S grasps Y only if, for each of a sufficiently broad class of possible tokens t,

a. S is disposed such that were S to consider whether t exemplifies Y, S would most likely judge correctly whether t exemplifies Y.

b. S is disposed such that were S to implicitly consider the relevant type of t with respect to Y, S would most likely correctly and implicitly ascribe relevance to many features of t that are important for grounding/explaining the specific reliability measurement for that t’s type.33

RG2 accommodates our intuitions about reliability concept possession in everyday life much better than RG1.

As I mentioned above, the revised C1-C5 argument will then use modified versions of C1 and C2 to derive the conclusion that we're unjustified in accepting PRJ:

C1’ Our current degree of understanding regarding type relevance for reliable belief-formation is so low that we fail to satisfy the dispositional conditions of RG2.

---

33 See fn. 29. Like RG1, the tokens t in RG2 are complete token descriptions that an agent can reflect upon. Also, the “broad class” of possible tokens for which an agent must exemplify the tacit competences described in a. and b. is probably different for each reliability concept.
C2’ If C1’, then we fail to grasp the concept *reliable belief-formation*. If C1` is the correct description of our lack of understanding regarding relevant types, then reasonably, we wouldn’t grasp of the concept *reliable belief-forming process* after all.

However, we’ve been given little to no reason to believe that C1` is true. C1` is an empirical claim about the abilities and dispositions of most human beings. Defenders of the GPO haven’t specifically offered any arguments in favor of a thesis like C1`. To go one step further, it is important to point out that one reasonable argument that one might *try* to make in favor of C1` seems to be undermined by recent empirical findings.

Consider the following empirical observation that one *could possibly* make:

P1 Across the human population, either there is no substantial agreement on the important features of belief-forming process type descriptions for a wide variety of particular cases, or there is no substantial agreement on how reliable a given process is for a wide variety of particular cases.

P1 would offer strong support to C1`. Given this evidence of widespread systematic disagreement about type relevance and reliability, it would be quite unreasonable to hold that most humans have the competences and dispositions described in RG2.34

Unfortunately for this approach, there is good empirical evidence that makes P1 significantly less probable. Martin Jonsson has recently conducted a study with the following set-up. He took a sample of the human population and tested their intuitions on type relevance and reliability measurements.35 In the survey given to the sample subjects, each subject was presented with a series of cases of belief formation that they watched on a video

---

34 I think that divergence in the sample’s reliability ratings for each given case would reveal a tacit disagreement in the sample group about which process type they thought was relevant. This is because intuitions regarding reliability levels often times depend on how we conceive of the relevant type.

35 Jonsson (2013)
screen or read about, and then they were asked various questions about the belief-forming process tokens that they just witnessed. They were asked to first describe the sort of belief-forming process used in each case. The subjects were then asked how reliable the various belief-forming processes were at generating true beliefs. In addition, a different group of test subjects were asked about whether the characters acquired justification in the various cases of belief formation.\footnote{For the first group of subjects, I have in mind what Jonsson calls the “reliability condition” group of experiment 3. For the second group of subjects, I have in mind the “justification condition” group of experiment 3 (2013: 255).}

What Jonsson found is that there is significant “convergence” (i.e., agreement) on the descriptions people gave for particular token cases of belief formation, and significant convergence on which of the cases exemplified high or low reliability.\footnote{Ibid., 256-257.} In addition, the other group of subjects for the most part identified the same group of cases, that were selected as reliable instances of belief formation by the first group, as cases of justified belief formation. These results suggest that, by and large, the subjects were all at least tacitly sensitive to the same features of the token when they were considering the sort of reliability that is constitutive of justification. This seems to be good evidence against P1. I refer the reader to his article to examine the finer details of his experimental methodology. Granted, one could say that his sample was too small to be sufficiently representative, or one ought to hold out to see if the results can be replicated before she makes a judgment on the study. Fair enough. But at the very least, one should at least acknowledge that P1 faces an important empirical challenge. In sum, C1’ remains unmotivated.
1.6 The Theoretical Virtue Approach

The theoretical virtues are good-making properties of a theory that many consider to be indications of a theory’s truth. We’ve already discussed one of them in section 3: explanatory power. But philosophers and other theorists acknowledge at least one other theoretical virtue: elegance (i.e., simplicity). It is a matter of some debate as to what theoretical elegance amounts to. Nevertheless, a rough grasp of this concept is good enough for our purposes. Consider two theories of some property F. It is, intuitively, a point in theory 1’s favor if both theories explain just as many particular cases of F but theory 1 does all of this while invoking fewer concepts and entities than theory 2 invokes. Here, theory 1 has an advantage over theory 2 with respect to elegance.38 In addition, it’s reasonable that we’re more justified in accepting theory 1 than we are in accepting theory 2 if we realize that theory 1 has more overall theoretical virtue than theory 2 has. So, the overall degree of theoretical virtue possessed by a theory plays a determining role in fixing the degree of justification an agent might have for accepting that theory.

As I discussed in §1.3, our lack of progress on a relevance theory doesn’t indicate that PRJ has a problem with low explanatory power. It also doesn’t appear as if our relative lack of understanding of type relevance reveals any shortfall in elegance possessed by PRJ. But, the absence of a good relevance theory does show PRJ to be lacking in a different theoretical virtue: what I call explanatory depth. Consider a theory T of property F. The degree of T’s

---

38 What might it mean for the theoretical virtues to indicate the truth of a theory, to the extent that some theory has these virtues? Here, I think one plausible view on how they indicate truth is by showing a theory to have a high epistemic probability to the extent that it has a high net degree of theoretical virtue. See Swinburne (2001) for a defense of this view. It is also common to think that one is justified in believing claims that are epistemically very probable for them, and unjustified in believing claims that are epistemically improbable for them. In what follows, what I say doesn’t necessarily depend on this account of the relationship between the theoretical virtues, probability, and justification. All I need is that theoretical virtue determines, in some way, one’s epistemic justification for adopting a theory.
explanatory depth is the degree to which it offers supplemental explanations or analyses of the non-primitive fact-types that the theory invokes to play a determining role in whether F obtains. According to PRJ, whether a given type for a token is the relevant type is one of the facts that helps determine whether S gains a justified belief from that process token. Given that type relevance isn’t primitive, PRJ could increase its explanatory depth by having a supplemental theory of type relevance. One key reason that explanatory depth is a theoretical virtue is because high explanatory depth allows agents to more clearly understand and apply the theory.

Given that there is currently no good theory of type relevance, it follows that there is some lack or lacuna in explanatory depth for PRJ. At this point, it is not immediately clear how this insight might generate a defeater for PRJ. The following principle clearly would generate such a defeater:

Explanatory Depth 1 (ED1)
For any theory T, if T has at least as low explanatory depth as PRJ currently has in virtue of its lack of progress on a relevance theory, then the overall degree of theoretical virtue for T is too low for one to be justified in accepting T.

But, ED1 is clearly too strong, and would send any GPO argument formulation that invokes it into the over-generalization horn. This holds for two reasons. First, analogically, any philosophical theory that invokes truth or moral responsibility would, seemingly, be unjustified given ED1 (for reasons discussed on §4). But that doesn’t seem right. Secondly, ED1 doesn’t accommodate intuitions on how the virtue of explanatory depth interacts with the other theoretical virtues in grounding one’s justification for accepting a theory. Suppose that some theory T of property F explains all of the intuitive cases of F correctly in a very elegant way, and suppose we see this to be the case because we grasp and understand T. Our degree of justification for accepting T doesn’t appear to be drastically reduced upon realizing
that we lack an analysis of \( G \)—where \( G \) is some notion invoked by \( T \) for determining whether \( F \) obtains. While a failure of explanatory power—e.g., giving the intuitively incorrect verdict on a particular case—can significantly undermine our justification for accepting that theory, a lacuna of explanatory depth doesn’t carry that same undermining force. Hence, it doesn’t seem as if considering the current state of PRJ alone—regarding its lack of explanatory depth—will reveal some clear defeater for PRJ.

On the other hand, evaluations of explanatory depth are particularly probative in the context of comparing/contrasting the overall theoretical virtue of two or more competing theories. To begin, consider two competing theories of property \( F \): theory \( A \) and theory \( B \). Suppose, for the sake of argument, that the two theories are equivalent with respect to explanatory power and elegance, but \( A \) is superior with respect to explanatory depth—and suppose the subject \( S \) realizes this to be the case. In such a situation, all else being equal, I contend that \( S \) is unjustified in adopting theory \( B \). The main idea is quite simple: a lacuna in explanatory depth drives down a theory’s overall degree of theoretical virtue to at least some degree, and if theory 1 has a greater degree of theoretical virtue than theory 2, one is not justified in accepting theory 2.

An improved version of the explanatory virtue approach is as follows:

\[
\text{D1 } \text{Given that we haven’t found anything close to a complete set of relevance principles for PRJ, there is a significant lacuna of explanatory depth for PRJ.}
\]

\[
\text{D2 } \text{Because of this lacuna of explanatory depth, PRJ’s overall degree of explanatory virtue—with respect to explanatory power, depth, elegance—is less than the overall degree of explanatory virtue had by at least one other competitor theory of justification, and we can see this to be the case.}
\]

\[39\] By “all else being equal,” I intend to rule out cases in which a subject has some “extra-theoretical” reason to accept a given theory, like testimony for a super-reliable agent that a theory is true.
D3 For all theories and rational agents, if theory A has more explanatory virtue than theory B and S sees this to be the case, then S is unjustified in adopting theory B.

D4 Therefore, we’re not justified in holding PRJ.

Unlike ED1, D3 is a much more plausible principle. D1-D4, in my mind, constitutes the strongest version of GPO discussed in this chapter. Unfortunately, it still falls into the no-motivation horn. At this point, we’re left wondering why we should believe something as strong and detailed as D2. Given the content of D2, giving a proper appraisal of this premise would require a thorough examination of all the extant competitor theories of justification. To make matters worse, the informativeness of such an investigation would only go as far as our ability to determine, compare, and contrast the level of theoretical virtue possessed by various candidate theories. Do we currently possess rigorous ways to measure values like degree of explanatory power, degree of elegance, and degree of explanatory depth? It seems not. Even if we possessed such metrics, such an investigation would only prove helpful if we also knew the degrees to which each different theoretical virtue impacts the degree of overall theoretical virtue. As I gestured at above, it appears to be the case that an absence of explanatory power hurts the overall theoretical virtue of a theory more than an absence of explanatory depth. But making a judgment like D2, rationally, would require more precision than this general sort of observation.

But even if we possessed all of these metrics, it is still an open question at this point whether there exists a competitor theory that outclasses PRJ with respect to overall theoretical virtue. Take a comparison with Conee and Feldman’s preferred theory of justification, mentalist evidentialism (ME), for instance. ME says that doxastic justification is a matter of basing one’s
belief on one’s evidence which provided her with propositional justification.\footnote{See Conee and Feldman (2001) for a thorough presentation and discussion of mentalism.} On ME, a belief that p is propositionally justified for S iff believing p fits the total evidence S has.\footnote{See Conee and Feldman (1985). In particular, pg. 24 for their evidentialist account of doxastic justification, and pg. 15 for their account of propositional justification.} According to ME, one’s evidence amounts to her representational mental states, and ME takes evidential support of some proposition p to occur when p is the best explanation of the contents corresponding to the agent’s mental state evidence.\footnote{See Conee and Feldman (2008: 97-99) for a discussion of best-explanationism about evidential fit/support.}

Here’s one way in which relevance-theory concerns might fail to show that PRJ has less theoretical virtue than ME. It might be the case that ME itself needs to invoke the notion of a reliable belief-forming process, or something sufficiently analogous, in order to explain certain cases of justified belief formation. Michael Bishop, and in some ways Juan Comesaña, have produced arguments for this thesis.\footnote{See Bishop (2010) and Comesaña (2006). See Conee (2013) for a response to Bishop’s argument, and see Matheson (2015) for a response to Comesaña’s argument. In the next chapter, I’ll defend Bishop and Comesaña’s arguments from the counter-arguments of Conee and Matheson.} If this were the case, then ME would share the same sort of lacuna in explanatory depth that PRJ has, given that ME has no supplementary theory of type relevance.

However, I’m more concerned with another problem for a view like ME. Consider the following: in its theory of justification, ME invokes the crucial notions of evidence, evidence possession, and evidential fit (support). Perhaps it turns out that all the proposed accounts of at least one of these notions fail.\footnote{A handful of philosophers have criticized Conee and Feldman’s own accounts of these notions. First of all, consider the notion of evidence itself. Conee and Feldman think that pieces of evidence are mental states. But many, including Alex Arnold (2011) and Timothy Williamson (2002), disagree with this. They each argue, respectively, for a view on which evidence consists of facts and a view on which evidence consists of propositions. Secondly, consider Richard Feldman’s doctrine of evidence possession (see Feldman (1988)). He adopts the view that one’s evidence at a given time t is whatever they are thinking about at time t. Alex Arnold argues that this} Were this to be the case, then ME would possess an
important absence of explanatory depth of its own unique sort. Such an absence of explanatory depth would, presumably, reduce ME’s overall theoretical virtue just like that of PRJ.45

But even if this is not the case, it could still be that PRJ is overall more theoretically virtuous because of its vastly superior explanatory power and elegance. The point is, coming to justifiably believe a premise like D2 would involve checking each competitor theory of PRJ for these ways in which they could end up carrying problematic lacunas of theoretical virtue. Interestingly enough, a thorough discussion of this particular issue is absent from the articles defending the GPO. At this point, I conclude that we’ve been given insufficient reason to believe something as strong as D2.

1.7 Conclusion

Here, at the end of our discussion of the GPO, I have two closing remarks. The first pertains to how we ought to think about the prospects that GPO could eventually be developed into a strong objection to PRJ. The second pertains to evaluating the extent to which the GD dilemma undermines the GPO.

1.7.1 The Future Prospects for the GPO

First, our investigation into the various ways to formulate the GPO reveals that any reasonable attempt to salvage and repair the GPO will be under imminent threat from future...
advances in a theory of type relevance. As was demonstrated in the discussion of the *non-existence of relevant types approach*, we’ve been given no good reason to believe that the reliabilist won’t or can’t make much more progress on a correct relevance theory. Given that all the GPO approaches covered in this chapter invoke, as a premise, our lack understanding or insight into type relevance, any such argument would be refuted once the reliabilists make more significant progress on a relevance theory.

But more importantly, the overgeneralization horn of GD raises an important problem that defenders of the GPO must avoid—namely, the problem of invoking principles that generate defeaters for *their own* favorite views of justification as well. For instance, presumably, Conee and Feldman would *not* be at all pleased if the considerations brought to bear against PRJ undermined ME as well. But as we’ve seen, each of the reasonable approaches to formulating GPO runs the risk of also undermining aspects of ME. Starting with the non-existence of relevant types approach, I suppose that a defender of that approach would attempt to invoke some principle similar to NG that stands a better chance of not undermining such a large amount of other philosophical views. Perhaps it would be a principle that still, in some way, specifies that *some sort* of low theoretical progress can undermine a theory. But if the GPO defender trying to repair this argument is a mentalist, she’ll have to be careful to not invoke a principle that would also undermine ME. We can ask, is it the case that mentalism has made sufficient theoretical progress on the notions of evidence, evidence possession, evidential support (fit), and basing in order to avoid the conclusion that ME is probably false as well? For example, Conee and Feldman have a best-explanationist account of evidential support: for the belief that p to be evidentially supported by evidence e is for p to *best explain* e.

Presumably, whether or not some thesis t best explains some other set of claims partially

---

46 Conee and Feldman (2008: 97-99)
depends on the elegance or simplicity with which t explains those claims. But this means that ME invokes the concept of theoretical elegance as playing a central role in the theory. But it doesn’t seem as if anyone has successfully presented a well-worked out theory of \textit{simplicity/elegance} (in the theoretical virtue sense). To give an example of one unresolved issue, we don’t currently possess helpful principles that determine which factors contribute more centrally to a theory’s simplicity: invoking fewer concepts/types to explain some data, or invoking fewer particulars. There is probably much more to be done in building a complete theory of elegance. So, to the extent that mentalist defenders of GPO use a principles similar to NG to defeat PRJ, they risk undermining their own view. Considering the \textit{theoretical virtue approach}, we’ve already seen how a principle like D3 could undermine a theory like ME to the extent that ME either lacks explanatory depth, or falls short of some other theory’s overall theoretical virtue for some other reason (shortage of explanatory power, or elegance).\footnote{Considering the \textit{no-grasp} approach, ME might be in trouble for a different reason. For reasons mentioned in fn. 43, it is doubtful that mentalists have made much progress on a theory of evidence possession, and it also seems quite implausible that humans are that good at giving anything close to complete descriptions of one’s total evidence for many particular cases of belief-formation. This is because, for many everyday cases of probabilistic belief-formation, one’s body of evidence for particular propositions like, \textit{it is probably the case that the grocery store will be open after work today}, is rather vast and hard to characterize. Large bodies of evidence are often grasped by the subject in varying degrees of consciousness or dispositional storage. The danger is that whatever principle similar to RG1 that the GPO defender might use for generating a defeater for PRJ might plausibly depend on a general necessary condition on grasping epistemic concepts that sets the bar so high that we couldn’t be said to grasp the concept \textit{possessed evidence} either.} Here, I’m just using the one competitor theory ME as an example. But it points to a bigger concern that every kind of theorist who tries to develop the GPO must accommodate—namely, to avoid invoking principles that also undermine one’s own theory of justification.

1.7.2 The GD Dilemma, and the Current State of the GPO

At this point, there are two possible conclusions one might draw from my discussion: First, and more modestly, we’ve yet to be presented with any good reason to think PRJ is
defeated based on our lack of understanding of type relevance. Secondly, and more ambitiously, we have good reason to doubt that any sort of good GPO argument against PRJ could ever be formulated. In other words, the GPO is a non-starter. I think the first and more modest conclusion has been rather thoroughly established. In §3, I showed that the way the defenders of the GPO explicitly frame the argument (the theory incompleteness approach) succumbs to the GD dilemma.

But I also think that the considerations in this chapter support the more ambitious conclusion that the GPO is a non-starter. Beyond the theory incompleteness approach, I go on to show that three other reasonable approaches to formulating the GPO succumb to the GD dilemma as well. To be sure, whether I’ve successfully supported this more ambitious conclusion depends crucially on whether the candidate formulations of the GPO examined in this chapter in fact represent the most reasonable general strategies for formulating a defeater for PRJ based on our current lack of progress on answering the generality problem question. While I don’t have a positive argument for the representativeness of the four approaches and specific formulations that I investigate in this chapter, I do think their representativeness has strong intuitive appeal. In other words, upon considering the breadth and nature of the argument formulations I present above, it seems reasonable that they do adequately represent the plausible ways one might try to formulate the GPO.

To conclude, I believe that our inquiry to this point highlights a sobering result for one who would press the generality problem as a clear devastating objection to PRJ (or process reliabilism about warrant for that matter). As we saw above, much of the rhetoric used to describe the severity and force of the generality problem is quite extreme, e.g., it is a “serious problem” for the theory, “process reliability theories of justification and knowledge look hopeless,” etc. This sentiment is profoundly misplaced, given that the most plausible
formulations of the GPO either prove far too much, or make claims about reliabilism that we have no reason to accept.
CHAPTER 2

A DEFENSE OF PARRYING RESPONSES TO THE GENERALITY PROBLEM

2.1 Introduction

Process reliabilism is an important contemporary externalist approach for analyzing epistemic concepts. Both a process reliabilist theory of warrant and a process reliabilist theory of justification have been presented and defended in the literature.¹ Process reliabilism about justification and process reliabilism about warrant share this common feature: they both claim that the agent’s use of a reliable belief-forming process to arrive at her belief is the key ingredient that grounds the belief’s possession of the relevant epistemic property (whether warrant or justification).²

While there have been many criticisms of reliabilist theories of warrant and justification, I want to address one particular objection that’s become rather popular throughout the past three decades: the generality problem.³ The generality problem begins with a puzzle. Processes

¹ By warrant, I use this term in the functional sense much like Plantinga does, where warrant is the state the possession of which grounds one’s having knowledge that p so long as she also possesses belief that p and p is true (Plantinga (1993: v)).

² For an early formulation of a reliabilist account of justification, see Goldman (1979:13-14) for an account prima facie justification and (pg. 20) for the additional no defeater condition (for ultima facie justification). For an account of warrant in which reliability figures as the central grounding feature, see Goldman (1986:44-5) presents a reliabilist theory of warrant, on which warrant requires that an agent’s belief-forming process has both “local” and “global” reliability.

³ See Goldman and Beddor (2015). In their overview of the work done on reliabilist epistemology, they include the generality problem as one of the top six “problems,” or, “objections” to reliabilism.
can be thought of as general repeatable types, or as precise one-off tokens. Plausibly, only the former can be evaluated for reliability or unreliability. As both defenders and critics of reliabilism have noticed, a given process token exemplifies many types. A given case of seeing a red ball at close distance—and then coming to believe a red ball is in front of me on the basis of that visual experience—exemplifies the following types: [belief formation], [visual belief formation], [visual belief formation in good light at a close distance], etc. Reasonably, not all of these types are measured for reliability in determining whether the target belief has warrant or justification. Indeed, perhaps only one of them is relevant for determining justification or warrant. But which type is the relevant one? This very question constitutes the starting off point for the generality problem.

According to the literature, this generality problem question turns into the generality problem objection to reliabilism upon realizing that humans—including the best philosophers—do (and have done) a poor job answering the generality problem question. It is our lack of understanding into the notion of type relevance that constitutes the key premise in the generality problem objection. Over the course of the past three decades, various philosophers have

---

4 Richard Feldman (1985) and Goldman (1979) are key figures who highlighted this important distinction for making sense of reliabilism. Most philosophers agree that types, rather than tokens, are the entities that can be measured for reliability. Although recently, Juan Comesaña (2006) articulates a way in which tokens could be evaluated for reliability with respect to some space of possible worlds as a reference class. Even if this way of conceiving measuring a token for reliability makes sense, Comesaña, correctly, recognizes that framing reliabilism like this doesn’t get the reliabilist out of the generality problem. The reliabilist still would have to provide an account of which possible worlds were contained within the reference class used to evaluate the token’s reliability. See 2.2.3 for a discussion regarding the close connection between the reference class for measuring reliability and the relevant type.

5 For example, presumably, [belief formation] isn’t the type that gets evaluated for reliability. Consider someone who, sadly, is a BIV, such that their perceptual beliefs are all systematically false. Such a person, presumably, can still have reliable and justified intuition beliefs in a priori claims. But, if the relevant type for these intuition belief tokens was just [belief formation], then the reliability measurement would be significantly decreased from all of the false perceptual beliefs formed by the agent (since perceptual belief formation is contained within the more general category of belief formation).

6 See Feldman (1985:160) for the introduction of the term “relevant type,” which has become standard terminology in the generality problem literature to denote the type exemplified by a token whose reliability measurement determines whether or not a given epistemic property is exemplified.
presented theories of process type relevance—i.e., theories of what determines type relevance for belief-forming process tokens.\(^7\) These theories are typically constituted by principles or conditions stating which sorts of features of a token are “held fixed” in the relevant type descriptions for those tokens. Critics have been quick to point out crippling difficulties with the relevance theories that have been proposed. Earl Conee and Richard Feldman are the arch defenders of the generality problem objection, and it turns out that others are inclined to agree with them in their critical stance towards the relevance theories currently on offer.\(^8\) Conee and Feldman conclude that the generality problem objection makes “process reliability theories of justification and knowledge look hopeless.”\(^9\)

Here, it is important to point out that very little work has been done to formulate exactly just how our lack of ability to answer the generality problem question generates a defeater for reliabilism. How might such an argument go? It is beyond the scope of this chapter to attempt to formulate an argument that represents the strongest version of the generality problem objection.\(^10\) However, I contend that proponents of the generality problem objection hold

---


\(^8\) For the most thorough and popular criticism of many of the relevance theories on offer, see Conee and Feldman (1998) and then supplementary critical articles by Brueckner and Buford (2013), Conee (2013), Dutant and Olsson (2013), Matheson (2015), and Conee and Feldman (2002). According to Feldman (1985), a solution to the generality problem falls into the no-distinctions problem if it types processes too broadly, so that intuitively justified cases of belief formation get ruled as unjustified (or vice versa). A solution to the generality problem falls into the single-case problem if it types processes too narrowly, such that any true belief will come out as justified, and any false belief comes out as unjustified (which is intuitively incorrect). Conee and Feldman (1998) argue that every candidate relevant theory presented in the literature either falls into one of these two problems, or fails to deliver specific verdicts on which type is relevant.

\(^9\) Conee and Feldman (1998: 24). It turns out that others are inclined to agree with them. For example, see Plantinga (1993: 28-29), Matheson (2015).

\(^10\) Indeed, in chapter 1, I argue that successfully formulating an argument against reliabilism about justification from a premise involving our lack of understanding of type relevance doesn’t look promising. It seems like any reasonable formulation either invokes principles that would generate a defeater for \textit{any philosophical theory} (and hence we should reject these principles), or invokes premises that we have little to no reason to believe.
that the strongest formulations of this objection all in some way (either explicitly or implicitly) depend on the following key claim: Reliabilist theories have an explanatory burden to produce an informative relevance theory because they invoke type relevance as playing an important explanatory role, and that failure to fulfill this burden significantly undermines the reasonableness of reliabilism. Presumably, one who defends this key idea thinks that it follows from a more general principle about explanatory burdens that come with invoking relevant process types, rather than following from some other particular features of reliabilist epistemic theories. We can formulate this general principle as follows:

Process Type Burden (PTB)
Necessarily, for all epistemic theories, if an epistemic theory invokes relevant process types (as playing a grounding or explanatory role) that share a similar degree of complexity to the relevant types invoked by reliabilism (either about justification or warrant), then that epistemic theory has an explanatory burden to provide an informative theory of type relevance.

Hence, in some way, we should conceive of the most reasonable formulations of the generality problem objection as being PTB-based arguments.

One rather straightforward way to respond to PTB-based arguments that readily suggests itself is simply to find and defend a correct relevance theory. If reliabilists make significant progress on a relevance theory in the future, then any PTB-based arguments will at that point be undermined. Call this response strategy the “head-on” response.

---

11 Here’s an illustrative quote that references the sorts of considerations that are present in the way defenders of the generality problem objection explain how the argument works:

Of course, the arguments of this paper do not show that no acceptable version of the reliability theory can be constructed. However, it is fair to say that The Problem of Generality is a serious problem for the theory… To make the reliability theory plausible, then, some other way must be found to specify processes, some way that assures that only reliable processes operate in cases in which one's evidence does support a belief adequately and only unreliable processes operate when one's evidence fails to support a belief. While it may be possible to come up with a general account of processes that satisfies this requirement, I believe that the prospects for doing so are not good. (Feldman 1985: 172 emphasis mine)
But there’s another way to respond to PTB-based arguments. Let’s say that a given epistemic theory satisfies PTB just in case it satisfies the antecedent of PTB, i.e., just in case it too invokes relevant process types roughly as complex as those invoked by reliabilism as playing some important explanatory role. If it can be shown that other reasonable candidate epistemic theories satisfy PTB, then it follows that PTB would generate an objection to these other theories insofar as they too lack successful relevance theories. Showing that some other theory T1 satisfies PTB as well would constitute a sort of *tu-quoque* objection to those who both hold to T1 and defend the generality problem objection against reliabilism. Call this sort of response to PTB-based arguments a *parrying response*. Let’s say that a parrying response R applies to a given theory T if and only if R successfully shows that T satisfies PTB. Parrying responses can be stronger or weaker depending on the number of competitor theories to which they apply. For simplicity’s sake, call a *strong* parrying response a parrying response that applies to many epistemic theories. Strong parrying responses, in addition to constituting a *tu-quoque* response to individual epistemic theories to which it applies, offer further reason to reject the generality problem objection in virtue of showing that the generality problem objection simply *proves too much*. Objections that undermine almost every candidate view in a domain can often be reasonably dismissed.

Given the differences between warrant and justification, one can technically distinguish a generality problem objection that uniquely targets process reliabilism about justification from a generality problem objection that targets reliabilism about warrant. Because of this, one can also distinguish two different sorts of parrying responses:

*Justification Parrying Responses* to the generality problem objection argue that other candidate theories of justification satisfy PTB.

*Warrant Parrying Responses* to the generality problem objection argue that other candidate theories of warrant satisfy PTB.
It is possible that one kind of argument might constitute a strong warrant parrying response, while failing to constitute a strong justification parrying response (and vice versa). This would be the case if, for instance, the argument successfully shows that many candidate theories of warrant satisfy PTB, while failing to show that many candidate theories of justification satisfy PTB.

In this chapter, I’m concerned with developing three sorts of parrying responses that have recently received attention in the literature, and arguing that they are in fact strong parrying responses. In §2.2, I present Michael Bishop’s reflective justification parrying response, which I take to be both a warrant parrying response and a justification parrying response. I then respond to Earl Conee’s strategy for undermining Bishop’s parrying response. In §2.3, I engage with Juan Comesaña’s head-on response to the generality problem, and then briefly recount Alex Arnold and Jon Matheson’s criticisms of it. After taking stock of how Comesaña’s head on response fails, I formulate two distinct new parrying response arguments one could glean from his discussion of the generality problem. The first parrying response exploits the notion of evidence possession, and constitutes both a warrant and justification parrying response. The second parrying response invokes the notion of competent basing, and constitutes only a warrant parrying response. I argue that both of these, just like the reflective justification parrying response, are strong parrying responses.

2.2 Reflective Justification and Bishop’s Parrying Response

2.2.1 Bishop’s Parrying Response

Michael Bishop thinks that every good epistemic theory of justification needs to account for how reflective justification can contribute to one’s doxastic justification for believing some
target proposition p. According to Bishop, S’s belief that p has reflective justification just when “S’s belief is justified on the basis of S’s knowledge that she arrived at it as a result of a highly (but not perfectly) reliable way of reasoning” (Bishop 2010: 286). So, reflective justification is a sort of inferential justification.12 More precisely, S’s belief that p is reflectively justified if and only if S competently infers p from a chain of reasoning constituted by one premise that identifies a particular belief-forming method M as being reliable, and another premise identifying that M delivers the judgment that p, where S is justified in believing these premises and the premises are true.13

Here’s a testimonial example of reflective justification inspired by one of Bishop’s cases (2010: 289).

SUSAN
You know that Susan studied US history in graduate school. You have good reason to believe that she is a reliable testifier in the domain of US history. From this you come to justifiedly believe that

\[ S_1 \text{ Deferring to Susan about US history is a reliable way to form beliefs about US history.} \]

You ask Susan a US history question, and she delivers the answer that p. You then believe that

\[ S_2 \text{ Deferring to Susan about US history would lead me to believe } p. \]

---

12 Reflective justification should not be confused with higher order justification (i.e., justification to believe one is justified in believing p). Here, Bishop is introducing reflective justification as a technical term that refers to a kind of belief formation that he thinks is rather common in normal everyday life.

13 Technically, the inference pattern in a case of reflective justification could include other supporting premises to the two main ones given here. Also, in the quote on 286, Bishop claims that it is one’s knowledge of the premises constitutive of a reflective justification that’s important, rather than mere justified belief. In my own reconstruction of reflective justification, I only require that the agent justifiedly believe these two premises and that the premises are true. I do this for two reasons: First, I think knowledge of the two premises is too demanding for an account of reflective justification. Secondly, my weaker account of reflective justification still gets Bishop’s argument off the ground, and invokes necessary conditions on reflective justification that would also be accepted by those who prefer the more demanding knowledge condition as well.
Then, you competently perform the inference from S1 and S2 and come to hold p as a result.

We could multiply examples here. For instance, one could replace a reliable testifier about US history with a reliable thermometer giving temperature readings in a particular region and then make the relevant question about the temperature for that region. Bishop claims that, in everyday life, we frequently solve problems and form beliefs in ways that exemplify the general pattern of reflective justification. He also claims that inferences that satisfy the requirements for reflective justification clearly confer justification on their target beliefs (288).

Given that reflective justification in fact confers justification on our beliefs, Bishop argues that any plausible epistemic theory has to account for how cases like this confer justification (289). Now, consider S1. Every epistemological theory will need to capture how facts like S1 ground or explain why competently performing a reflective justification inference confers justification on the target belief. Yet notice, the notion of a reliable way to form beliefs about x is a constituent of S1. The phrase “a way to form a belief” just refers to a belief-forming process. Also, belief-forming processes are only reliable or unreliable considered as belief-forming process types. Here, the generality problem emerges: what is the correct relevant process type referred to in S1? What does the type, deferring to Susan about US History, amount to? Is it believing whatever she says about US history? Is it believing whatever she says about US history in cases in which she’s not drunk? Is it believing whatever she says about US history in cases in which she is sober and not sleep-deprived?, etc. The list goes on. But only one of these is the relevant type description for the method referred to by S1. Hence, given that there’s one relevant type referred to in S1, and that any good theory of justification will need to capture how the S1-S2 inference confers justification, then any candidate theory of justification
will have to invoke relevant process types. Hence, any reasonable candidate theory of justification satisfies PTB. I take this to be the essence of Bishop’s argument.

Technically speaking, it is clear that the reflective justification parrying response doesn’t apply to all theories of justification. Consider phenomenal conservatism, which claims that a belief that p is prima facie justified for S if it seems to S that p is true. Bracketing concerns about ultima facie justification, it is open to a phenomenal conservative to simply claim that the premises of a reflective justification inference don’t ground the justification S gets for p in any central way. Rather, the only entity responsible for grounding S’s justification is the *seeming that* p which S receives upon completing the inference. This being the case, the phenomenal conservative could deny that the reflective justification parrying responses shows that her view satisfies PTB.

Nevertheless, I think that the reflective justification parrying response applies to a great many views of justification, thus making it quite strong. In addition to being a justification parrying response, I think it also constitutes a strong warrant parrying response. The reflective justification inference pattern also seems to be satisfied in many cases of inferential warrant. Given that any reasonable theory of warrant needs to account for how inference with premises like S1 can confer warrant on a target proposition, there’s good reason to believe that any reasonable theory of warrant satisfies PTB as well.\(^\text{15}\)

\(^{14}\) See Huemer (2001: 99) for a similar formulation.

\(^{15}\) Consider once again the phenomenal conservative about justification. While there are many phenomenal conservatives about justification, there seem to be fewer phenomenal conservatives about warrant. So, reflective justification can still get traction as a warrant parrying response for many individuals who are phenomenal conservatives about justification.
2.2.2 Conee’s Rejoinder

Earl Conee demurs. He denies that countenancing reflective justification inferences commits a theory to invoking relevant belief-forming process types. Conee does think that cases of reflective justification give “an evidential role to certain justified beliefs about reliability” (Conee 2013: 760). However, he denies that these justified beliefs attribute reliability to particular belief-forming process types.

I’ll fill out Conee’s account below, but let me flag here that Conee’s response implies that Bishop has technically mischaracterized cases of reflective justification. Clearly, S1 does attribute reliability to a particular belief-forming process type. S1 claims that a certain “way” of forming beliefs is reliable. Conee’s strategy entails that beliefs like S1 aren’t actually playing a role in generating reflective justification.

So, if reliable belief-forming processes aren’t constituents of premises in reflective justification inferences, what other epistemic factor is involved? Conee claims that reflective justification inferences involve premises that ascribe reliability to things like tests (e.g., a thermometer) and testifiers rather than to belief-forming processes. So, instead of S1, Conee thinks that the relevant proposition constitutive of one’s reflective justification in SUSAN is

\[ S1^* \quad \text{Susan is a reliable testifier about US history.} \]

One might now wonder about the specific nature of the epistemic support that a claim like S1*, in conjunction with S2, gives to the target proposition \( p \). Conee explains as follows:

The evidential relevance of the reliability propositions to the justification of the target beliefs in the [test and testimony] cases is to give the believers broadly inductive evidence for the truth of those beliefs. Nearly enough, each justified reliability proposition asserts that the corresponding believed proposition has a certain contingent property that is shared mostly by truths. Having justification for those propositions that place the target beliefs in contingent classes that contain mostly truths gives the believers routine inductive reasons to think that the target beliefs are true. (761 emphasis mine)
So, by Conee’s lights, believing $S1^*$ and $S2$ in the Susan case is a special case of having good \textit{inductive} support for a target proposition $p$.

2.2.3 Bishop’s Parrying Response Defended

I’ll grant, for the sake of argument, that Conee’s inductive support proposal shows how cases of reflective justification needn’t invoke relevant \textit{belief-forming} process types. However, this move doesn’t really undermine Bishop’s parrying response. This is because the inductive support proposal invokes a kind of reliability that applies to a different sort of process: what I call \textit{information delivery} processes.

To begin, consider the concepts \textit{reliable testifier} and \textit{reliable test}. What are these? What does it take to be a reliable testifier or a reliable test? First, it is important to recognize what testifiers and tests do: they communicate/deliver information. In other words, these testifiers or tests are important constituents in events characterized by some piece of information being delivered. Testifiers and tests can deliver information that’s either true or false. As a first pass, a reliable testifier is one that reliably delivers \textit{true} information. But what does it mean for a testifier to reliably deliver true information? Importantly, the concept \textit{reliable testifier} is always relativized to a particular set of \textit{conditions}. For instance, when we say that Susan “is reliable,” what we mean is that she’s reliable given a very specific characterization of her testimonial acts. For instance, she’s reliable with respect to certain sorts of contents she might deliver—namely, contents about US history. Also, she is reliable only in certain circumstances. She’s not reliable at answering questions about US history when she’s starving, totally dehydrated, just woken up from a deep sleep, being paid to lie to everyone she talks to, about to skydive, having a panic attack, or having just taken large amounts of marijuana. So, the idea that “Susan
is reliable” is more precisely framed as “Susan is reliable about a given domain of claims when asked in circumstances of type C.” This is what we actually mean when we note Susan’s reliability.

The same goes for the concept reliable thermometer. With a good thermometer one can buy at the store, no one thinks it is reliable at delivering the correct temperature in temperature ranges above 700 degrees Celsius. Indeed, at those temperatures the thermometer would be melting! It is also the case that a standard mercury based thermometer won’t be reliable under conditions of extraordinary air pressure. So again, our idea of a reliable thermometer is more precisely seen as a thermometer that is reliable for a given domain of temperatures under circumstances of type D. Both the concept of a reliable thermometer and the concept of a reliable testifier share circumstance and domain qualifications.

Next, consider the metaphysical workings of the reliability measurement that pertains to whether some testifier is reliable. When we evaluate for reliability here, we’re invoking a certain sort of ratio: the ratio of cases in which the testifier delivers true information to those in which she delivers false information, across a certain class of cases/situations. The class of (actual or possible) cases, with respect to which a reliability measurement is taken, is called the reference class. For evaluating the reliability of a testifier, the cases constituting the corresponding reference class are comprised of particular information delivery process events, in which that testifier is delivering some piece of information. But our investigation above also reveals that two other important features are held fixed in all the reference class cases corresponding to the reliability measurement of a testifier: a relevant circumstances description, and a relevant domain description.16

Importantly, belief-forming process types share the same sort of relationship with corresponding reference classes. Consider once again the case of visually coming to believe

16 The same also holds for reference classes corresponding to tests.
there's a red ball in front of me. When we're wondering whether S's belief-forming process is reliable, we're only concerned with the ratio of truth to falsehood S would have across a class of possible belief-forming cases (i.e., possible belief-forming events) in which S is forming judgments with a particular sort of content in a particular kind of circumstance. For instance, in this particular case of visual belief formation, it is irrelevant whether S is competent when it comes to abstract mathematical belief formation, or how accurately S would form visual beliefs while also undergoing brain surgery.

Here, I posit the following relationship between the relevant type and the reference class: all of the event particulars that constitute the reference class for a given token t exemplify t's relevant type. Seeing this is rather straightforward. It doesn't make sense for the reference class for a belief-forming token to include cases that don't exemplify that token's relevant type description, given that the reliability measurement for that token is determined by the ratio taken across all the particulars of the reference class. Hence, if a theory invokes a reliability measurement taken across some reference class, it follows that the theory thereby invokes a relevant type description that's exemplified by all of the event particulars constituting the reference class. It also follows that a theoretical account of what determines the features

---

17 Given the characterization of the generality problem according to Feldman (1985) and Conee and Feldman (1998), the problem for the reliabilist is that, although we can know the features of a given token, we don't know how to perform the reliability measurement for that token to determine whether that token confers justification or warrant on the target belief. For the token case of seeing and coming to believe in the presence of a red ball (discussed in the introduction), we don't know whether to perform the reliability measurement with respect to [belief-formation] in general, or with respect to [perceptual belief-formation], or with respect to [visual-belief formation], etc. But the more precise way to formulate this puzzle is in terms of the reference class corresponding to the token: we don't know whether the token's reference class is comprised of particular cases of all sorts of belief formation, or just with particular cases of perceptual belief-formation, or just with cases of visual belief-formation, etc. If we could completely determine the features of a token's reference class, then this would effectively answer the questions Conee and Feldman raise to get the generality problem going.
held fixed across a token’s reference class suffices for being an account of what determines a token’s relevant type.¹⁸

As we’ve seen, both the concept of a reliable belief-forming process and the concept of a reliable testifier correspond to respective reference classes of cases in which some content and condition description is held fixed throughout. If my above observation regarding the relationship between relevant process type descriptions and reference class descriptions is right, then it is the case that one is committed to there being relevant process types of some sort if one invokes reliable testifiers. Here, I’ll simply call them relevant information delivery process types. Clearly, these relevant process types play a central explanatory role in Conee’s account of reflective justification.

This discussion highlights how explicating reliable testimony doesn’t seem in any way importantly different or less complex than explicating reliable belief formation. Therefore, Conee’s response to Bishop hasn’t succeeded in showing how a reasonable candidate epistemic theory might escape satisfying PTB while still accounting for reflective justification.¹⁹

¹⁸Wallis (1994:251–262), argues that the relevant type and the reference class are distinct, and hence there are actually two questions raised by the generality problem: First, what is the relevant type for a token? Secondly, what is the reference class for measuring reliability for a token? He claims that making progress on the generality problem requires keeping these two questions (and their answers) distinct in our theorizing. Here, however, he doesn’t specifically argue why, methodologically, it is important for addressing these two questions separately. As I’ve articulated above, if one offers an account of the reference class, one thereby also succeeds in offering an account of the relevant type—given that all of the reference class cases share the relevant type description in common.

¹⁹It is not surprising that Conee’s inductive support theory of reflective justification invokes rather complicated process type descriptions. Inductive support is a sort of probabilistic support. Probability theory comes out of measure theory. When we consider the probability that some state G will obtain, in some way we’re comparing the “measure” of relevant situations in which G obtains with the measure of relevant situations in which G doesn’t obtain. But the “relevant situations” here constitute a reference class, without which all talk of probability won’t make sense. Since any probabilistic judgment necessarily invokes a reference class, Alan Hajek notes that every theory of probability has the reference class problem (2007)—namely, the problem of specifying the reference class for a given probability measurement. Reichenbach (1949) formulates the problem of determining the reference class in terms isomorphic to the generality problem for process reliabilism.

If we are asked to find the probability holding for an individual future event, we must first incorporate the case in a suitable reference class. An individual thing or event may be incorporated in many reference classes, from which different probabilities will result. This ambiguity has been called the problem of the reference class. (374)
2.3 Comesaña’s Response to the Generality Problem

2.3.1 Comesaña’s Head-On Response, and a Simple Parrying Response

The ingredients for a rather simple parrying response to the generality problem objection can be seen in Juan Comesaña’s head-on solution to the generality problem. Comesaña claims that any good epistemic theory of justification needs to have an account of doxastic justification, not just propositional justification. The important difference between propositional and doxastic justification is illustrated by the following two sorts of cases. First, S can possess evidence that adequately supports believing a proposition without S’s actually believing that proposition. This is a case of having propositional justification without doxastic justification in virtue of never forming the belief. Second, consider someone who possesses good empirical evidence to believe p: a rare species of eagle is currently nested in a particular region of the Andes Mountains. Suppose the agent believes p, but sadly believes p on the basis of flipping a coin rather than on his good evidence. Here, the agent’s belief is not doxastically justified, even though believing the proposition is propositionally justified for him given his evidence. Most epistemologists agree that in order to have doxastic justification, one must base her belief on the evidence she has that sufficiently supports belief that p.

In addition, it is reasonable that a belief B’s being based on evidence E at the very least entails that S possesses E and that B was non-deviantly caused by the mental states that

---

20 Much like the propositional justification/doxastic justification distinction, there is a similar distinction for warrant: possessing a warrant for believing p vs warrantedly believing p (which amounts to basing p on your warrant). While no one thinks these distinctions are identical, it appears as if Conee and Feldman have a notion of possessing propositional justification that comes close to possessing a warrant. So, even though I’ll frame the discussion in this section in terms of propositional/doxastic justification, the lessons learned here could equally apply to a discussion of warrant.
represent $E$. Hence, as Comesaña points out, every epistemic theory needs to posit the following sort of belief-forming process in order to adequately account for doxastic justification: the *basing on evidence* process.

Comesaña continues,

> It follows that we have good reasons to believe that any adequate epistemological theory needs to appeal, either implicitly or explicitly, to the notion of a belief’s being based on certain evidence. In the next section I will argue that that notion is all we need to solve the generality problem. If this is so, then any adequate epistemological theory is going to have the resources to solve the generality problem. (2006: 33)

Comesaña thinks that the correct answer to the generality problem is revealed once the reliabilist sees *basing on evidence* as the central feature of relevant belief-forming process types.

> Given that there will always be some evidence that the belief is based on, the process that generates the belief will always instantiate a case of the type-schema *producing a belief that $p$ based on evidence $E$*. (2006: 37)

Comesaña claims that this solution to the generality problem “should be accepted not only by Conee and Feldman, but by anyone who thinks that an epistemological theory is incomplete without an appeal to the basing relation” (38).

If Comesaña’s head-on solution to the generality problem succeeds, then PTB-based arguments fail. For what it is worth, Comesaña’s head-on response raises a simple parrying response to the generality problem objection as well. If Comesaña is right, then every candidate epistemological theory invokes the process type *basing belief that $p$ on evidence $E$* (where $p$ and $E$ here are variables). If this type *just is* the relevant type that reliabilists invoke in their theory, then every candidate epistemological theory invokes the same sorts of types invoked by reliabilism. Hence, every candidate epistemological theory satisfies PTB.
2.3.2 Refuting Comesaña’s Head-on and Simple Parrying Responses

But there are problems with Comesaña’s treatment of the generality problem. First, his head-on response falls prey to some compelling objections. To begin, Alex Arnold points out that typing processes according to the producing a belief that \( p \) based on evidence \( E \) schema delivers incorrect verdicts for which types are relevant in cases of believing necessary truths (2011: 128). Suppose Goldbach’s conjecture is true, and someone comes to believe it by basing the belief on the claim lollipops are delicious. It seems as if this belief would be unjustified, but the type producing a belief that Goldbach’s conjecture is true based on evidence \( E \) for any \( E \) (including lollipops are delicious) will yield a reliably formed belief. This is because every metaphysically possible belief-forming scenario which satisfies this type will be a scenario in which a true belief is formed. After all, Goldbach’s conjecture is true in every possible world/situation, including the possible situations in which one comes to be believe Goldbach’s conjecture on the basis of some other claim. This means that the reference class corresponding to this type will manifest a perfect reliability measurement. Hence, believing Goldbach’s conjecture on the basis of lollipops are delicious would end up being justified (perhaps maximally so), and this is clearly the wrong result. Perhaps there’s some patch one could make for the view to correct this problem with necessary truths.

There’s an even more problematic feature with Comesaña’s treatment of the generality problem, which undermines both his head-on solution, and the simple parrying response one could glean from it. As Jonathan Matheson points out, Comesaña’s head-on solution doesn’t tell us which possible situations/events—that satisfy the description producing a belief \( p \) on the basis of evidence \( E \)—are epistemically relevant for the reliability measurement and belong in the reference class (2015: 467). Does the reference class include every possible belief-forming
situation characterized by the description *believing p on the basis of E*? It is doubtful that reliabilists would invoke a reference class this broad.

Consider the following example: Someone is walking around in her environment and comes upon a red ball and subsequently gains the evidence E1 of a visual experience representing a red ball in front of her. Suppose that, on the basis of E1, she comes to believe that p1: *there is red ball*. It is doubtful that the correct reliability measurement is taken across all possible worlds in which S comes to believe p1 on the basis of E1. Presumably, the reliabilist thinks that certain external factors (from S’s environment) are relevant to whether or not S’s belief in this case is reliably formed. These external factors might include whether or not there are hidden red lights present in the area that tend to illuminate white objects to look red, whether people have set up red ball facades in the general area of S, whether or not there’s a malicious demon in S’s environment who likes to deceive agents like S, etc. According to reliabilism, then, there’s some description (C) of S’s environment that must be held fixed in the reference class in order to determine the intuitively correct reliability measurement. Therefore, the crucial information required for measuring reliability, for a given token, is captured in the following schema instantiated by the token: *believing p on the basis of E in circumstances C*, not just *believing p on the basis of E*. But how do we determine how narrowly or broadly to construe C for each token case? Unfortunately, Comesaña doesn’t say anything about the circumstances C description that’s held fixed across a reference class. This being the case, Matheson helpfully points out that Comesaña’s head-on solution to the generality problem is woefully incomplete.

Drawing on this shortcoming, we can see that the simple parrying response, raised by Comesaña’s head-on response, fails as well. Before explaining why it fails, it is instructive to present Matheson’s favored theory of justification: Evidentialism. For our purposes here, it
is helpful to see evidentialism as denoting a family of specific views that all analyze justified belief in terms of evidence, evidence possession, evidential support, and basing rather than the reliability of belief-forming process types. What differentiates the various specific evidentialist theories are their differing accounts of evidence, evidence possession, and evidential support. As a matter of fact, the most ardent supporters of the generality problem objection against reliabilism are evidentialists, e.g., Conee and Feldman. It is also the case that these evidentialists who defend the generality problem objection, for the most part, formulate their evidentialist theories to be internalist in nature.²¹ Technically speaking, not all evidentialist theories are internalist, but for simplicity’s sake in this chapter, we’ll let “evidentialism” denote an internalist family of views.²²

Against the simple parrying response one could draw from Comesaña’s head-on response, Matheson argues,

While evidentialism does claim that a certain belief-forming process is relevant for doxastic justification, evidentialism does not claim that the justification of a belief results in any way from any property of that belief-forming process type. In contrast, evidentialism claims that it is properties of the evidence that are relevant. Since according to evidentialism no property of the belief-forming process—believing P on the basis of E—is relevant to the justificatory status of a belief, the question of which possible worlds are relevant to assessing some property of that belief-forming process type simply does not arise. Since no property of the process type is relevant, no questions surrounding how to evaluate any such property are relevant either. (2015: 467-8)

Matheson is saying that evidentialists don’t think that S’s justification depends at all on a reliability measurement taken across a class of possible belief-forming events in which S bases p on E. Rather, justification is simply determined by an internal relation between E and p

²¹ By internalism about justification, I merely mean the thesis that internal duplicates are also justification duplicates. See Conee and Feldman for a specific defense of this internalist supervenience thesis for their version of evidentialism (2001: 3-5).

²² For examples of externalist evidentialist views, see Arnold (2011: 162-172) and, arguably, Alston (1985) and (1988).
itself. It follows from this that evidentialists needn’t invoke anything like the *circumstances C* description to hold fixed across a reference class. However, the circumstances C component of the reference class is a crucial theoretical posit that plays a central role in grounding justification and warrant verdicts for the reliabilist. Given that the evidentialist doesn’t posit a complex reference class with this *circumstances C* qualification, she lacks a significant explanatory burden that the reliabilist possesses. I take this to be a rather compelling refutation of the simple parrying response one might glean from Comesaña’s account.

Nevertheless, I contend that Comesaña’s simple parrying response points, in a less straightforward way, to two stronger parrying responses that the reliabilist could make. Here, I’ll specifically argue that these parrying responses even apply to reliabilism’s main foil in this debate: evidentialism. First, Comesaña’s arguments highlight the importance of accounting for evidence *possession*. It turns out that explaining this phenomenon generates a strong justification and warrant parrying response. Secondly, there’s a certain *kind* of basing that’s required for warrant, and explicating this more robust form of basing generates a strong warrant parrying response that is effective against evidentialist theories of warrant.  

2.3.3 Evidence Possession and a Parrying Response

Like Comesaña claims, it is quite plausible that every good theory of justification and warrant will need to account for how one’s evidence factors in to whether and how she knows or justifiably believes some claim. This means that every good epistemic theory needs to have some account of evidence possession. In what follows for this section, for simplicity, I’ll only

---

23 By an “evidentialist theory of warrant,” I mean a theory that roughly holds that *basing* *p on good enough evidence E* is a crucial necessary component to having warrant. These views might also posit an additional anti-luck condition as well.
write in terms of justification, but I take the relevant upshots of this discussion of evidence possession to apply equally well to warrant.

To begin, understanding evidence possession requires some idea about what evidence itself is. Some hold that propositions are evidence, and some hold that facts are evidence. On these two views, the account of evidential support is rather straightforward, since intuitively it seems like propositions and facts themselves are the sorts of entities that can stand in evidential support and evidential disconfirmation relations to other claims one may or may not believe. Clearly, it is not the case that every fact or every proposition is automatically possessed by someone as her own evidence. Factual and propositional accounts of evidence typically impose a mental representation requirement on evidence possession: in order for S to possess evidence E (where E is a fact or proposition), S must have a mental state that represents E as being true. Here, these mental states amount to beliefs or representational experiences (like perceptual experiences). Conee and Feldman’s particular evidentialist view, called mentalism (M), on the other hand, rejects factual and propositional accounts of evidence. They claim that the representational mental states themselves are the evidence that an agent has. On M, evidential support for some claim p happens in virtue of the representational contents of an agent’s mental state evidence standing in supporting relations to p.

Here, I won’t engage in arguments about which account of evidence is correct. But for simplicity moving forward, I will write in terms of a propositional view of evidence. One can easily translate what I say below into a mentalist account of evidence in the following way:

---

24 See Arnold (2011:144-155) for a defense of the claim that facts are evidence. Arnold distinguishes facts from propositions in the following way: facts are what true propositions are about, and facts can be partially concrete (rather than fully abstract) if some of their constituents are concrete. For instance, the fact that earth revolves around the sun is partially concrete in virtue of the earth and the sun being concrete objects. See Dougherty (2011) and Williamson (2002) for a defense of the view that evidence is propositional.

take sentences of the form “S possesses proposition e as evidence” to actually mean “S possesses the mental state m, that represents content e, as evidence.”

So, we can ask the following question: what are the sorts of contentful mental states that can represent the propositions constitutive of one’s evidence? There seem to be two reasonable candidates to choose from:

- **Occurrent Thesis (OT)** Necessarily, S’s evidence for p at t can only consist of propositions that S is occurrently representing at t.

- **Moderate Thesis (MT)** Necessarily, S’s evidence for p at t can consist of propositions that S is representing either occurrently or dispositionally at t.

I think there’s good reason to reject OT. Alex Arnold argues that the OT thesis cannot generate the correct justification verdicts on a variety of particular cases. For one, he argues that, intuitively, OT delivers the incorrect result regarding justified belief (and knowledge) while the subject is sleeping (Arnold 2011: 95). Intuitively we know many things while we’re asleep. It is not as if our knowledge vanishes when we sleep. But, it is also the case that for large portions of dreamless sleep, we aren’t conscious, and hence have zero occurrent mental states. But this means that we have no evidentially supported beliefs while we sleep, and hence have no justification and knowledge at those times. This can’t be right. Richard Feldman, who as of 1988, supported OT, responds to this argument by positing a distinction between dispositional knowledge and occurrent knowledge (1988: 263). S dispositionally knows p just in case S would occurrently know p were S to consider p. Here, the process of consideration would bring evidence occurrently to mind (and thus literally bring the evidence back again into S’s possession according to OT). One problem Arnold notices with this response to the sleeping case, is that it radically over ascribes dispositional knowledge to agents (2011: 98). There seems to be an important difference between one’s dispositional knowledge (knowledge
that one has stored \textit{dispositionally}) and one's dispositions to know things (knowledge one doesn't currently \textit{have} in any important sense, but instead has a disposition to have once one considers the claim for the first time). For instance, most of the time people have dispositional knowledge of their names, of the towns they grew up in, etc. But most people don't know—occurrenctly or dispositionally—things like \textit{no elephants are even numbers}, \textit{Alpha Centauri isn't made of cheese}, \textit{the number of blades of grass in Nebraska is either even or odd}, etc. This is because most people have never even entertained these rather obvious propositions. They have a disposition to know them, but they don't dispositionally know them already. Feldman's early response to the sleeping case entails that we do in fact dispositionally know all of these obscure truths, and this is rather counter-intuitive.\footnote{See Audi (1994) for helpful discussion on the distinction between dispositional beliefs and dispositions to believe. It is this distinction here that ultimately grounds the distinction between dispositional knowledge and dispositions to know.}

There's another sort of case that significantly counts against OT. It is the case of rather complex and lengthy inference. Consider cases of calculating a very long math problem in one's head—a problem that involves many discrete steps. Also, consider cases in which one makes a rather complex inference to the best explanation, either in scientific practice, or in everyday life. For these cases of belief formation, the body of evidence on which we base our conclusions tends to be quite expansive. It is very unlikely that, at the moment of forming the belief in the conclusion of these inferences, humans are occurrently representing all of the evidential claims on which they're basing the conclusion. Human ability to occurrently represent (at a single moment) many complicated claims is quite limited. Supposing that basing and forming a belief in the conclusion takes place at time $t$, it is much more plausible that in cases of lengthy inference, agents call the premise/evidence claims to mind throughout the moments \textit{leading up to} time $t$. But if this is the correct account of how evidential basing
occurs for complex inference, and if complex inferences can confer justification, then OT is false.

Lastly, Arnold raises a case having to do with defeating evidence. Intuitively, our evidence can both epistemize and defeat the justification/warrant for beliefs that we have. To use a familiar example, suppose one possesses, dispositionally, a lot of great evidence that supports the claim that some widgets aren’t actually red, but are actually white and being irradiated by tricky red light. In addition, suppose S understands and believes—dispositionally—that this is in fact what her stored evidence supports. But in a moment of quick forgetfulness or carelessness, S comes to believe, on the basis of visual experience, that the widget in front of her is red. Intuitively, at this moment, S lacks justification for believing this claim. The reasonable explanation for this is that S isn’t responding correctly to all of the evidence in her possession. According to her total evidence, her belief is undermined. In other words, her total evidence gives her a defeater for her belief. Although, if OT is true, the justification of the belief at the moment of belief formation wouldn’t be undermined, because she wouldn’t have the undermining evidence (given that it is not occurrent to her). This, however, seems to be the incorrect response to a case like this.

All this to say, there’s good reason to accept MT, and allow contents possessed dispositionally to serve as one’s evidence. But once one accepts MT, one’s epistemic theory

---

27 Pinning down exactly what Conee and Feldman think nowadays about evidence possession is a bit more difficult. As of 2001, in “Internalism Defended,” they describe mentalism as having the following implication: The justificatory status of a person’s doxastic attitudes strongly supervenes on the person’s occurrent and dispositional mental states, events, and conditions (2).

This suggests that they are seeing dispositional mental states as candidates for being evidence one has for a target proposition. By 2008, in “Evidence,” Conee and Feldman seem to not take a stand on the issue of MT vs. OT. They say that within the broader evidentialist camp, one can distinguish many precise evidentialist theories by their precise views of evidence possession (2008:89). Some of these will hold to OT on one end of the spectrum. On the far other end of the spectrum, some evidentialist theories will hold that mental states buried so deeply within a person—that only the most in-depth psycho-analysis will bring them to consciousness—can still count as possessed evidence. And between these two poles, there are (hypothetically) many precise evidentialist theories corresponding to each degree of difficulty in bringing a stored mental state to occurrent awareness. Conee and Feldman don’t take a stand here on which one of these precise theories is correct. Here, I might
takes on an explanatory burden that is very similar to the reliabilist’s burden to explain relevant process types. Consider what we typically mean when we say that someone has a disposition to $\Phi$. First, we typically don’t mean these dispositions to be unconditional, but rather, to only hold in certain conditions. For instance, when we say that a child is disposed to clean up after herself, we do not take this disposition to hold for every possible circumstance the child could find herself in. For instance, we wouldn’t take our claim that a child has such a disposition to clean up after herself to be falsified if it turned out to be the case that the child wouldn’t clean up after herself if the child were to find herself in excruciating pain or paralysis. Dispositions, as we commonly invoke them, only hold in certain circumstances of type T. Secondly, for the most part, we don’t take the dispositions we invoke in everyday life to be infallible. In other words, saying that the child has a disposition to clean up after herself in normal circumstances T doesn’t commit us to the claim that every time the child finds herself in T, she’ll clean up after herself. That seems far too demanding. Rather, when we invoke dispositions to $\Phi$, we typically mean something more modest and fallible, like,

In situations of type T, S probably $\Phi$s.

It seems as if there are at least two kinds of necessary conditions on dispositionally possessing evidence E for a belief that p. Considering both helps generate a parrying response to the generality problem objection.

Necessarily, S dispositionally has evidence E for her belief that p at t only if

i. S brings E to occurrent awareness in a sufficiently large percentage of possible futures of type T from moment t.

point out that, if there’s not a special burden for the evidentialist to present a theory of evidence possession, then it is a bit ad hoc to say that the reliabilist has a burden to present a theory of type relevance without further explanation.
ii. S occurrently represents E as epistemically relevant to p in a sufficiently large percentage of possible futures of type U from moment t.

Condition (i) serves to accommodate the following intuition: the contents of mental representations that are so far “at the back of our minds” (and perhaps near the point of being absolutely forgotten) aren’t candidates for being part of one’s evidence. For some evidence E to be part of one’s total evidence—that could play some defeating or epistemizing role—E must, in some sense, be within the range of S’s broad perspective and reflective grasp. Condition (ii) serves to accommodate the intuition that S doesn’t possess E as evidence for p if S lacks any sort disposition to think E is relevant at all to whether p is true. It is simply not enough to be disposed to bring E to conscious awareness quite easily. For instance, suppose Tim knows what a fire is, and knows what smoke is, but lacks any disposition to think that smoke has anything, causally, to do with fire. On an occasion where Tim possesses, in his memory, the belief that there is smoke in a neighboring room, it would not be the case that he possesses this content, dispositionally, as evidence for the proposition that there’s a fire in the other room.28

The parrying response to the generality problem objection is generated due to the fact that (i) and (ii) invoke the concepts of cases of type T and cases of type U. What exactly are these? Plausibly, they refer to reference classes of some sort. They probably include cases in which S is considering whether p or considering what could be said in favor or against p, but there’s no reason to think that they’re limited to cases like this. Also, it is clear that cases in which S is having a seizure, or cases in which S is struggling to breathe, aren’t contained in the U class or the T class. Intuitively, S’s ability (or lack thereof) to recall E and see E’s relevance to p in these wild possible cases doesn’t determine whether S possesses E as evidence for p. Other

28 See Arnold (2011: 173-177), who suggests and defends a similar agential dispositional account of evidence possession.
than these general insights regarding \( U \) and \( T \), we don’t have much an account of \( U \) and \( T \). I contend that explicating these reference classes, pertaining to dispositional evidence possession, is importantly similar to explicating the reference classes corresponding to the belief forming process tokens invoked by process reliabilism.

Given than many theories of warrant and justification will invoke dispositionally possessed evidence in some way, we can conclude that many reasonable candidate theories of warrant or justification will satisfy PTB. If this is the case, then we’ve shown that there’s another strong warrant and justification parrying response to the generality problem objection.\(^{29}\)

2.3.4 Competent Basing and a Warrant Parrying Response

One can also see an additional warrant parrying response upon considering the specific sort of basing that’s required for having knowledge. Not just any basing process can generate a warranted belief, as the case below will show. I’ll call the sort of basing that’s necessary for warrant competent basing.

PARTY

Jim and Sandra are returning to their house after a night of being away, and as they show up in the morning, they are perceptually acquainted with the following pieces of information at time \( t_1 \). Call this information set \( A \), which they both come to believe on the basis of their perceptual acquaintance with it.

<table>
<thead>
<tr>
<th>Set A</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Windows are broken.</td>
</tr>
<tr>
<td>A2</td>
<td>Red solo cups are strewn everywhere.</td>
</tr>
<tr>
<td>A3</td>
<td>Passed-out people are everywhere (on couches, the yard, floors).</td>
</tr>
<tr>
<td>A4</td>
<td>Half-eaten pizzas are dispersed throughout the house.</td>
</tr>
</tbody>
</table>

\(^{29}\) Granted, this parrying response won’t apply to some phenomenal conservatives, who might just say that one’s evidence that justifies a given belief just is a seeming that one occurrently has at the moment of belief formation. Regardless, there are far fewer phenomenal conservatives about warrant, so many more will find the evidence possession warrant parrying response compelling.
A5 A sound system has been set up.

A6 It smells like beer everywhere.

They also believe (in memory storage) each claim in the following set of information. Call this information set B.

Set B

B1 We desire for our housemates to stop throwing parties.

B2 Our housemate, Steve, wears nice clothing.

B3 People in this college town like owning pets.

After coming to believe set A at t1, Jim and Sandra both (individually) are prompted to consider the following question at t2: What happened at our house last night? Upon consideration, both Jim and Sandra form the belief, at t3, that

N There was a wild party at our house last night.

Jim and Sandra form N on the basis of set A. However, Jim and Sandra are peculiar cognitive agents, but for different reasons. Their respective peculiarities manifest themselves in the unique ways they each form the belief N on the basis of A.

Jim’s case: Jim based N on set A, but only by luck. Modally, looking at Jim’s dispositions, at the point of time (t2) at which Jim starts considering what happened at his house last night, Jim would base his answer to the question “What happened at my house last night” on Set B in the vast majority of the nearby (modally) possible futures from t2.

Sandra’s case: Sandra bases N on set A, and (unlike Jim) has a stable disposition to base answers to questions like what happened at my house last night on (relevant) sets of data like set A. Although, Sandra adopts N on the basis of Set A only by luck. Modally, looking at Sandra’s dispositions at t2, in most nearby possible futures, Sandra forms all sorts of other beliefs (instead of N) about the relevant question on the basis of Set A (e.g., that there was no wild party at our house; that the cleaning service came to our house; that there was a chess tournament at our house; etc.)

Intuitively, both Jim and Sandra lack warrant for believing N on the basis of set A.
The problematic nature of Jim’s belief formation highlights an important necessary condition on warrant in cases of belief-formation in which the subject bases a belief on other claims she mentally represents.

W1 Necessarily, S has warrant for believing p on the basis of some set of claims represented by S only if it is not the case that, given S’s dispositions with respect to a reference class F of nearby possible futures, it is sufficiently likely that S bases the belief that p on a set of claims that fails to support p.

Given the example, Jim clearly fails to satisfy W1. With respect to the nearby possible futures at t2, he is disposed such that it is sufficiently likely that he’ll base his belief that p on claims that don’t support p. His having this modal feature undermines his warrant for p.

The problematic nature of Sandra’s belief formation reveals another important necessary condition on warrant for cases of belief formation that involve basing on other claims mentally represented by the subject.

W2 Necessarily, S has warrant for believing p on the basis of some set of claims represented by S only if, given S’s dispositions with respect to a reference class G of nearby possible futures, it is sufficiently likely that S correctly represents whether the basis set of claims supports p.

Once again, given the details of the case, Sandra fails to have the disposition described in W2.

It is reasonable to view W1 and W2 as necessary conditions on competent basing—the sort of basing that is required for warrant in cases where agents base their belief on contents that they represent either doxastically or with other representational states. Insofar as any plausible candidate theory of warrant must invoke W1 and W2, it must also invoke reference classes F and G. At the outset, we have no reason to believe explicating F and G will be any easier than explicating the reference classes corresponding to the reliabilist’s belief-forming process types. Therefore, there is good reason to believe that many theories of warrant will satisfy PTB.

In response, one might try to deny that an account of competent basing will invoke reference classes F and G. The envisioned response denies that there’s actually anything
dispositionally important for warrant in cases where the target belief p is based on some set H of claims mentally represented by the agent. Rather than invoking W1 and W2 in our theory of competent basing, we can instead invoke other mental states to play a comparable role. This alternative account holds that, in addition to H actually supporting p and S basing p on H, competent basing also requires having the following mental state as well.

A1 Necessarily, S has warrant for believing p on the basis of some set of claims H represented by S only if, at the moment of basing, S occurrently has a seeming or occurrent belief with the content H supports p, and lacks seemings or occurrent beliefs of the form H doesn’t support p.

Let’s call this response strategy the awareness response.

I contend that many won’t find the awareness response persuasive as an alternative account of competent basing. The awareness response entails that Jim and Sandra could both have had warrant in the case above even if everything in the case stayed the same except for adding that, at t3, they each happened to have the seeming A supports N. But this is highly counter-intuitive, given the problematic cases of belief formation that occur in the majority of nearby possible belief-forming events. Suppose this seeming comes to them on a complete fluke of an event—something like a random shot of gamma radiation hits their brains at just the right time in just the right place, generating the seeming at t3.

Granted, phenomenal conservatism and perhaps other epistemic theories allow seemings generated like this, no matter how unlikely it was that they arose, to ground an agent’s possessing justification at the moment of belief formation. They would read a case like this as one of an agent’s acquiring an important piece of evidence just in the nick of time, and then this evidence serves to justify the target belief. This is why I am not framing the upshot of PARTY as a powerful justification parrying response. But, I doubt that there are many internalists or externalists about warrant who would view the awareness response as salvaging
warrant for people with modal profiles like Jim and Sandra in the moments leading up to belief formation. Intuitively, there is a sort of *stable competence* that warrant requires—whether it is a perceptual competence, an intuition based competence, or a competence in grasping one’s evidence and what that evidence supports. But this stability seems to be dispositional in nature, where outcomes in nearby possible worlds play a crucial grounding role in whether the dispositional conditions are satisfied. All this to say, the nature of competent basing still seems to generate a strong warrant parrying response to the generality problem objection.

2.4 Conclusion

In this chapter, I’ve shown that there are strong justification and warrant parrying responses to the generality problem objection. This being the case, many who would present the generality problem objection as a defeater to process reliabilism do so at the peril of their own epistemic views. In addition, we’re left with good reason to think that PTB is a principle that proves too much, and should be abandoned. But if this is the case, then the most plausible versions of the generality problem objection are undermined insofar as they rely on PTB. At this point, the defender of the generality problem is faced with three options. First, she could formulate a compelling version of the generality problem objection that doesn’t rely on PTB. Secondly, she could attempt to formulate reasonable and explanatorily powerful theories of warrant and justification that somehow avoid the parrying responses I defend in this chapter. Thirdly, she could give up on the generality problem objection. I contend that options one and two are rather daunting, and that many will find option three the most attractive.

---

30 Many might view the latter as a particular sort of intuition competence.
CHAPTER 3

AGAINST SUBJECT-BASED THEORIES OF RELEVANCE

3.1 Introduction

Process reliabilism is an important contemporary externalist approach to analyzing epistemic concepts.¹ Both a process reliabilist theory of warrant and a process reliabilist theory of justification have been presented and defended in the literature.² Process reliabilism about justification and process reliabilism about warrant share this common feature: they both claim that the key ingredient that grounds whether an agent’s belief that p possess the relevant epistemic property is whether the agent used a reliable belief-forming process to arrive at the belief.³ In what follows, for simplicity’s sake, I’ll only write in terms of reliabilism about warrant (henceforth, PR), but it’s reasonable to take at least some of the lessons of this chapter to apply to reliabilism about justification as well.

¹ Process reliabilism is an externalist approach in the sense that it invokes features of the world external to the agent’s mind and outside of the agent’s reflective or conscious access to ground the presence or absence of these epistemic properties for a given belief.

² By warrant, I use this term in the functional sense much like Plantinga does, where warrant is the state the possession of which grounds one’s having knowledge that p so long as she also possesses belief that p and p is true. (Plantinga 1993: v)

³ By “key” ingredient, I take this broad construal here to be consistent with the reliable process condition being the only ingredient that generates the relevant epistemic property. I use the term “key ingredient” here so as to include reliabilist theories that hold that process reliability is a crucial part of warrant (or justification), but that there’s some interesting no-defeater condition—that’s not analyzed in terms of reliability—that must also be satisfied in order for an agent to have warrant (or justification). For an account of warrant in which reliability figures as the central grounding feature, see Goldman’s account (1986:44-5), on which warrant requires that an agent’s belief-forming process have both “local” and “global” reliability.
In this chapter, I’d like to focus on the specific nature of what in particular gets evaluated for reliability according to PR. In a way, there’s an important puzzle for the reliabilist to sort out. One might think that the object that gets evaluated for reliability, according to PR, is rather straightforward: it’s simply the process used by the subject. But things are more complicated here. Processes can be thought of as general repeatable types, or as precise one-off tokens.\(^4\) Plausibly, only the former can be evaluated for reliability or unreliability.\(^5\) As both defenders and critics of PR have noticed, a given process token exemplifies many types. A case of seeing a red ball at close distance, and then coming to believe there’s a red ball there on the basis of that visual experience, exemplifies all of the following types: [belief formation], [perception], [vision], [vision in good light], [vision in good light at a close distance], [vision in good light at a close distance with a medium sized object], etc. Reasonably, not all of these

---

\(^4\) Here, I’ll flag one important feature of the relevant literature: there’s been no effort to give a precise account of what a belief-forming process token is. Everyone has more or less adopted the same sorts of conventions in using this term, and allowing this notion to stay at the rough intuitive level. Perhaps this is for good reason. A belief-forming process token seems to be a certain sort of event. Alas, the metaphysics literature on the analysis and individuation of events is voluminous and vexed. This much seems clear from the way the generality problem literature proceeds: a belief-forming process token is an event that has a maximally precise description. So, it’s individuated by the proposition believed, the agent who comes to believe, the time and place at which the belief is formed (where the place description is maximally precise). The state of the environment leading up to the moment of belief formation also seems to be part of the belief-forming token description, and this gets a maximally precise description as well. This appears to be the rough idea of how this literature has used the notion of a belief-forming process token. But there are some more unanswered specific questions one might ask. For instance, does the environment description component of the token include all of the history leading up to the moment of belief formation, or just part of it?\(^5\) It seems odd to say that part of the belief-forming process token description is a causal world history story leading back to the big bang. But if it doesn’t go back all the way, one might wonder what grounds the token environment description’s stopping at some other point in the past relative to the belief-forming moment. Saying more here would surely lead right back to the metaphysical puzzles surrounding event individuation in general.

---

\(^5\) Richard Feldman (1985) and Goldman (1979) are key figures who highlighted this important distinction for making sense of reliabilism. Most philosophers agree that only types are the sorts of things that can be measured for reliability. Although, recently Juan Comesaña (2006) articulates a way in which tokens could be evaluated for reliability with respect to some space of possible worlds as a reference class. Even if this way of conceiving measuring a token for reliability makes sense, Comesaña, correctly, recognizes that framing reliabilism like this doesn’t get the reliabilist out of the generality problem. The reliabilist still would have to provide an account of which possible worlds were contained within the reference class used to evaluate the token’s reliability. As I argue in 3.2, providing an account of the reference class suffices for answering the generality problem as normally construed (as a problem of determining a relevant type).
types are measured for reliability in determining whether the target belief has warrant. Indeed, perhaps only one of them is “relevant” for determining warrant. But which type is relevant? In other words, which features of the token are “held fixed” in the relevant type description? Here, it’s important to point out that this very question constitutes the basis of the famed \textit{generality problem} for reliabilism. According to the literature, this generality problem question turns into the generality problem objection to PR upon realizing that humans—including the best philosophers—do (and have done) a poor job answering the generality problem question. It’s the nature of our lack of understanding into the notion of type relevance that constitutes the key premise in the generality problem objection.

Here, it’s important to point out that very little work has been done to formulate exactly just how our lack of ability to answer the generality problem question generates a defeater for reliabilism. It’s beyond the scope of this chapter to attempt to formulate an argument that represents the strongest version of the generality problem objection. However, responding to the generality problem has motivated many attempts to formulate a \textit{relevance theory} over the

\footnote{For example, presumably, [belief formation] isn’t the type that gets evaluated for reliability. Consider someone who, sadly, is a BIV, such that their \textit{perceptual} beliefs are all systematically false. Such a person, presumably, can still have reliable and warranted \textit{intuition} beliefs in a priori claims. But, if the relevant type for these intuition belief tokens was just [belief formation], then the reliability measurement would be significantly decreased from all of the false perceptual beliefs formed by the agent (since perceptual belief formation is contained within the category of [belief formation]).}

\footnote{The specific language of \textit{relevant} and \textit{irrelevant} types was introduced to the literature by Feldman (1985: 160). This way of talking about the type that gets measured for reliability has become standard throughout the generality problem literature, and I’ll continue that usage here.}

\footnote{Earl Conee and Richard Feldman are the arch defenders of the generality problem objection (Feldman (1985); Conee and Feldman (1998) and (2002)), and it turns out that others are inclined to agree with them. See Plantinga (1993: 28-9) and Matheson (2015).}

\footnote{Indeed, in chapter 1, I argue that successfully formulating an argument against PR from a premise involving our current lack of understanding of type relevance doesn’t look promising. It seems like any plausible approach to formulating this argument either requires one to invoke principles that would generate a defeater for \textit{any} theory (and hence we should reject these principles), or to invoke premises that we have little to no reason to believe.}
A relevance theory presents conditions or principles that determine which features of tokens are held fixed in their corresponding relevant type descriptions. It’s also the case that over the course of the past three decades, critics of PR have pointed out crippling difficulties with many of the extant relevance theory proposals.\(^{10}\)

In this chapter, I’ll examine a particular kind of relevance theory that, while present in the earlier discussions of the generality problem, has re-emerged in the contemporary literature: what I call subject-based relevance theories. What unifies all subject-based relevance theories is a commitment to the view that elements of the subject’s perspective determine which type exemplified by her token is relevant for measuring reliability. Interestingly enough, subject-based relevance theories for the most part haven’t been met with much of any sort of critical response like other kinds of relevance theories have. Offering a critical response to subject-based relevance theories is the goal of this chapter.

In the literature, one can see two different kinds of subject-based theories put forward: belief-relevance theories, and practical interest-relevance theories. Belief-relevance theories claim, roughly, that features of an agent’s beliefs ground which type is relevant. Practical interest-relevance theories claim that the practical and cognitive goals of an agent determine which type is relevant. In §3.2, I present and explain belief-relevance theories. In §3.3, I offer some objections to belief-relevance theories. In §3.4, I present and discuss practical interest-relevance theories. In §3.5, I highlight some key problems with practical interest-relevance theories.


\(^{11}\) For criticisms of Heller and Alston, see Conee and Feldman (1998). For a criticism of Adler and Levin, see Conee and Feldman (2002). For a criticism of Becker, see Brueckner and Buford (2013). For a criticism of Comesaña, see Matheson (2015). For a criticism of Bishop, see Conee (2013). For a criticism of Beebe, see Dutant and Olsson (2013).
theories, arguing that we have no good reason to accept them. In §3.6, I take stock of the findings in earlier sections, ultimately concluding that subject-based theories are a dead end, and that the relevance theorist ought to take a different approach to answering the generality problem question.

3.2 Belief-relevance Theories

Broadly speaking, belief-relevance theories claim that an agent’s beliefs about which features constitute her epistemically relevant type determine, in some way, what the features of her relevant type really are. There are two precise (and distinct) belief-relevance theories I’ll concern myself with below.

Sufficiency Belief Determination (SBD)
Necessarily, for all possible tokens t, if the subject believes G is a property constitutive of the relevant type for her token t, then G is a property constitutive of her relevant type for token t.

Necessity Belief Determination (NBD)
Necessarily, for all possible tokens t, if G is a property constitutive of the relevant type for a given token t, then the subject believes that G is a property constitutive of the relevant type for token t.

Of course, there are other specific belief-relevance principles one could posit, but I take these two here to be the most important and instructive principles for evaluating whether our beliefs about type relevant play any sort of important role in actually determining type relevance.\textsuperscript{12}

I now turn to some putative examples of SBD or NBD appearing in the generality problem and relevance theory literature. I’ll begin with Ernest Sosa. In \textit{Knowledge in Perspective}, Sosa writes,

\footnote{\textsuperscript{12} For example, SBD and NBD hold that the agent’s beliefs about type relevance fix the nature of a token’s relevant type in a 1-to-1 fashion. For instance, according to SBD, if S believes feature G is constitutive of t’s relevant type, then G is a constituent of t’s relevant type. It is conceptually possible to formulate a belief-relevance principle in which the contents of an agent’s beliefs wouldn’t fix the nature of the relevant type in this specific 1-to-1 fashion. However, I take my discussion of SBD and NBD to ultimately cast doubt on whether such a modified belief-relevance principle could be extensionally correct as well.}
Subject S believes proposition P at time t out of intellectual virtue only if there is a field of Propositions F, and there are conditions C, such that: (a) P is in F; (b) S is in C with respect to P; and (c) S would most likely be right if S believed a proposition X in field F when in conditions C with respect to X. (1991: 138)

For clarity’s sake, I should point out that Sosa doesn’t explicitly formulate his theory of warrant in terms of processes generating reliably-formed beliefs. Instead, Sosa writes in terms of reliable faculties, or, intellectual virtues, generating warranted beliefs—where a faculty is a belief-forming capacity for contents of certain field F that an agent possesses when she finds herself in circumstances C (where C can describe both the internal constitution and external environment of the agent). One’s forming a belief out of intellectual virtue is a crucial necessary condition on warrant for Sosa. However, I think Sosa’s virtue theory can be helpfully translated in terms of processes: For any token process t of a faculty, characterized by F and C, generating a belief that p for S, t’s relevant type description holds features F and C fixed (as well as the fact that it was S who formed the judgment). This type must be reliable in order for t to confer warrant on the target belief.

Here’s one place in which Sosa specifically articulates the generality problem for his view:

One has a faculty only if there is a field F and there is a set of circumstances C such that one would distinguish the true from the false in F in C. But of course whenever one happens to have a true belief B, that belief will manifest many such competences, for many field/circumstance pairs F/C will apply. How then can one rule out its turning out that just any true belief of one’s own is automatically justified? To my mind the key is the requirement that the field F and the circumstances C must be accessible within one's epistemic perspective. (1991:274)

---

13 Ibid. (270)

14 This will come as no surprise to many readers, who are well aware that Sosa’s virtue epistemology is now often described under the broader family of views called virtue reliabilism.

15 Sosa, in the text, (1991:138) seems reticent to categorize his view as a “historically reliabilist” view, which is how he views Goldman’s standard process reliabilist framework. This is because he wants his theory to be able to account for the warrant conferred by instantaneous (and hence, “non-historical”) mental processes, like grasping one’s own existence in the cogito. While it’s by no means clear that there even can be instantaneous belief-forming processes, I’ll be using the term “reliabilism” to include possible cases of instantaneous belief formation. So, Sosa’s theory fits within the broader reliabilist camp.
Sosa grants that any particular token belief-forming process exemplifies an infinite number of different field and circumstance descriptions. According to some of these field and circumstance descriptions, the belief is reliably formed, and according to others, the belief is formed unreliably. It’s clear that Sosa just thinks that one F and one C are relevant for each token. We can now ask, what makes it the case that a given F and C are in fact the relevant ones?

In the quote above, I take Sosa to mean that the relevant field and circumstances must be accessible within the agent’s perspective. Sosa’s specific language of one’s perspective being “the key” to determining which F and C are relevant for a given token at least strongly suggests that Sosa, in these passages, is presenting some version of NBD. John Greco agrees with this interpretation of Sosa.16 I must point out that interpreting Sosa here is tricky because in the nearby context, he also presents two other sorts of factors that play some role in determining type relevance: whether a given type description corresponds to a natural kind, and whether a given type description plays some part in serving the agent’s practical needs/goals.17 While NBD is the most natural interpretation of the block quoted passage above, it is possible that Sosa intends for each of the three factors he invokes (in responding to the generality problem) to be individually sufficient for determining the relevant type.18

---


17 See Sosa (1991: 282) for the presentation of the practical interests/goals approach, and (236) for his presentation of the natural kinds approach.

18 One other point that complicates interpreting the block quote from 274 is that, in the immediate context, Sosa hints that he’s concerned about reflective knowledge rather than animal knowledge (275). In Sosa’s framework, reflective knowledge is the sort of knowledge that humans have (at least most of the time) that’s more epistemically valuable and robust than the sort of knowledge animals have. Reflective knowledge, for Sosa, does require an agent to have a second order perspective on her first order belief formation (and its reliability). But even for animal knowledge at the first order, there must in fact be reliability with respect to a relevant F and C exemplified by the token. In the block quotation on (274), Sosa appears to invoke one’s perspective as a determining factor of the relevant F and C at the first order.
Next, while Alvin Goldman no longer subscribes to this view, he did at one time put forward the “normal worlds” theory of epistemic justification. While the view he presents is a theory of justification, it’s still instructive for our inquiry into warrant. One might adopt an analogous position pertaining to the relevant types for warrant. Goldman writes,

We have a large set of common beliefs about the actual world: general beliefs about the sorts of objects, events, and changes that occur in it. We have beliefs about the kinds of things that, realistically, do and can happen. Our beliefs on this score generate what I shall call the set of normal worlds. These are worlds consistent with our general beliefs about the actual world.... Our concept of justification is constructed against the backdrop of such a set of normal worlds. My proposal is that, according to our ordinary conception of justifiedness, a rule system is right in any world \( W \) just in case it has a sufficiently high truth ratio in normal worlds (1986: 107).

One might wonder what the relationship is between a set of normal worlds and the nature of the relevant type. I think Goldman can best be interpreted as claiming that the normal worlds constitute the reference class of situations across which the belief-forming process is evaluated for reliability. Consider now the notion of a reliability measurement. Intuitively, the degree of reliability for a process depends on an important sort of ratio: the ratio of instances in which some outcome x occurs to instances in which outcome y occurs, taken across some class of possible cases. This “class of possible cases” used for the reliability ratio measurement is what philosophers (and statisticians) call the reference class. With respect to the concept of reliable belief-formation, the pertinent outcome x is true belief formation, outcome y is false belief formation, and the reference class of possible cases is constituted by particular possible belief-forming process token events. So, for each token belief forming process with a reliability measurement, there’s a corresponding reference class over which that reliability measurement is taken.

There’s an important relationship between the reference class of a token, and a token’s relevant type: All the particular events that constitute the reference class satisfy the token’s
relevant type description. It doesn’t make sense for a token’s degree of reliability to depend or partially depend on a truth/falsity ratio measured across a class of cases that don’t even satisfy the relevant type description. So, it follows that any principles that place constraints on the relevant type thereby place constraints on the reference class for the reliability measurement. Ultimately, the motivation behind invoking any relevance principles at all is to place the right sorts of constraints on the reference class, so that the correct reliability measurements are delivered for particular cases. Reasonably, this is a central goal motivating Goldman’s normal worlds theory of reliability.

In Goldman’s description of the reference class, he appeals to the beliefs that “we” have in determining what the normal worlds are. The content of the beliefs “we” have about the typical environments and conditions in which “we” form our beliefs determines the class of normal worlds—and thus, determines the content of the reference class description. He claims that the beliefs we have “generate” the set of normal worlds. There are two important issues to flag here. First, who does “we” refer to in Goldman’s account? For a given token case t and subject S, is the “we” who determines the normal worlds for that t just S herself, or does “we” refer to some larger group of people? If it’s just S, then it will be easy to map Goldman’s normal worlds theory on to either SBD, or NBD. If “we” refers to some group of individuals, such a simple mapping won’t work. Nevertheless, it is clear that Goldman at

---

19 First, offering an account of the reference class suffices to answer the questions that get the generality problem going to begin with. Given the characterization of the generality problem according to Feldman (1985) and Conee and Feldman (1998), the problem for the reliabilist is that, although we can know the features of a given token, we don’t know how to perform the reliability measurement for that token to determine whether that token confers justification or warrant on the target belief. For the token case of seeing and coming to believe in the presence of a red ball (discussed in the introduction), we don’t know whether to perform the reliability measurement with respect to [belief-formation] in general, or with respect to [perceptual belief-formation], or with respect to [visual-belief-formation], etc. But the more precise way to formulate this puzzle is in terms of the reference class corresponding to the token: we don’t know whether the token’s reference class is comprised of particular cases of all sorts of belief formation, or just with particular cases of perceptual belief-formation, or just with cases of visual belief-formation, etc. If we could completely determine the features of a token’s reference class, then this would satisfactorily answer the questions Conee and Feldman raise to get the generality problem going.
least thinks that the beliefs of at least one agent can determine the features of the relevant type for some token t. So, if “we” refers to some group, then analogous principles to SBD and NBD could be crafted to capture Goldman’s view. Secondly, while it’s not absolutely clear, the language of “just in case” and “generates” in Goldman’s presentation of the normal worlds account strongly suggests that the views “we” have about the normal worlds is both necessary and sufficient for determining the features held fixed across all of the reference class.

Whether or not my specific interpretive points regarding Sosa and Goldman are correct, it’s clear enough that views sufficiently close to NBD and SBD have been advanced in the broader reliabilist literature. In what follows, I present what I take to be cases that undermine both NBD and SBD, casting serious doubt on whether one’s beliefs about type relevance play any sort of interesting role in determining type relevance.

3.3 Against Belief-relevance Theories.

Let’s begin with SBD. Consider the following case. Suppose James is in fake barn county and non-inferentially forms the belief that p1: there’s a barn. Suppose James also has the background belief: The relevant type for my barn-existence judgment is partially constituted by my being in an environment with lots of barns and lots of farming. The latter belief is false. In fact, he’s in an area with few barns, little farming, but many individuals who want to trick travelers into thinking

20 Also, Jarrett Lepplin takes a similar approach to grounding which type is relevant:

In consulting the encyclopedia, I consult a book. The latter is not a reliable method; there are bad books. This is a problem of generality. Any method I use to form beliefs satisfies a description general enough to ruin its reliability. But under the more general description, it need not be the method I use intentionally. Again, the test of intention is what one does under changed conditions: replace my copy of Paul Bocuse with Joy of Cooking, and see if I use it. If I do not trust Joy of Cooking where I would have trusted Paul Bocuse, then my method is not to consult a cookbook as such, but is more specialized. (2007:35)

Lepplin defends the view that the relevant type is the type the agent intends for her process token to exemplify. Commonly, intentions are seen to be states, constituted by both beliefs and desires, which play a central role in explaining action. The doxastic/representational component of intention puts Lepplin’s strategy in the belief-relevance theory category.
there’s lots of farming. James’ p1 belief is unwarranted, and we take it to be unwarranted because the belief-forming process he uses is unreliable. That is, his token’s relevant type is unreliable. Intuitively, the correct relevant type description would include details of the staggeringly high ratio of fake barns to real barns in the nearby area. This fact, together with other features of the relevant type, make it the case that the relevant type is unreliable. But, if the content of James’ background beliefs about his environment were sufficient for determining the features of the relevant process type, then a staggeringly high ratio of fake barns to real barns wouldn’t be held fixed in the relevant type—thus making the relevant type reliable after all. But this is, intuitively, the wrong verdict on this case.

Engaging with Goldman’s normal worlds view, suppose we interpret the “we” to refer to the community of rational agents in S’s surroundings. Applying the case of James to this formulation of Goldman’s theory, I think there’s good reason to deny that the views of this kind of group could suffice for determining the features of a token’s relevant type. Consider a small farming community made up of rational agents who all greatly value the practice of visually recognizing barns in their environment. It turns out that they all, on the basis of really good evidence, believe that the ratio of real barns to fake barns in the country around them is 500:1. In fact, they’ve been terribly misled by their great evidence, and the ratio is actually 1:37. Suppose it’s also the case that mostly everyone in this community believes the following claim: In this community, the barn-existence judgments we make have relevant types partially constituted by our being in an environment with a real-barn to fake barn ratio of 500:1. Suppose, by sheer luck, in the history of the town, the agents often just happen to be looking at the one or two real barns around them.

21 Here, I’ll leave to the side the concern that such an account of normal worlds would entail that subjects in worlds where they’re the only person either wouldn’t be able to know anything (in virtue of there being no community to fix the normal worlds) or would determine the features of the normal worlds all on their own (given that we’d view their community to include just one member: just themselves).
when they make their barn recognition judgments. Even though these beliefs would be true, and even though the agents believe on good evidence that their environment has warrant/reliability-conducive features, they still don’t have warrant for their barn identification beliefs. Their beliefs about their environment and their relevant types for barn-existence judgments, aren’t sufficient for fixing the features of the normal worlds which constitute the actual reference class for their barn identification judgments.

James’ case above was a case of the subject having a false belief about the type in virtue of having a false belief about the token. But one might try to salvage a sufficient belief-relevance theory by simply requiring that the agents, who determine the relevant type with their beliefs, only have true beliefs about the token. Consider the following modified version of SBD.

SBD* Necessarily, for all possible tokens t, if the subject believes G is a property constitutive of the relevant type for her token t, and G actually is a property constitutive of the token t, then G is a property constitutive of t’s relevant type.

However, counter-examples to SBD* readily suggest themselves as well. Consider this case from Michael Bishop. Note, “BFPT” in the following quote stands for relevant belief-forming process type.

After years of neglect, Paula spends much of the day, from noon to 4:55, balancing her checkbook, performing a few thousand simple arithmetic calculations. She spends the next 5 minutes coming to various absurdly optimistic beliefs about her future. She tries to justify these beliefs by appealing to a BFPT that she knows is reliable. She argues: “My beliefs about my future will be produced by process tokens that are instances of a highly reliable BFPT. That BFPT consists of the process tokens that share the property of having produced my beliefs between noon and five…..” Of course, despite Paula’s appeal to an extremely reliable BFPT, her absurdly optimistic beliefs are not justified. (2010: 289)

Here, Paula correctly identifies a feature of her process token—namely, that it’s a token of hers that occurred between noon and five that day. Nevertheless, this feature description doesn’t constitute the relevant type description. The fact that she believes that it does
constitute the relevant type *doesn’t make it the case* that it in fact constitutes the relevant type. Likewise, we could alter the fake barn case with James and stipulate that he happens to hold the following (albeit obscure) background belief about his relevant type as he forms his (luckily true) barn identification judgment: *For measuring the reliability of my barn-identification judgment, it’s relevant that my belief-forming process involves optical nerve stimulation from light waves that bounced off of a real barn.* While James’ belief correctly identifies a feature of his token, it’s clear that this feature is *not* held fixed in the relevant type. Holding this feature fixed would generate a reliable type, but intuitively, this is the wrong result for James’ case. These fake barn cases, as well as the Paula case, provide good reason to reject both SBD and SBD*.

There’s also good reason to reject NBD. The main argument I present here is from John Greco. He calls it the psychological plausibility objection. I present his argument below, as well as some further developments of my own.

As I mention above, Greco interprets Sosa as defending NBD.22 In Greco’s argument against NBD, he uses the term “second order beliefs” to refer to beliefs an agent might have about her relevant type used in coming to believe p. Greco writes,

> The Psychological Plausibility Objection now amounts to this: It seems implausible that such second-order beliefs exist in typical cases of human knowledge, even subconsciously, and even in ways that are not verbalizable by the knower. (2004:101)

I think Greco is correct. It’s quite reasonable to think that adults and even children know many things. But by and large, adults and children don’t possess descriptive beliefs about what their relevant belief-forming process types are.23 Small children probably have no informative

---

22 Greco (2004:98-99)

23 Charitably, I take Greco to be saying that adults and children don’t possess beliefs that are either *explicitly or implicitly* about what their relevant belief-forming types are. Trivially, most people don’t have *explicit* beliefs about relevant types, because most people (except for philosophers) don’t explicitly have the concept of *relevant belief-forming process type.* Most people only possess this concept implicitly.
views about how they form their beliefs at the time at which they’re formed. The same goes for most adults. But according to NBD, the subject must have beliefs about the nature of her relevant type in order to have a relevant type. One can only have a reliably formed belief if one’s token possesses a relevant type. Hence, according to NBD, it follows that most children and adults lack reliability (and thus, warrant) for at least most of their beliefs. This is, intuitively, the wrong result.

Granted, there are many cases in which adults might form a belief at time t1, and then later at time t2, reflecting on their past, form some beliefs about how their belief-forming process worked at t1. But surely, these adults could have warrant for their beliefs formed at earlier times before they form their reflective beliefs at a later time.24

One might put their foot down, claiming that if we humans (even children) introspected enough, we’d see that we do have some beliefs about our relevant belief-forming process types when we form beliefs. I don’t think this is true, but even if it were, it wouldn’t demonstrate the truth of NBD. According to NBD, for every feature G constitutive of the relevant type, we must believe, of G, that it is relevant. In other words, NBD entails that we must believe the complete description of the relevant type in order for that type to in fact be relevant. It’s extremely implausible that our beliefs (if we have them deep down) about type relevance are this detailed. This point also undermines one potential attempt to escape Greco’s objection: A belief-relevance theorist might claim that, at the moment of belief formation, all an agent needs in order to have a token with a relevant type is a disposition to form true beliefs regarding

---

24 See Greco (2004:101). Greco also cites psychological research on blindsighters. These are individuals who reliably receive information from their external environment. However, there’s no phenomenal character to the information coming in, so these individuals have not even the beginnings of an idea as to where this information is coming from, and hence no idea regarding the causal details constitutive of the relevant belief-forming process types they use when they reliably form their blindsight beliefs. Blindsighters present a problem for Sosa’s perspectival view, given that it’s reasonable to think that blindsighters can know propositions believed as a result of their blindsight faculties.
the relevant type. Indeed, perhaps this is just what Sosa meant all along, given his use of the term “accessible.” In response, it’s doubtful that every case of knowledge is accompanied by such dispositions to have true beliefs about one’s belief forming process type. But even if such dispositions are widespread, it’s doubtful that these dispositions to believe have contents that are anywhere close to detailed enough to specify all of the features of the relevant type.\footnote{Indeed, there would be no generality problem if us humans had dispositions to believe the complete type descriptions for all of our tokens.}

One might grant this point, but then retreat to a less ambitious necessitating belief relevance principle:

\[\text{NBD*} \quad \text{Necessarily, for all possible tokens } t, \text{ if type } T \text{ is the relevant type for } t, \text{ then the subject } S \text{ must believe that some subclass } X - \text{of all the features constituting } T - \text{are held fixed in the relevant type for token } t.\]

As I mention above, there’s good reason to think that humans don’t have to have any beliefs about their relevant types in order to simply have reliable belief formation. But if one rejects this idea and goes in for some intermediate theory like NBD*, the question that immediately arises is, what sorts of properties make it in to the subclass X? Are there any principles that determine how much of the total relevant type description one must know? For example, consider a fake barn county case. How might NBD* be applied here? Perhaps it’s the case that the agent must believe she’s using a visual belief-forming process, but not the case that she needs to have any beliefs about the ratio of real barns to fake barns in the area. The latter property is constitutive of the relevant type, but not in subclass X. Of course, I’m just speculating here as to the workings of NBD* as it pertains to this example.

NBD* without further development seems rather ad hoc. If not all of the features of relevant type T must be identified by the agent, why must subclass X be identified in order for T to be the relevant type? At best, NBD* is left with the following problem: it’s
underdeveloped to the point where we’re not in a position to evaluate whether NBD* is
extensionally correct with respect to our pre-theoretical warrant intuitions. I doubt much
further development or discussion on NBD* will occur, because most readers will probably
be sympathetic to Greco’s original line of attack, pointing to the plausibility of many cases of
knowledge for which agent’s lack any second order beliefs about their process type at all. As
things stand, I find any *necessitating* belief-relevance theory to be unmotivated.

In sum, the argument against NBD shows that beliefs about type relevance aren’t even
present in most cases of knowledge to play any sort of grounding or explanatory role. Next,
the arguments against SBD showed that there’s no *necessary* grounding connection (applying to
all cases) between the content of one’s beliefs about type relevance and the actual features of
the relevant type. These arguments seem to suggest (while not entailing) that one’s beliefs
about type relevance *have nothing to do with* determining the actual features of relevant types.
Hence, belief-relevance theories in general seem to be the wrong approach for making
theoretical progress on type relevance.\(^{26}\)

\(^{26}\) Of course, there is conceptual space left open for a belief-relevance theory that’s not directly covered by
the SBD and NBD. For instance, both of those theses are ones that apply “to all possible tokens t.” Perhaps
even if SBD and NBD are false, there are other versions of these principles that do correctly apply to just some
\textit{subclass} of all possible tokens. I certainly haven’t definitively ruled this out by any of my arguments above.
However, I do think that the above discussion serves to illustrate that whatever subset of possible tokens the
revised SBD and NBD principles apply to won’t be sufficiently representative of most everyday instances of
warrant-conferring belief formation. Regarding any sort of necessitating belief relevance principle, we saw that
we lack beliefs (occurent or dispositional) regarding most or all of the features that constitute the relevant types
of our tokens. Regarding any sort of sufficing belief-relevance principle, the cases discussed above highlight how
our beliefs about type relevance don’t end up determining anything about the features of the relevant type (in
virtue of their content about type relevance) for fairly common cases of belief formation. At this point, we have
good reason to doubt that any revised version of SBD or NBD that applied to a restricted class of tokens could
be developed that correctly describes the nature of type relevance for most cases of knowledge and belief
formation.
Practical Interest-relevance Theories

Practical interest-relevance theories claim that the *practical interests* of agents determine the relevant type for a belief-forming process token. As I mention above, there are places in which Ernest Sosa appears to advocate this sort subject relevance theory.

Concerning the generality problem, now, what should one say about the field $F$ and the conditions $C$ relative to which one can attain pertinent animal justification, through one's tendency to be right in $F$ and $C$? What restrictions may appropriately be placed on $F$ and $C$?... The likely solution is to be sought, as I see it, in the requirements that $F$ and $C$ must fulfill if (i) $F$ and $C$ are to be *usefully generalized upon by us* as the epistemic community of the subject $S$ (assuming he is one of us - and, if not, then knowledge attributions may need to be indexed or relativized to such communities); and if (ii) $F$ and $C$ are to be *usefully generalized upon by the subject himself* as he bootstraps up from animal to reflective knowledge (1991: 283-4 emphasis mine).

First, we need to get clear on what Sosa means here by “useful.” Useful for what? Sosa gives us a clue in a preceding passage.

We care about justification because it tends to indicate a state of the subject that is important and of interest to his community, a state of great interest and importance to an information-sharing social species. What sort of state? Presumably, *the state of being a dependable source of information over a certain field in certain circumstances*. In order for this information to be obtainable and to be of later use, however, the sort of field $F$ and the sort of circumstances $C$ must be projectible, and must have some minimal objective likelihood of being repeated in the careers of normal members of the epistemic community. For it is through our cognizance of such relevant $F$ and $C$ that we grasp the relevant faculties whose possession by us and others makes us dependable informants and cognizers. What is more, it is precisely by grasping how one does oneself have such animal aptitude over a certain field $F$ in certain circumstances $C$ that one bootstraps up to a higher level of reflective justification. (281-2 emphasis mine)

Here, Sosa lays out two important interests we have as agents: finding and identifying dependable informants in a particular domain, and ourselves being dependable belief-formers in a given domain. Given these interests, the property of justification is important to us because this property marks the beliefs of those who form their judgments in a way that a dependable source of information would. As I mention above, Sosa’s comments on justification could be just as applicable to the notion of warrant, and I proceed under this
assumption. Given these specific texts, the central fact Sosa invokes for determining the relevant field $F$ and conditions $C$ for a given subject $S$’s token belief-forming process (for coming to believe $p$) is the following:

\begin{itemize}
  \item Dependability ($D$) Selecting $F$ and $C$ as relevant successfully allows us to sort and determine whether people who form a belief that $p$, like $S$ did in her token, are dependable on matter $p$.
\end{itemize}

John Greco advances a similar interest-based view. Going back to $D$, one might reasonably ask what sort of dependability is at play. After all, the specific nature of the $F$ and $C$ that’s held fixed in the relevant type is determined by the pertinent sense of dependability. Judgments of dependability are relative to specific standards and contextual conditions. But what are the standards and contextual conditions implicit in $D$? Greco goes a bit further than Sosa, and specifically claims that it’s the practical interests at play in a given case that determine the specific sort of dependability that’s relevant for evaluating the reliability of a given case of belief formation.

[I]t is the relevant practical reasoning context, which may be that of the subject, the attributor, or some third party. Specifically, the relevant parameters [of the type] are set by the interests and purposes that are operative in the relevant practical reasoning context. (2010: 79 emphasis mine)

Greco’s view is shown to be quite similar to Sosa’s as he explains the implications of his strategy for two examples. In the first case (case 1), imagine a government employee, driving through fake barn county, on the job counting barns in an area to determine property taxes. This fellow has no idea that he’s in fake barn county. When he spots the one actual barn in the distance and counts it on his spreadsheet in response to his barn-representing visual experience, we don’t have the intuition that the employee knows there’s a barn there. In case 2, imagine a farmer who owns a farm in fake barn county. Unlike the government employee, this farmer knows that he owns the only real barn in the area. When this farmer is leading his
cattle back to the shelter and has, on his own farmland, a barn-ish visual experience and forms the belief that’s a barn, we do want to say that he has knowledge. Greco says his strategy explains our differing reactions to these cases as follows:

Relative to the purposes at play in the first practical reasoning context, [the employee] is not a good source of information about there being a barn yonder, and information from him to that effect would not be actionable. But relative to the purposes at play in the second practical reasoning context, [the farmer] is a good source of information about there being a barn yonder. (80 emphasis mine)

For Greco, like Sosa, the central property that determines the relevant type is that of being a good, or, dependable source of information. The practical interests at play in a given case serve to determine what “being a dependable source of information” amounts to for that case, and hence, determine the specific features of the relevant type for a given case. In the employee case, the practical interests at play ground the relevance of a type for this token that ends up yielding a low reliability measurement. In the farmer case, the practical interests at play ground the relevance of a type that yields a high reliability measurement. In this way, the goals or practical interests at play in a given case determine which type is relevant. 27

Before I offer my objection to practical interest-relevance theories, it’s important to distinguish two different sorts of practical interest-relevance theories. One way of dividing the conceptual terrain is in terms of whose practical interests and goals matter in the given case: the belief-forming subject’s practical interests, or the knowledge attributer’s practical interests. Given this distinction, we can formulate the following two practical interest-relevance theories.

27 Frederick Schmitt, while not attempting to present necessary and sufficient conditions on type relevance, does present “constraints” on type relevance that he takes as describing “the pragmatic conditions of epistemic evaluation” (1992:142) These constraints include pragmatic factors like the salience of certain features to a given context, as well as maximizing utility vis-a-vis the end of adopting ways of evaluating belief that lead us form more true beliefs in the long run (1992: 143,157) Hence, Schmitt seems to endorse a sort of practical interest relevance theory.
SS  Necessarily, for all subjects and possible tokens, the practical interests of the
subject S, regarding the nature of a kind of dependability S may or may not have,
suffices to determine the relevant type T for S’s belief-forming process token.

AS  Necessarily, for all subjects, attributers, and possible tokens, the practical
interests of the warrant attributer (for a given token of the subject S), regarding
the nature of a kind of dependability S may or may not have, suffices to
determine the relevant type T that determines the truth or falsity of the
attributor’s warrant attribution to S’s belief-forming process token.\(^{28}\)

To get clear on SS and AS, one’s practical interests regarding the nature of some sort of
dependability suffices to determine a relevant type for a given token in the following way:

With one’s practical interests, one values judgments formed in ways that give the subject some
kind K of dependability with respect to the content judged. K here might be characterized by
a sort of content F as well as a sort of belief-forming situation C a subject might find herself
in. According to SS, the subject’s valuing beliefs that manifest dependability with respect to
contents F in circumstances C suffice for it’s being the case S’s token (that exemplifies F and
C) has a relevant type constituted by F and C. For AS, the attributer’s valuing of judgments
that manifest dependability with respect to F and C makes it the case that the token being
considered must be reliable with respect to a reference class constituted by F and C in order
for the attributer’s reliability attribution to be true.\(^{29}\)

\(^{28}\) Here, AS isn’t so much a condition on having warrant, but rather articulating a truth condition on warrant
attributions. SS is more clearly a subject-based relevance principle, but investigating AS is instructive for two
reasons: First, it makes contact with one of the suggestions made by Greco in the quoted passage above (79).
Second, it encapsulates one interesting way in which the concept of knowledge might be connected to practical
interests.

\(^{29}\) Given Sosa’s other discussions on type relevance, particularly his apparent support of some sort of a
belief-relevance theory, it’s not completely clear whether Sosa would accept something quite as strong as SS and
AS, given that these are sufficing conditions. One might interpret Sosa’s comments here, in the context of all
the things Sosa says about type relevance, as suggesting that practical interests play some important grounding
role that falls short of sufficing. Regardless, Sosa’s views are important for setting up the standard sort of
dependability based practical interest relevance theory. With Greco, it’s much more clear that he means
something as strong as AS or SS.
3.5 Against Practical Interest-relevance Theories

To begin, there’s good reason to believe that AS and SS are false. I’ll address AS first. As I’ll try to make clear, it’s certainly possible that practical interests (of some group or individual) could fix a sense of “dependable” to correspond to a relevant process type that is intuitively far-off from a process type that corresponds to the actual reliability measurements warrant. Here’s an example. Suppose that there’s a tour group, who happens to be walking through fake barn county. Also, for the short time interval of a few seconds \( t_1 \)–\( t_3 \), the tour group happens to all be facing in the general direction of the one real barn in fake barn county. Even though there are facades all around, their eyes just happen to be pointed toward the one real barn in this short period of time. It seems possible for Stephanie, a perceptual psychology graduate student, to be given the—albeit odd—assignment to evaluate the barn-recognition dependability of the tour group relative to just the time interval \( t_1 \)–\( t_3 \). Suppose Stephanie comes to value judgments that manifest dependability with respect to barn recognition in the tour-group’s situation just across moments \( t_1 \)–\( t_3 \). More precisely, suppose that this is the only sort of dependability that Stephanie values.\(^{30}\) Here, given the practical interests operative in Stephanie’s research, she is going to find the individuals of the tour group to be dependable barn recognizers with respect to this sense of dependability. The only barn beliefs they’re likely to form in this short time period are correct barn identification beliefs. But now, suppose that, during \( t_1 \)–\( t_3 \), Stephanie says (or thinks to herself) that “The tour members warrantedly form the belief that’s a barn.” Intuitively, this claim of hers is false, even though the subjects she’s thinking about are dependable in the sense that she values. The tour group subjects still exemplify a kind of unreliability with respect to barn-existence judgments that undermines the

\(^{30}\) Suppose that the only thing she cares about is this one little psychological experiment.
truth of Stephanie’s warrant attribution. This example shows that not all interest-dependent
senses of being a good or dependable source of information correspond to the kind of dependability that
marks agents who have warrant.

We can adjust the case slightly to see the implausibility of SS as well. Suppose Barney is
one of the members of the above tour group, and for some very odd idiosyncratic reason that
he’s adopted, it is of dire importance to him that he is dependable at detecting barns only in
the \( t_1 \)–\( t_3 \) time interval. He doesn’t care at all about what happens after that. If he were to find
out that he wasn’t so dependable in that time interval, he would take his own life. Having
beliefs that manifest this odd sort of dependability is the only thing that is important thing to
him. At time \( t_1 \), he looks at a real barn, and on the basis of a barn-ish visual experience, comes
to believe (truly) that there’s a barn in front of him.

It is the case that, given where he’s positioned at \( t_1 \), he would be a near perfectly
dependable barn identifier within the time period of \( t_1 \)–\( t_3 \). But it still doesn’t seem as if Barney
knows, at \( t_1 \), that there’s a barn in front of him. The reference class corresponding to his visual
barn identification beliefs within the \( t_1 \)–\( t_3 \) time window isn’t the same reference class as the one
with respect to which he’d need to be reliable in order to have warrant. Cases like this illustrate
that the subject’s practical interest-based notion of dependability doesn’t suffice to determine
the relevant type and reference class for the actual warrant reliability measurement.

I think these tour-group counter-examples to AS and SS are rather decisive. But much
like SBD and NBD, AS and SS are formulated to apply to all possible tokens. Perhaps there’s
some subclass of tokens for which subject or attributer practical interests do determine features
of the relevant type. I’m inclined to think that the most plausible cases that seem to illustrate
how cognitive and practical goals might impact type relevance are cases in which agents
possess interests that are explicitly connected to their own concept of knowledge. Here, the
sort of case I have in mind is taken from Blake Roeber’s discussion of intellectual interests and knowledge. Roeber’s cases involve agents changing their view on the requirements for warrantedly believing p, while holding fixed both their credence in p and the amount of evidence (both doxastic and experiential) that ground the agent’s belief in p. Suppose that at time t1, an agent forms the belief that p based on evidence E. But then, after taking some epistemology classes, at time t2 S comes to hold that knowing p demands quite a bit of evidence—more evidence than E. At t2, she doesn’t relinquish her belief or change her credence in p, but she does come to view her belief that p as one that lacks warrant. Intuitively, even if S’s professor was wildly mistaken about how high the evidential demands of warrant are, it still seems as if, given S’s new views on the demands of warrant and S’s own doubt about whether she knows p, S lacks warrant for p at t2.

But in virtue of what is S incapable of knowing p at t2? An answer one might give is that the interests of S have now changed. She values true and warranted beliefs, and her beliefs about what it takes to get warranted beliefs have radically changed. Thus, the sorts of beliefs she values has changed. A defender of a practical interest-relevance theory might say that this change of goals undermines S’s warrant at t2 in virtue of making a different type relevant for the token S used to originally form her belief that p. According to the new relevant type, S’s belief that p wasn’t formed with sufficient reliability for warrant.

Before proceeding, I should at least mention that it’s a point of contention between contemporary epistemologists whether or not cases like Roeber’s are ones in which the agent

---

31 Roeber (MS) 9-12.

32 What we see here is a case that might seem to support a modified version of SS:

SS* Necessarily, for all subjects and possible tokens, the practical interests of the subject S, regarding the nature of a kind of dependability S thinks is constitutive of having warrant suffices to determine the relevant type T for S’s belief-forming process token.
actually loses warrant. But even granting that Roeber’s cases are ones in which the agent loses warrant, I think cases like his fail to offer compelling support for a practical interest-relevance theory. This is because there are other reasonable explanations for why the agents in these sorts of cases are incapable of warrantedly believing p. These other reasonable explanations account for a loss of warrant in Roeber-style cases without invoking a change in the agent’s relevant process type. First, it might be the case that what changes for these agents is not the relevant process type with respect to which the agent must be reliable, but rather the threshold, or, degree, of reliability required for having warrant. Perhaps with her new beliefs on the nature of knowledge, she now (at t2) has to have 99% reliability in the reference class in order to know, rather than 95% reliability (or something around there) that she needed to attain before (at t1). Secondly, it might be the case that the relevant type and threshold for warrant are unchanged for S at t2, but the absence of knowledge arises due to S gaining something like a rationality defeater for p. Remember, S comes to believe not just p, but also I don’t know p. Perhaps this latter belief makes it the case that her warrant for believing p is defeated. Perhaps there are other accounts for the absence of warrant in the Roeber cases. I raise these two here to simply indicate that this sort of case for practical interest-relevance theories is inconclusive at best.

---

33 See Alston (1989: 178-179). In this passage, Alston argues that justification isn’t a necessary condition on knowledge after all. The case he uses is one in which an agent has extraordinarily good evidence to believe that his perceptual faculty is wildly unreliable, and he believes this. Unfortunately, he’s misled by this evidence, and his perceptual faculty is perfectly fine. One day, he’s crossing the street, and sees a car barreling towards him. In a quick moment, his philosophical reflection stops and his perceptual belief-forming faculty kicks in, generating the belief that the car is coming towards him, and on this basis quickly moves himself out of the way to dodge the car. Alston thinks this is a case in which the agent knows that the car is coming towards him, even though he’s not justified: he has great undercutting evidence for this proposition. One could generate a similar example to respond to the Roeber style cases above. Perhaps it’s the case that a philosophy teacher could make you (mistakenly) think that the epistemic support requirement for knowledge is extraordinarily high. Suppose that coming to believe this philosophical claim doesn’t change the credences you place in everyday propositions you believe. Someone sympathetic to Alston’s style of reasoning might just say that in this situation, you do in fact know things--like there’s a red door in front of me, or the things your history teacher tells you about simple civil war battle facts—on the basis of vision or testimony when you find yourself naturally coming to believe them in everyday circumstances, regardless of your false theoretical views about the nature of knowledge.
So what was behind the initial intuitive pull of Greco’s cases that motivate his practical-interest based relevance theory? I do think that Greco’s warrant verdicts on his two cases are correct. The employee not does warrantedly make a barn-identification judgment, but the farmer does warrantedly make a barn-identification judgment, even though they’re both in fake barn county and both make correct judgments. I even agree with Greco that the relevant types in each case are distinct—thus making their reference classes distinct as well. Unlike Greco, I deny that the best explanation for this difference in type relevance has anything to do with anyone’s practical interests or cognitive goals regarding some sense of dependability that could be applied to the judgments in these cases.

To begin, given the details of the first case, in most of the nearby possible barn-identification situations the employee could find himself in, he makes false barn identification judgments. This is because in the actual token barn identification situation, he’s in fake barn county and has no idea about this, and this fact is reflected in nearby possibilities. On the other hand, in most of the modally nearby barn-identification situations the farmer finds himself in, he makes correct barn identification judgments. This is because, in his actual token case, the farmer knows that most of the other surrounding barn shaped objects are fake, so he’s disposed to either withhold judgment on whether those objects are barns or judge (correctly) that they aren’t barns, and this fact is reflected in the nearby possibilities to his token case. While I haven’t argued for it here, I think it’s at least quite plausible that the following features of these two tokens are held fixed in their respective relevant types: the identity of the agent making the barn identification judgment, and the fact that sort of judgment being made is a barn identification judgment. These facts, as I discuss in section 2, make it the case that the reference classes for each respective token are distinct and non-overlapping. I also think the following principle has much intuitive appeal: the reference class for a given token is
populated by belief-forming events that are *modally close* to that token event. One can see, that if these three principles of the reference class and type relevance hold, then the relevant types for Greco’s two cases will be distinct, and their respective reliability measurements will differ just how Greco says they will.

Notice, the explanation for why their reliability measurements and reference classes differ has nothing to do with whether the judgment of the farmer or the judgment of the employee satisfies some standard of dependability that’s determined by the practical interests of anyone. The explanation only invokes nearby possibilities, the content of the token judgment, and the identity of the subject in the token case.\(^{34}\) We’ve yet to be given any independent motivation for invoking practical interest-relative dependability criteria to explain why the reference classes, types, and reliability measurements differ for Greco’s two cases.

To this point, one might think that practical interest-relevance theories haven’t been shown to be a dead end. One reason is that the accounts considered above invoke practical

\(^{34}\) Also, as modal theorists have recognized, the idea of modal closeness and modal distance only makes sense given that the possibilities around the actual case (@)) stand in a particular sort of *ordering relation* to @. The ordering relation determines what makes one possibility further away from (or closer to) @ than another one. Some have claimed this this ordering relation—which is a relation of similarity—can vary depending on practical considerations at play in the context of utterance or thought (see Stalnaker (1968: part II)). Here’s one way in which one might think that type relevance for warrant is determined by practical matters and cognitive goals—it’s these goals that determine the ordering relation of worlds, and thus, which possibilities make it in to the reference class. Ultimately, I think this attempt to vindicate a practical interest-relevance theory is misguided. I’ll grant the sake of argument that features of the context of utterance (or thought) can partially determine whether someone is a *dependable* source of information about some matter. Consider once again the case of Stephanie’s research. One might think that whether she’s correct in claiming that a member of the tour group is dependable depends on the following factors: a reliability measurement taken across some reference class N of possible situations in which that tour member makes a certain kind of judgment, and that the possible situations that make it in to that reference class N are possibilities that are relevantly nearby relative to an ordering relation r1 that’s determined by the specific practical interests corresponding to the sense of dependability Stephanie cares about. Now, however, suppose that Stephanie wanted to attribute knowledge to a member of the tour group. Considering the details of the case, this knowledge attribution would still be clearly false, even though there’s some sense of dependability D1 that she cares about according to which it’s correct to say that the tour member has dependability D1. The relevant upshot for our purposes is to realize that, when one makes a knowledge ascription, a different modal ordering relation r2 comes to have salience, such that, for cases like the tour group, a class of nearby worlds M according to this other ordering relation includes many cases of false barn-identification.
interests in the service of fixing a sense of *dependability*, and that in turn determines the reference class. But it’s conceptually open to craft a practical interest-relevance theory which doesn’t invoke dependability, but rather some other notion. I’m not sure how such accounts would go. All this to say, the accounts currently in the literature, which *do* invoke dependability, don’t appear to have much going for them. More speculatively, I think the tour group cases provide *prima facie* reason to think that the cognitive goals or practical interests of the attributers or subjects just aren’t the sorts of things that can fix reference classes for warrant.\(^{35}\)

3.6 Conclusion

As we’ve seen, reasonable attempts to formulate how beliefs or cognitive goals determine type relevance don’t ultimately pan out. Both types of subject-based relevance theories appear to be dead ends. I think the results of this investigation reveal something deeper about the fundamental nature of the sort of reliability constitutive of warrant: whether a particular belief-forming case exemplifies warrant reliability is an importantly *objective* matter. In other words, it doesn’t depend on an agent’s views about the world or her practical interests. Epistemic reliability is instead a matter of a certain sort of *objective* relationship that obtains between an agent’s representations of the world and the world itself, regardless of whether the agent has any higher order views *about* her acquaintance with the world, and regardless of her practical interests about gathering information. While what I say in this chapter doesn’t strictly entail

\(^{35}\) One might also notice that principles AS and SS invoke a “sufficing” sort of determination. One might think that, even if AS and SS have been shown to be problematic, more modest versions of a practical interest-relevance theory could be salvaged. Perhaps, for instance, one could formulate a principle that only invokes “partial-determination” by practical interests. Obviously, the specific details of such an account would need to be further worked out. At this point, I think that cases like the odd tour group should at least make us skeptical that such a modest account could be successfully developed. The tour group cases seem to give the impression that practical and cognitive goals at play in a given case of belief formation have nothing to do with the reference class pertaining to warrant.
this objectivity thesis, I do think the ways in which belief and practical interest-relevance theories fail point to (in a probabilifying way) the underlying objective nature of epistemic reliability. Oddly enough, the majority of relevance theories on offer frame reliability in this objective way. Many of them point to causal features of the token as the kinds of features that must be held fixed in the relevant type, regardless of whether the agent has any views or goals regarding her belief-forming process type. As I mention in the introduction, the current lineup of these kinds of theories have either been shown to be extensionally incorrect or not informative enough. Nevertheless, given the results of this chapter, pursuing an objective relevance theory is the way to go, whether that means refining the current causal theories, or looking to other kinds of objective features, like modality and nearby possible worlds, for crafting a relevance theory.

36 For example, Kelly Becker defends the following relevance theory: “[T]he relevant process type is the narrowest, content-neutral process [type] that is causally operative in belief production” (2008:263). In addition, William Alston holds that the relevant types are “psychologically realist types” that are “maximally specific, in that any difference in input that is registered by the function indicates a different function (1995: 24). For additional psychological and causal theories, see Beebe (2004), Adler and Levin (2002), Goldman (1986).

37 As I argue in chapter 4 and 5, the key to formulating a correct relevance theory is to adopt relevance principles that hold fixed the psychological and modal properties of the token in the relevant type.
CHAPTER 4

CAUSAL RELEVANCE PRINCIPLES WORTH RejectING

4.1 Introduction

Process reliabilism is an important contemporary externalist approach to analyzing epistemic concepts.¹ Both a process reliabilist theory of warrant and a process reliabilist theory of justification have been presented and defended in the literature.² Process reliabilism about justification and process reliabilism about warrant share this common feature: they both claim that the key ingredient that grounds whether an agent’s belief that p possess the relevant epistemic property is whether the agent used a reliable belief-forming process to arrive at the belief.³ In what follows, for simplicity’s sake, I’ll only write in terms of reliabilism about warrant (henceforth, PR), but it’s reasonable to take at least some of the lessons of this chapter to apply to reliabilism about justification as well.

¹ Process reliabilism is an externalist approach in the sense that it invokes features of the world external to the agent’s mind and outside of the agent’s reflective or conscious access to ground the presence or absence of these epistemic properties for a given belief.

² By warrant, I use this term in the functional sense much like Plantinga does, where warrant is the state the possession of which grounds one’s having knowledge that p so long as she also possesses belief that p and p is true. (Plantinga 1993: v)

³ By “key” ingredient, I take this broad construal here to be consistent with the reliable process condition being the only ingredient that generates the relevant epistemic property. I use the term “key ingredient” here so as to include reliabilist theories which hold that process reliability is a crucial part of warrant (or justification), but that there’s some interesting no-defeater condition—that’s not analyzed in terms of reliability—that must also be satisfied in order for an agent to have warrant (or justification). For an account of warrant in which reliability figures as the central grounding feature, see Goldman’s account (1986:44-5), on which warrant requires that an agent’s belief-forming process have both “local” and “global” reliability.
In this chapter, I’d like to focus on the specific nature of what in particular gets evaluated for reliability according to PR. In a way, there’s an important puzzle for the reliabilist to sort out. One might think that the object that gets evaluated for reliability, according to PR, is rather straightforward: it’s simply the process used by the subject. But things are more complicated here. Processes can be thought of as general repeatable types, or as precise one-off tokens. Plausibly, only the former can be evaluated for reliability or unreliability. As both defenders and critics of PR have noticed, a given process token exemplifies many types. A case of seeing a red object at close distance, and then coming to believe there’s a red object on the basis of that visual experience, exemplifies all of the following types: [belief formation], [perception], [vision], [vision in good light], [vision in good light at a close distance], [vision in good light at a close distance with a medium sized object], etc. Reasonably, not all of these

---

4 Here, I’ll flag one important feature of the relevant literature: there’s been no effort to give a precise account of what a belief-forming process token is. Everyone’s more or less adopted the same sorts of conventions in using this term, and allowing this notion to stay at the rough intuitive level. Perhaps this is for good reason. A belief-forming process token seems to be a certain sort of event. Alas, the metaphysics literature on the analysis and individuation of events is voluminous and vexed. This much seems clear from the way the generality problem literature proceeds: a belief-forming process token is an event that has a maximally precise description. So, it’s individuated by the proposition believed, the agent who comes to believe, the time and place at which the belief is formed (where the place description is maximally precise). The state of the environment leading up to the moment of belief formation also seems to be part of the belief-forming token description, and this gets a maximally precise description as well. This appears to be the rough idea of how this literature has used the notion of a belief-forming process token. But there are some more unanswered specific questions one might ask. For instance, does the environment description component of the token include all of the history leading up to the moment of belief formation, or just part of it? It seems odd to say that part of the belief-forming process token description is a causal world history story leading back to the big bang. But if it doesn’t go back all the way, one might wonder what grounds the token environment description’s stopping at some other point in the past relative to the belief-forming moment. Saying more here would surely lead right back to the metaphysical puzzles surrounding event individuation in general.

5 Richard Feldman (1985) and Goldman (1979) are key figures who highlighted this important distinction for making sense of reliabilism. Most philosophers agree that only types are the sorts of things that can be measured for reliability. Although, recently Juan Comesaña (2006) articulates a way in which tokens could be evaluated for reliability with respect to some space of possible worlds as a reference class. Even if this way of conceiving measuring a token for reliability makes sense, Comesaña, correctly, recognizes that framing reliabilism like this doesn’t get the reliabilist out of the generality problem. The reliabilist still would have to provide an account of which possible worlds were contained within the reference class used to evaluate the token’s reliability. As I argue in 4.2.1, providing an account of the reference class suffices for answering the generality problem as normally construed (as a problem of determining a relevant type).
types are measured for reliability in determining whether the target belief has warrant.\textsuperscript{6} Indeed, perhaps only one of them is “relevant” for determining justification. But \textit{which} type is relevant?\textsuperscript{7} In other words, which features of the token are “held fixed” in the relevant type description? Here, it’s important to point out that this very question constitutes the basis of the famed \textit{generality problem} for reliabilism. According to the literature, this generality problem \textit{question} turns into the generality problem \textit{objection} to PR upon realizing that humans—including the best philosophers—do (and have done) a poor job answering the generality problem question. It’s the nature of \textit{our lack of understanding} into the notion of type relevance that constitutes the key premise in the generality problem objection.\textsuperscript{8}

Here, it’s important to point out that very little work has been done to formulate exactly just \textit{how} our lack of ability to answer the generality problem question generates a defeater for reliabilism. It’s beyond the scope of this chapter to attempt to formulate an argument that represents the strongest version of the generality problem objection.\textsuperscript{9} However, responding to the generality problem has motivated many attempts to formulate a \textit{relevance theory} over the

\textsuperscript{6} For example, presumably, \[belief formation\] isn’t the type that gets evaluated for reliability. Consider someone who, sadly, is a BIV, such that their \textit{perceptual} beliefs are all systematically false. Such a person, presumably, can still have reliable and warranted \textit{intuition} beliefs in a priori claims. But, if the relevant type for these intuition belief tokens was just \[belief formation\], then the reliability measurement would be significantly decreased from all of the false perceptual beliefs formed by the agent (since perceptual belief formation is contained within the category of [belief formation])

\textsuperscript{7} The specific language of \textit{relevant} and \textit{irrelevant} types was introduced to the literature by Feldman (1985: 160). This way of talking about the type that gets measured for reliability has become standard throughout the generality problem literature, and I’ll continue that usage here.

\textsuperscript{8} Earl Conee and Richard Feldman are the arch defenders of the generality problem objection (Feldman (1985); Conee and Feldman (1998) and (2002)), and it turns out that others are inclined to agree with them. See Plantinga (1993: 28-9) and Matheson (2015).

\textsuperscript{9} Indeed, in chapter 1, I argue that successfully formulating an argument against PR from a premise involving our current lack of understanding of type relevance doesn’t look promising. It seems like any plausible approach to formulating this argument requires one to either invoke principles that would generate a defeater \textit{for any theory} (and hence we should reject these principles), or to invoke premises that we have little to no reason to believe.
course of the past three decades. A relevance theory presents principles that determine which features of tokens are held fixed in their corresponding relevant type descriptions. Call these *relevance principles*. It’s also the case that over the course of the past three decades, critics of PR have pointed out crippling difficulties with many of the extant relevance theory proposals.

In what follows, I’ll investigate what I take to be one of the more popular sorts of relevance principles that various relevance theories invoke: causal relevance principles. Causal relevance principles all have the following general form: Necessarily for all possible tokens, if a token has causal features of kind X, then causal features of kind Y are held fixed in the token’s relevant type for measuring reliability. Here, a causal feature of a token is just a fact that’s a constituent in the complete causal explanation for how the agent’s token belief gets formed. I’m particularly concerned with causal relevance principles that determine that a particular kind of causal feature must be held fixed in the relevant type: contingent causal non-mental features (CNs). A CN is causal feature of a token that doesn’t obtain necessarily and that’s also not constituted by a mental state of the subject nor any physical state that grounds (given the psycho-physical laws) some mental state of the subject. It’s plausible that some CN features are held fixed in relevant types for a wide variety of tokens. In order to account for these relevant types, some theorists have adopted CN relevance principles: causal relevance principles that determine that at least some of a token’s CN features must be held fixed

---


fixed in the relevant type for measuring reliability. In this chapter, I argue that there are no true CN relevance principles. This result falsifies the many relevance theories that invoke CN relevance principles.

In §4.2, I introduce in more detail the notion of a causal relevance principle, and the main motivation for adopting causal relevance principles. I then explain CN features, and the motivation for adopting CN relevance principles. Next, I present the two different kinds of relevance theories that have historically posited CN relevance principles: specific causal relevance theories and non-specific causal relevance theories. In §4.3, I present and defend my argument for why all CN relevance principles are false. In §4.4, I conclude by suggesting a new approach for capturing which token CN features are held fixed in the relevant type description.

4.2 Causal Relevance Principles and CN Features

4.2.1 Causal Relevance Principles Motivated and Characterized

A given feature F is a feature of a token event t if and only if F is a fact that’s a constituent of the overall token event t. I define a causal feature of a token belief-forming process as follows:

Z is a causal feature of a token t, where t is a case of S coming to believe p, if and only if Z is a constituent of the complete causal explanation for how S comes to form the target judgment p in the token case t.

I don’t have a rigorous analysis of constituent of a complete causal explanation for how something occurs. Basically, the complete causal explanation for how a given token judgment that p occurs is the complete causal chain (in the token case) that terminates in the act of judging that p. I at least mean this much by a constituent of the complete causal explanation for how the judgment that p occurs.
For all facts $F$ and belief-forming tokens $t$, $F$ is a constituent of the complete causal explanation for how a given belief-forming token $t$ occurs only if

a. $F$ is a state or event constituted by some object or objects exemplifying a set of properties, where at least one of these properties is an active causal relation of causing some other state or event.

This means that facts constitutive of a complete causal explanation for a token are, importantly, positive actors in the causal story for how the token judgment comes about. For example, the following sorts of facts cannot be constituents of a complete causal explanation for how a token comes about:

- the non-existence of a given entity.
- the non-occurrence of some sort of state or event.

Causal relevance principles claim that some causal feature from the token must be held fixed in the relevant type. But what's important about holding causal features fixed in the relevant type? Why invoke a causal relevance principle to begin with? Understanding the answer to this question highlights the central motivation for adopting causal relevance principles in one's relevance theory. Consider the case of a human forming a visual belief that there exists some ball in front of him.

**BALL**

Sam is a normal 20-year old adult human sitting on a park bench day-dreaming at time $t_1$. At $t_2$, Sam wonders if there's a ball nearby him, and decides to snap out of his daydream and attend to his visual field to see. As a matter of fact, right in front of him at $t_2$, there is a ball in plain sight. He attends to his visual field at $t_3$ and comes to have the visual experience representing the ball in front of him. At time $t_4$, he bases the belief that there's a ball on that visual experience. Sam has no reason or disposition to believe there's anything irrational or odd about this case of belief formation, and is not in a position to acquire information that might suggest this.

Plausibly, **BALL** is a case of everyday standard reliable and warranted belief formation. Consider now the notion of a reliability measurement. Intuitively, the degree of reliability for a process depends on an important sort of ratio: the ratio of instances in which some outcome
x occurs to instances in which outcome y occurs, taken across some class of possible cases. This “class of possible cases” used for the reliability ratio measurement is what philosophers (and statisticians) call the reference class. With respect to the concept of reliable belief-formation, the pertinent outcome x is true belief formation, outcome y is false belief formation, and the reference class of possible cases is constituted by particular possible belief-forming process token events. So, for each token belief-forming process with a reliability measurement, there’s a corresponding reference class over which that reliability measurement is taken.

There’s an important relationship between the reference class of a token, and a token’s relevant type: All the particular events that constitute the reference class satisfy the token’s relevant type description. It doesn’t make sense for a token’s degree of reliability to depend or partially depend on a truth/falsity ratio measured across cases that don’t even satisfy the relevant type description. So, it follows that any principles that place constraints on the relevant type thereby place constraints on the reference class for the reliability measurement.¹³

Ultimately, the motivation behind invoking any relevance principles at all is to place the right sorts of constraints on the reference class, so that the correct reliability measurements are delivered for particular cases.¹⁴ Given that, intuitively, BALL is a case of reliable belief

---

¹³ First, offering an account of the reference class suffices to answer the questions that get the generality problem going to begin with. Given the characterization of the generality problem according to Feldman (1985) and Conee and Feldman (1998), the problem for the reliabilist is that, although we can know the features of a given token, we don’t know how to perform the reliability measurement for that token to determine whether that token confers justification or warrant on the target belief. For the token case of seeing and coming to believe in the presence of a red ball (discussed in the introduction), we don’t know whether to perform the reliability measurement with respect to [belief-formation] in general, or with respect to [perceptual belief-formation], or with respect to [visual-belief formation], etc. But the more precise way to formulate this puzzle is in terms of the reference class corresponding to the token: we don’t know whether the token’s reference class is comprised of particular cases of all sorts of belief formation, or just with particular cases of perceptual belief-formation, or just with cases of visual belief-formation, etc. If we could completely determine the features of a token’s reference class, then this effectively would satisfactorily answer the questions Conee and Feldman raise to get the generality problem going.

¹⁴ At this point, one might wonder, what is the relationship between the extension of the relevant type description for a belief-forming process token and the reference class for a belief-forming process token? Both are classes of possible belief-forming event particulars that all share some set of properties in common with the
formation, there’s strong motivation to invoke relevance principles that guarantee that there won’t be a high measure of false-belief-forming scenarios in the reference class. Notice, if it’s the case that the reference class did contain belief-forming situations in which, for instance, Sam is envatted as a BIV and forms false judgments about the presence of a ball, or Sam is being manipulated by an evil demon to form false judgments about the presence of a ball, then these cases of false belief formation would drive down the reliability measurement corresponding to Sam’s actual belief formed by the token in BALL. But since it’s clear that Sam’s belief is highly reliable, it’s reasonable to think that there are true relevance principles that in some way block these obscure, far-off possible belief-forming situations from entering the reference class. One way in which these obscure cases could be blocked is if there were true relevance principles stating that certain kinds of causal features from the token must be held fixed in the relevant type (and thus, held fixed in the reference class as well). For example, suppose there’s a true relevance principle that determines, for the BALL token case, that the token causal feature Sam’s eyeballs took in light waves from the external world must be held fixed in the relevant type for measuring reliability. If this were the case, then there’s no way that the standard BIV or demon scenarios make it in to the reference class to lower the reliability measurement. Eyeballs receiving light waves aren’t a part of the causal explanation for belief formation in BIV or demon cases. Remember, causal features are constituents of the complete causal explanation for how the token judgment occurs. Hence, holding a given token causal feature x fixed in the relevant type entails that x is a constituent of the causal explanations for each of the particular belief-forming events that comprise the reference class.

---

token process being evaluated. One might think that these two classes are identical. While I’m sympathetic to this view, it’s beyond the scope of this chapter to argue for this identity. This much seems rather straightforward: the reference class for a token is either identical to the relevant type’s extension, or the reference class is a subset of the relevant type’s extension. One of these two options must be correct, because every possible belief-forming event particular that constitutes the reference class must exemplify the relevant type description.
Upon considering cases like BALL, it’s plausible that BALL has the reliability measurement it does in virtue of having a relevant type in which certain causal features from the BALL token are held fixed. Remember, the goal of the relevance theorist is to find the principles that determine which token features are held fixed in relevant types. First, one might be attracted to the idea that a relevance principle—presenting the conditions under which a given causal feature is held fixed in the relevant type—articulates the explanation for why that causal feature is held fixed in all of the reference class cases. Secondly, one might think that whether some causal feature is held fixed in the relevant type is solely a function of the causal features of the token. Causal relevance principles have this very structure. In other words, according to causal relevance principles, the token’s complete causal explanation determines on its own whether some causal feature is held fixed in the type. Hence, causal relevance principles have the following form:

**CP1** Necessarily for all possible tokens t, and all possible token causal features, if t has a complete causal explanation constituted by causal features of kind K, then t’s causal features of kind L (which either completely or partially constitute K) are held fixed in t’s relevant type for measuring t’s reliability.

Here’s an example of a causal relevance principle:

**EB** Necessarily, for all possible tokens t, and all subjects S, if t’s complete causal explanation includes—as a constituent—the causal feature S’s eye balls took in photons from the outside world and activated S’s optical nerve, then t’s relevant type exemplifies the causal feature S’s eye balls took in photons from the outside world.

Return to the general form that causal relevance principles take (CP1). The following general theorem follows from CP1.

**CP2** Necessarily for all possible tokens t, and all possible causal token features, if there’s a true causal relevance principle stating that necessarily all tokens exemplifying causal features K have causal features L held fixed in their relevant types (for measuring reliability), then it follows that:

if t1 and t2 both share the same causal explanation E1 and E1 exemplifies K, then the relevant types for both t1 and t2 are both constituted by causal features of kind L.
In other words, causal relevance principles entail that all tokens that share the same complete causal explanation will all possess relevant types sharing the same sort of causal feature—the one mentioned in the consequent of the causal relevance principle itself.

4.2.2 Contingent Causal Non-Mental Features (CNs)

Here, I’d like to draw attention to a specific sort of token causal feature: Contingent causal non-mental features (CNs).

Necessarily, for all features $F$, tokens $t$, and complete causal explanations $E$, a feature $F$ is a CN of $t$ if and only if

a. $F$ is a causal feature constituent of $E$.

b. $F$ is not constituted by a state or event that obtains necessarily.

c. $F$ is not constituted by a mental state of the subject.

d. $F$ is not constituted by a physical state which, by the psycho-physical laws, grounds a mental state of the subject.

Given this definition, mental states of the subject (e.g., beliefs) cannot be CN features, nor are other phenomenological or psychological features of the subject. Interestingly enough, the causal feature *Sam’s eyeballs took in light waves from the outside world* is a CN of the token in BALL. It itself is not a mental state of Sam, and presumably it grounds no mental state of Sam’s given the psycho-physical laws.\(^{15}\) Here are some other CNs from the token in BALL:

1. Sam’s retina receives light waves coming from an environment well-lit by the noonday sun.

2. Sam’s retina receives light waves that didn’t travel through any wavelength-distorting substances in the air.

3. Sam’s non-damaged optical nerve transferred neural signals to the brain.

4. The light entering Sam’s eye comes from an outside environment with no fake ball-facades within 500 miles of Sam.

\(^{15}\) While an eye’s taking in a photon may cause a mental state (e.g., a visual belief), there’s not much reason to believe it grounds the presence of a mental state for $S$ given the psycho-physical laws. Presumably, smaller portions of the brain are responsible for this.
Call a CN relevance principle any causal relevance principle that determines that at least some CN feature must be held fixed in the relevant type. EB is an example of a CN relevance principle, given that the feature it claims must be held fixed in the relevant type (in its consequent) is itself a CN. One can discern a two-part motivation for adopting CN relevance principles into one’s relevance theory. First, CN relevance principles ground reference classes that accommodate our reliability and warrant intuitions on a host of cases—especially normal everyday cases like BALL, in which part of the causal explanation for the token judgment involves some interaction with the subject’s physical environment. Secondly, CN relevance principles, like all causal relevance principles, are quite elegant and explanatorily powerful. After all, they claim that whether some CN is held fixed in the relevant type (and thus, the reference class) only depends on the causal features of the token itself. If these dependence relations are true, it’s a particularly informative and straightforward insight into the nature of epistemic reliability itself. By my lights, this two-part motivation is the best explanation for why many relevance theorists are keen to adopt CN relevance principles to begin with. Next, I’ll present the two broad families of relevance theories that have adopted CN relevance principles.

4.2.3 Two Sorts of Causal Relevance Theories

First, there are what I call specific causal relevance theories. Specific causal relevance theories hold that for every possible token, that token’s entire causal explanation gets held fixed in its relevant type. Given that specific causal relevance theories claim that every feature constituting the token’s causal explanation is held fixed in the relevant type, it follows that any specific causal relevance theory will entail the following principle:
Necessarily, for all possible tokens \( t \), and for all possible CN features \( U \), if \( t \)’s complete causal explanation is at least partially constituted by \( U \), then \( U \) is held fixed in \( t \)’s relevant type.

Consider Kelly Becker’s response to the generality problem: “[T]he relevant process type is the narrowest, content-neutral process [type] that is causally operative in belief production.”\(^{16}\) A type is content neutral just in case the specific judgment \( S \) makes on the target matter \( p \) (belief or disbelief) is not held fixed in the type. By “narrowest causally operative process type,” Becker means the type that holds fixed each of the causally contributory factors of the token that led to the belief.\(^{17}\) On Becker’s view, it follows, for any possible token, that each of that token’s CN features will be held fixed in the relevant type for the reliability measurement of that token.

Specific causal relevance theories have come under significant fire in the past two decades. As many point out, specific causal relevance theories have the problem of fixing relevant types

\(^{16}\) Becker (2008:263)

\(^{17}\) Alston and Beebe endorse similar views to Becker’s. Alston writes,

Instead I will point out how the reasons given by Feldman and Plantinga for doubting that any types satisfy this requirement can be seen to fail if we restrict ourselves to psychologically realist types. And I will end with a few considerations that suggest that the belief-forming functions in the human psyche might well be epistemically homogeneous. In this discussion I will assume that the functions in question are maximally specific, in that any difference in input that is registered by the function indicates a different function (1995: 24 emphasis mine).

Here, Alston’s psychological realist view identifies the relevant type with the psychological function used by the agent in the token case. He views psychological functions as (at least partially) constituted by the many input/output steps that comprise and entire belief-forming process, and the mechanisms that realize these functions (12).

Next James Beebe presents a two-step strategy for solving the generality problem. First, one must take the token and narrow it down into which “information processing type” it belongs (2004:180). This occurs by examining which information-processing problem the token solves, which information-processing algorithm it used, and what cognitive architecture it used (180). Call the information-processing type determined by these three criteria type \( A \). More narrowing is required in the second step. Beebe writes,

\[ [T] \text{he relevant process type for any process token } t \text{ is the subclass of } A \text{ which is the broadest objectively homogeneous subclass of } A \text{ within which } t \text{ falls. } A \text{ subclass } S \text{ is objectively homogeneous if there are no statistically relevant partitions of } S \text{ that can be effected. (181) } \]

The “statistical relevance” at play here is relevance to generating true belief. Given this sense of statistical relevance, every feature from the token that in any way increases the likelihood of the agent’s forming a belief that \( p \) will count as statistically relevant. All of these features will constitute the broadest subclass of \( A \) that is objectively homogenous. Taking out or adding in any one of these details from the token would create a statistically significant different subclass in \( A \).
that are too narrow. These overly narrow types make it the case that nearly every case of true belief ends up being formed with near perfect reliability, and any false belief ends up being formed with near perfect unreliability. This is a rather lousy result for specific causal theories, seeing as how it’s quite reasonable to think that one can form true beliefs in epistemically unreliable ways. Consider applying Becker’s account to the standard fake barn county case. Henry is in fake barn county, about to form the true belief that object O is a barn when he’s looking at the one real barn in the area (O). His belief is true, yet it doesn’t seem warranted. Nevertheless, his barn belief is generated in normal lighting conditions, from a normal distance, by a perceptual experience caused by a barn. More specifically, his token exemplifies the following CN feature:

The light waves that stimulated the optical nerve of Henry’s eyes reflected directly off of something with the physical composition of a real barn.

Specific causal relevance theories would have us hold this fact fixed in the type as well, given that it is a detail that partially constitutes the causal explanation for how the token judgment was formed. But if this fact is held fixed in the type (and reference class), along with all the other constituents of the token’s causal explanation, then given that the belief is one of ascribing barn-ness to an object, the belief-forming process gets, at the very least, a remarkably high reliability rating. But this is the wrong result. Given cases like this, it’s clear that specific causal relevance theories fail.

---

18 Feldman (1985) originally calls this problem of determining types that are too narrow the single case problem for a relevance theory.

19 This case, inspired by Ginet, originally appeared in Goldman (1976).

20 See Dutant & Olsson (2013) for a criticism of Beebe’s account, and Brueckner & Buford (2013) for a criticism of Becker’s account. Here, my criticism of specific causal views slightly differs from those of Dutant, Olsson, Brueckner, and Buford, but remains similar in spirit and general approach.

21 Others, especially Conee and Feldman (1998) have pointed out an equally serious problem for cases of belief formation in necessary truths that all specific causal relevance theories face. To begin, any time an agent
Secondly, there are non-specific causal relevance theories. The relevance principles of non-specific causal relevance theories only determine that certain kinds of token causal features are held fixed in the relevant type. They do not entail that, necessarily, all causal features (for each token) are held fixed in the relevant type. This being the case, they avoid the problems of over-specificity that we saw with Becker’s view. Charles Wallis presents a non-specific causal theory like this. He construes answering the generality problem as giving a specification of the cognitive task performed by the subject.\textsuperscript{22} Specifying the cognitive task involves specifying these three components:

\textsuperscript{22} One might interpret Ernest Sosa as invoking CN relevance principles in a non-specific causal relevance theory. He writes,

At a minimum, for S to believe P at t out of intellectual virtue, there must be a field of propositions F such that P is in F, and there must be conditions C such that S is in C at t (with respect to P), and such that S is nonologically (but not tautologically) likely to be right if S believes a proposition in field F when in conditions C. (1991:278 emphasis mine)

Earlier, he explains having an intellectual virtue (with respect to an example of visual belief) as follows,

Because subject S has a certain inner nature (I) and is placed in a certain environment (E), S would most likely be right on any proposition X in field F relative to which S stood in conditions C. S might be a human; I might involve possession of good eyes and a good nervous system including a brain in good order; E might include the surface of the earth with its relevant properties, within the parameters of variation experienced by humans over the centuries…F might be a field of propositions specifying the colors or shapes of an object before S up to a certain level of determination and complexity…C might be the conditions of S's seeing such an object in good light at arm's length and without obstructions. (139)
an idealized target function between input and output types

a specification of the nomic correlations (including statistical correlations) that underlie the behavior of both the [cognitive] system and the relevant objects within the domain

a specification of the relevant process by reference to the system’s dispositions, viewing these dispositions as a strategy or set of strategies for generating outputs and other associated behavioral responses (if any) from inputs relying on certain nomic correlations (1994: 265 emphasis mine).

Wallis is keen to point out that W1-W3 have a general enough description, for each token, so as to not plunge his theory into the same problem faced by Becker and other specific causal theorists. For W2, the nomic correlations Wallis has in mind are the causal correlations that obtain between the agent’s cognitive apparatus and the objects out in the agent’s environment of habitual cognitive task performance. Wallis makes it clear that the “environment” description for many cognitive processes has a physical component. This being the case,

---

Given that Sosa holds that believing out of intellectual virtue is at least partially constituted by being disposed to most likely be right about some sort of claim in some sort of circumstance, it’s rather natural to see Sosa as offering an account of warrant that can be rephrased in terms of reliable processes. Here, we can read Sosa as claiming that, for a given token t, we must hold fixed features (I), (E), (C), and (F)—relative to t—in t’s relevant type description.

Given that Sosa specifically invokes a nomological high likelihood of forming correct judgments on matters (F) while in circumstances described by (I), (E), (C), it’s fitting to interpret (I), (E), (C), and (F) as causal features of the token. Consider Sosa’s notion of inner nature (I). It’s clear from what Sosa says in the previous quote, that the goodness of an eye and the goodness of a nervous system, in this context, refer to these structures’ causal capacities to reliably generate true beliefs. Importantly, the causal features of an agent’s eyes and an agent’s surrounding physical environment (E) are CN features of the token.

Sosa indicates in some places that practical interests and an agent’s beliefs might play some role in determining the relevant type for an agent (282, 236). So, for a case like visual belief formation, practical interests or background beliefs might partially determine the specific characterization of the subject’s eyes or environment that are held fixed in the relevant type. But, if Sosa thinks that any case of visual belief formation must have a type with some physical environmental and cognitive features held fixed, regardless of the agent’s specific beliefs or interests, then Sosa would be invoking some very general CN relevance principle. I won’t argue that this is the correct interpretation of Sosa, but it’s at least a live interpretation.

23 Ibid. (245-251)

24 Ibid. (266-267)

25 Ibid. (267). Here, Wallis uses an example of visual knowledge for a forest frog. Here, he says the reference class cases over which the frog is evaluated for reliability are determined by the frog’s “environments of habitual performance.” Hence, if a frog regularly lives in a physical forest, then the reference class cases will all be cases that occur in a physical forest.
Wallis’ relevance theory invokes relevance principles that hold CN features fixed in the relevant type. Notice, Wallis doesn’t indicate that whichever CNs are held fixed in the relevant type is a function of anything more than the causal features that obtain in the token case. Hence, we can reasonably interpret Wallis’ relevance theory as invoking CN relevance principles.

One central problem for Wallis’ account is that he doesn’t offer any insight into how narrowly or broadly to characterize the environment of a subject in the relevant type description. In other words, there’s a rather important lack of informativeness in Wallis’ account. But this lack of informativeness isn’t, in and of itself, a reason to abandon a non-specific causal relevance theory. Perhaps all that’s needed is more attention to offering more detailed analyses of the concepts non-specific causal theories invoke. Unfortunately, there’s good reason to believe that more informativeness couldn’t rescue a theory like Wallis’. Insofar as his theory invokes CN relevance principles, his theory is false. In the next section, I’ll argue that all CN relevance principles are false. This result undermines every specific causal relevance theory and every non-specific causal relevance theory that invokes CN relevance principles.

4.3 Against CN Relevance Principles

My argument proceeds as follows. In step 1 of the argument, I present a token case, LUKE, that has the same causal explanation as BALL. However, I’ll argue that, unlike BALL, LUKE’s relevant type has no CNs held fixed in it. Given CP2, it follows that there are no true CN relevance principles that explain or determine how any CNs might be held fixed in BALL’s relevant type. This is consistent with there still being CNs held fixed in BALL’s relevant type and reference class. It’s just that there cannot be any true CN relevance principles to explain
why those CN’s are held fixed in BALL’s type. In step two of the argument, I argue that the
pair of cases in step 1 gives us a general formula we can use for coming up with—for any
complete token causal explanation Eₐ containing any set of CN features—another possible
token that exemplifies Eₐ that fails to have any CN features held fixed in its relevant type. But
if this can be done for every possible token causal explanation with any CNs, then by CP2, it
follows that there are no true general CN relevance principles.

4.3.1 Step One

Here, I should flag that my argument operates with the following presupposition, which I
find intuitively plausible: it’s not the case that, necessarily for all tokens, the features of a token
that are epistemically relevant for being held fixed in the reference class stretch back in time
to the beginning of the agent’s life, or even further back than that. For each human belief-
forming token t, there’s some point of time R that’s prior to the terminal moment of judgment
itself (the last moment of t) with the following feature: All contingently obtaining features of
t that obtain before R don’t matter for the reliability measurement of t, and hence aren’t held
fixed in t’s reference class and relevant type. Call this point in time the token’s time of relevance.²⁶

For instance, consider cases of simple non-inferential perceptual belief formation like
BALL. Imagine a parallel situation exactly like BALL except the subject, Dan, like the classic
swamp-man cases, spontaneously pops into existence due to a freak quantum event, at t₁. Call
this case SWAMP BALL. At t₁, Dan and Dan’s environment are in a particle for particle,
state for state, qualitatively identical position that Sam and Sam’s environment are in at t₁ (in

²⁶ Of course, the duration between the time of relevance and the moment of judgment differs from token
token. Also, perhaps it’s incorrect to claim that there’s some particular instant that’s the moment of relevance.
Perhaps it’s more like a fuzzy/vague range of moments throughout which it’s indeterminate whether any of the
states at those times are held fixed in the relevant type. I can accept this. All my argument below requires is that
the obscure modally close belief-forming possibilities branch off from moments in time at or before the fuzzy
range of relevance moments (in the token’s timeline) occurs.
BALL). Dan goes on to form the belief *there’s a ball* in the exact same way Sam does. It’s reasonable that not only do these scenarios have the same degree of reliability, but that they could also have the same reference classes. More modestly, it’s implausible to think that pairs of tokens like this necessarily must have different reference classes due to their differences in what goes on leading up to $t_1$ in each particular token case. Perhaps one isn’t convinced that $t_1$ is the time of relevance for BALL. Fair enough. I do contend that the time of relevance for BALL couldn’t be that much prior to $t_1$ and it certainly isn’t anywhere close to as far back as Sam’s birth.27

In what follows, I’ll present a token case that has the exact same causal explanation as BALL. However, in this new case, the relevant type fails to hold fixed *any* CN properties.

LUKE

Luke lives in a (possible) world on a planet that’s almost qualitatively identical to Sam’s world. Luke is a human being with a normal adult human cognitive apparatus. One main difference between Sam’s world and Luke’s world, is that Luke’s world contains a very powerful demon creature (in the spirit realm) who has recently taken some twisted interest in making the normalcy of Luke’s life hang on a thread. The demon has created an objectively indeterministic machine—-in the same way nuclear decay and quantum laws are genuinely indeterministic—-that is active at each moment of Luke’s life. The machine also exists in the spirit realm, apart from the physical world. Here’s how the machine works at any moment of time in which it’s activated:

The demon designed the indeterministic machine so that, for any moment $t_i$ at which the machine is activated, there is an objective $0.001\%$ chance that the machine does nothing to alter the next moment in time $t_{ii}$. The next moment $t_{ii}$ will come about in the normal way according to the causal laws of Luke’s world. It’s also the case though, for any moment $t_i$ at which the machine is active, that there’s a

---

27 It’s also possible to conceive of another case like BALL, in which the subject, Pam, from $t_1$ forward, forms the belief *there’s a ball* with the exact same causal explanation for why, from $t_1$ on, Sam forms the belief. However, in Pam’s case, a different history (that’s just as long as Sam’s history) leads up to the moment $t_1$. Call this case ALTERNATIVE BALL. It’s certainly possible for Sam and Pam’s cases to differ in this way if we build in that causal determinism in both of their worlds is false. Perhaps a combination of quantum events occurs in Pam’s past but not in Sam’s. Further, suppose that by the time $t_1$ comes about in ALTERNATIVE BALL, the exact same causal states and dispositions obtain that obtain in BALL at the moment of $t_1$ (except in BALL we have Sam standing in place of Pam). It seems like in such a case, Pam and Sam have the same degree of reliability for their judgments, and also have the same reference classes. More modestly, it’s doubtful that any differences in their reference classes could be due to the differences in their respective pasts before $t_1$. 

114
99.999% objective chance that the machine activates a doomsday device that, by \( t_\text{in} \), will have destroyed every physical object in Luke’s world (including photons) except for the small part of Luke’s brain that’s responsible (given the psycho-physical laws of Luke’s world) for grounding his conscious experiences and basic mental causation abilities. Also, if the doomsday device activates at \( t_\text{i} \), then by \( t_\text{ti} \) the demon uses his powers to protect and sustain this small part of Luke’s brain, and will proceed to stimulate that part of Luke’s brain so as to give Luke all of the same phenomenological experiences Luke would have had from \( t_\text{i} \) forward (given the state of \( t_\text{i} \) and the laws of Luke’s world) had the doomsday device not been turned on. In other words, the demon will make sure Luke has the life experiences Luke would have had anyways from \( t_\text{in} \) onward had the doomsday device not activated.

Luke is now 20 years old, and has lived a life up to this point that is qualitatively identical to Sam’s life. Amazingly (and improbably) for Luke, the indeterministic machine has yet to activate the doomsday device. Today, Luke finds himself in a children’s park daydreaming at time \( t_1 \) and not paying attention to what’s going on around him. At time \( t_2 \), Luke wonders if there’s a ball nearby him, and decides to snap out of his daydream and attend to his visual field to see. As a matter of fact, right in front of him at \( t_2 \), there is a ball in plain sight. He attends to his visual field at \( t_3 \), and has the visual experience representing the ball in front of him. The next moment at \( t_4 \), Luke bases the belief there’s a ball on that visual experience. Throughout this entire span of time, luckily, the doomsday device doesn’t activate. The demon remains completely causally inactive in the explanation of Luke’s belief formation. Had the doomsday device activated at \( t_1 \), the demon would have generated the same visual experience in Luke (at \( t_3 \)) that Luke in fact had at \( t_3 \). Luke would then have gone on to (falsely) believe there’s a ball on the basis of that experience at \( t_4 \). Note: Luke has no reason or disposition to believe there’s anything irrational or odd about this case of belief formation, and is not in a position to acquire information that might suggest this, and this holds even if the doomsday device turned on.

I think it’s clear that Luke fails to have warrant for his belief at \( t_4 \). This intuitive verdict constitutes that first key premise in an argument for the conclusion that no CN properties are held fixed in Luke’s relevant type.

N1   Luke does not have warrant.

N2   The explanation for N1 is that the token \( t_\text{i} \) that Luke uses has an unreliable relevant type.
N3. Necessarily, the low reliability measurement for the relevant type of $t_L$ is at least partially grounded in the particular nature of the *class* of nearby possible belief-forming situations centered on $t_L$. Call this class of possibilities $t_L$’s modal profile.

N4. Necessarily, $t_L$’s modal profile grounds the low reliability measurement (mentioned in N3) *in virtue of* $t_L$’s reference class being at least partially populated/constituted by the class possible belief-forming situations contained in $t_L$’s modal profile.

N5. The nearby modal profile for $t_L$ includes possible belief-forming situations that don’t share any CN properties with the actual causal explanation of $t_L$ from the moment of $t_L$’s time of relevance up through the moment of judgment.

N6. At least some subclass of the reference class for $t_L$ consists in modally nearby belief-forming possibilities to $t_L$ that don’t share any CN features with $t_L$ dating back to $t_L$’s time of relevance. (N1-N5)

N7. Therefore, none of $t_L$’s CN features are held fixed in its relevant type. (N6 and the definition of a relevant type)

N7 (by entailment) follows from N6 and the definition of a relevant type. N6 follows from N1-N5. Seeing as how N1 has much intuitive plausibility, that leaves, N2, N3, N4, and N5 as the four premises that one might attempt to reject. However, each of these premises is independently plausible.

First, consider N2. By “reliable” or “unreliable” I mean these terms in the very straightforward sense that captures the rather intuitive thought that occurs to us when we reflect on LUKE—namely, that it’s highly likely that Luke will form a false judgment. Token $t_L$ must exemplify a relevant type with the right features in order for this likelihood claim to be true. This seems to be the correct reading of LUKE.

One might think it’s a live option to explain the no-warrant verdict in LUKE *not* in terms of a reliability difference, but rather in terms of defeat—a sense of epistemic defeat that cannot be analyzed in terms of reliability or unreliability. In the literature, the two main senses of
defeat that are seen as difficult to analyze in terms of reliability are rationality defeat and evidential defeat.

Rationality Defeat

S has a Rationality defeater (RD) for S’s belief that \( p \) iff were S to reflect (under normal circumstances) on one’s belief that \( p \), it would then seem to S that \( p \) is false or epistemically irrational to hold.\(^{28}\)

Evidential Defeat

S has an evidential defeater (ED) for S’s belief that \( p \) iff S’s total evidence for \( p \) makes it sufficiently probable for S that either \( p \) is false, or that her belief that \( p \) is irrational to hold.\(^{29}\)

While these are interesting notions of defeat in their own right, LUKE doesn’t exemplify either of them. It’s built in to the case that Luke isn’t disposed to think his belief is irrationally formed, and he’s not disposed to uncover any evidence that might suggest that it is irrationally formed.

N3 invokes the modal language of nearness and distance. The notions of “nearby” and “far off” possible worlds got their popularity from the old debates about the truth conditions of counterfactual sentences. One popular model that emerged was the Lewis/Stalnaker semantics, that says, roughly, that sentences of the form If event X had occurred, then event Y would have occurred, are true iff the nearest possible worlds in which X occurs are also worlds in which Y occurs.\(^{30}\) Let \(@\) refer to the actual situation/world. We can imagine a sphere of possible worlds with \(@\) “at the center.” The distance of the radius moving away from the “center” corresponds to the degree of modal distance from the center. Of course, these notions of modal “distance” and “nearness” are metaphorical. Here, modal distance from \(@\) is a sort of modal

\(^{28}\) This notion of defeat is similar to what Lackey (2008: 44-5) calls psychological defeaters.

\(^{29}\) This sort of defeater would probably fall under what Lackey calls a normative defeater (45). See Swinburne (2001), (2004) for a probabilistic account of epistemic justification with which ED would integrate nicely.

\(^{30}\) See Lewis (1973) and Stalnaker (1968).
dis-similarity to @, and modal closeness is a sort of modal similarity. Hence, worlds “farther out” from @ are less similar to worlds closer to @.

For our purposes here, rather than having @ (the center of the modal sphere) refer to the actual world, let’s have it refer to the token belief-forming event. Then, let’s have the modal sphere corresponding to the short (close) modal radius from @ be constituted by a class of possible belief-forming events rather than possible worlds. It’s this class of possible belief-forming events that constitutes the modal profile for token t.L

The plausibility of N3 is brought out by comparing LUKE with another example, namely with BALL. Unlike Luke, intuitively, Sam does have warrant for his belief at t4. The relevant type for Sam’s token is sufficiently reliable for warrant. Notice, however, that the only important difference between Sam’s token and t.L is a difference in their nearby modal profiles. They both share the exact same causal explanation for how their respective beliefs in fact come about. From this fact, we can infer that the nearby modal profile for t.L is doing significant grounding work in determining a low reliability measurement for t.L’s relevant type.

Looking at both Luke and Sam’s tokens, their respective classes of modally nearby belief-forming events, are quite different from each other. This is because, intuitively, roughly 99.999% of Luke’s nearby ball-existence judgments involve Luke coming to make false ball judgments on the basis of non-veridical sensory experience. On the other hand, Sam’s nearby ball-existence judgments involve him forming (at least for the most part) true beliefs on the basis of a veridical perceptual experience.\(^{31}\)

\(^{31}\) The details of Luke’s indeterministic machine is important. To see that these other belief-forming possibilities really are modally close to t.L the analogy with the sort of quantum indeterminacy physicists posit in our world is especially important. At the microscopic level, physicists say, roughly, that there’s just a brute objective probability that an electron will have a given velocity \(x\) and an objective probability that it will have velocity \(y\). That’s just how the probabilistic law works: nothing else would have to change in the world for the electron to have \(x\) rather than \(y\) (or vice versa). In the same way, consider LUKE. While the change from t1 to t2 would be quite extensive if the doomsday device were to turn on at t1, that doesn’t mean the possible futures
N4 states that the metaphysical explanation for why N3 is true is that the class of nearby belief-forming situations to \( t_t \) populates the reference class for \( t_t \). This explanation is a simple and elegant way for the modal profile of \( t_t \) to ground the low reliability measurement for \( t_t \)'s relevant type. It's important to point out that, the argument goes through even with a more modest premise:

\[ \text{N4*} \quad \text{Necessarily, } t_t \text{'s modal profile grounds the low reliability measurement mentioned in N3 in virtue of } t_t \text{'s reference class being at least partially constituted by a fraction of the class of nearby possible belief-forming situations to } t_t. \]

Remember, every reference class case exemplifies the relevant type description. Hence, showing that a given feature isn't held fixed in a token’s relevant type only requires demonstrating that one of the cases constituting the token’s reference class fails to exemplify that feature. It needn’t be the case that anything close to 99.999% of the reference class cases lack that feature.

Given how LUKE is described, and given our definition of a CN, N5 doesn’t seem open to doubt. LUKE makes it clear that all of the CN features that constitute the causal explanation of \( t_t \) are non-existent in 99.999% of the nearby belief-forming possibility space to \( t_t \) for moments at or after \( t_t \)'s moment of relevance. The only parts of \( t_t \)'s causal explanation that exist in these nearby possibilities are the causal features constituted by mental states of Luke. We can be confident of N5 even if we’re not exactly sure what \( t_t \)'s moment of relevance precisely is, because the indeterministic machine is activated for every moment of Luke’s life. Surely, \( t_t \)'s moment of relevance is not anywhere near as far back in time as Luke’s birth.

N1-N5 necessitate the truth of N6. Features of the token that obtain (in the token’s timeline) before the token’s moment of relevance aren’t held fixed in the relevant type. So,

\[ \text{after } t_1 \text{ in which everything gets destroyed are in some way modally far away. Given that the indeterministic machine operates like our quantum laws, these possible futures are very close for Luke at } t_1. \]
the fact that \( t_1 \)'s causal explanation exemplified some CN features in the distant past doesn't entail that any of these CN features are held fixed in \( t_1 \)'s relevant type. Looking at \( t_1 \)'s moment of relevance itself, there are nearby possible belief-forming situations that branch off from that moment that, as we've seen, have no CN features and yet—given N1-N4—are members in \( t_1 \)'s reference class. From N6, and the definition of a relevant type, we can infer N7.

Return now to the case of BALL. Given our initial discussion of the case, it's plausible that the token in BALL has the reliability measurement it in fact has in virtue of some CN properties of BALL being held fixed in its relevant type. But if N7 is true, given CP2, a relevance theorist cannot posit any CN relevance principles to explain why some CN features are held fixed in BALL's relevant type. The token in BALL and \( t_1 \) share the same complete causal explanation. But if this is the case, then there can be no true CN principles (claiming that any given CN feature must be held fixed in the relevant type) that explains why a particular CN feature is held fixed in BALL's relevant type. For someone optimistic that adopting CN relevance principles is the key to formulating an informative relevance theory that accommodates our reliability intuitions on many everyday cases of belief formation, this result for BALL should be rather sobering. But the results get far worse for someone broadly sympathetic to CN relevance principles.

4.3.2 Step Two

What we see in pairs of cases like BALL and LUKE, is a general formula for coming up with possible cases, for any complete causal explanation \( E \), that undermine the possibility of there being any true CN relevance principles that apply to tokens that exemplify \( E \). First, take a given complete causal explanation \( E_i \) that a given token could exemplify, where \( E_i \) is at least
partially constituted by a given CN x.\textsuperscript{32} Second, consider some token Y that in fact exemplifies E. Thirdly, build in to Y that there is some mechanism M—that’s not part of E—that grounds its being the case that, at or before Y’s moment of relevance, there are many modally nearby possible futures C\textsubscript{1}-C\textsubscript{n} branching forward (in time) that are very different from how events actually play themselves out in Y. Fourth, make it the case that at least one of C\textsubscript{1}-C\textsubscript{n} (call it C\textsubscript{1}) is a possible belief-forming scenario in which x doesn’t obtain in the causal explanation of the resultant judgment in C\textsubscript{1}, but that still involves S making a judgment on the target matter p. Fifth, build in that C\textsubscript{1} involves the token subject S having all of the same phenomenological and mental states leading up to the target judgment on p that S has in the actual token Y. Given the comparison between BALL and LUKE, and the plausibility of premises N1-N6 (and especially the key move in N4), there’s just as good reason to believe that C\textsubscript{1} is a member of the reference class for Y. This being the case, it follows that x is not a feature that’s held fixed in Y’s relevant type. I contend that coming up with a possible token case like this for any CN feature y and for any causal explanation z that might include y as a constituent is possible. But given CP2, it follows that there cannot be any true CN relevance principles explaining why y must be held fixed in any token that exemplifies z—and this holds for all CN features y and causal explanations z. Therefore, there are no true CN relevance principles.

Cases like LUKE show us that there are entities with causal powers present in tokens that have a tremendous impact on the token’s relevant type, even though these entities and their causal powers aren’t constituents of the causal explanation for the token. In LUKE, the demon and the indeterministic machine are such entities. They have a significant impact on the reference

\textsuperscript{32} Feature x could be a simple CN or a complex CN constituted by multiple simple CNs.
class (and indirectly, on the relevant type) in virtue of grounding some important modal facts about the token—namely, which possible events are modally nearby.

4.4 Conclusion

I haven’t argued that all causal relevance principles are false. It’s quite plausible that, necessarily, certain mental/psychological causal features of tokens must be held fixed in their respective relevant types. For instance, I find it quite plausible that if a given token is a case of visual belief formation caused in part by the subject experiencing visual phenomenology, then this mental causal feature must be held fixed in the relevant type.

Nevertheless, the conclusion of part 3 leaves us with a puzzle. Intuitively, it seems as if there are many cases (like BALL), where holding fixed some CN features in the relevant type appears to be a crucial explanation for the actual reliability measurements corresponding to these particular tokens. Presumably, we’d like our relevance theory to be able to explain why such CN features are held fixed in the relevant types for tokens like this. As we’ve seen, however, causal relevance principles don’t successfully explain why any CN features might be held fixed in the relevant type, because all CN relevance principles are false.

If CN relevance principles are off the table, how else shall we construct a relevance theory that accommodates our intuitive reliability judgments on particular cases? One critical lesson stands out from our analysis of the LUKE and BALL cases: in order to yield correct reliability verdicts for any given token, one’s theory must make the token’s modal properties a central determining factor.

Remember, the key difference (articulated in N3) between these LUKE and BALL is a difference in nearby modal profiles centered on their respective tokens. This difference ultimately explains why their reference classes are so different. This result indicates that
whichever CN features (if any) are held fixed in the reference class cannot merely be a function of just the causal features of the token themselves. It’s also a function of the token’s nearby modal profile, and we should adopt principles that place constraints on the reference class that reflect the important role modal facts play in determining the characteristics of the reference class. While these principles would undoubtedly be less elegant than CN relevance principles, they would be able to yield correct reference class determinations on cases like LUKE, as well as many others.
CHAPTER 5

ARITHMETICAL INTUITION, AND TWO NEGLECTED PARTS OF A
RELEVANCE THEORY

5.1 Introduction

The generality problem is commonly thought of as one of the most pressing issues for process reliabilism. Process reliabilism about justification and process reliabilism about warrant share this common feature: they both claim that the key grounding feature that grounds why an agent’s belief that p possesses the relevant epistemic property is the agent’s use of a reliable belief-forming process. In this chapter, I’ll focus on process reliabilism about warrant (henceforth PR), but surely many of the upshots of this chapter could apply to process reliabilism about justification. The generality problem for PR begins with a puzzle. Processes can be thought of as general repeatable types, or as precise one-off tokens. Plausibly, only the former can be evaluated for reliability or unreliability. As both defenders and critics of PR

---

1 By warrant, I use this term in the functional sense much like Plantinga does, where warrant is the state the possession of which grounds one’s having knowledge that p so long as she also possesses belief that p and p is true. (cite Plantinga: 1993: v)

2 I use the term “key ingredient” here so as to include reliabilist theories which hold that process reliability is a crucial part of justification (or warrant), but that there’s some interesting no-defeater condition that must also be satisfied in order for an agent to have justification (or warrant). The relevant kind of defeater I have in mind is not one that can be analyzed in terms of unreliability. For an account of warrant in which reliability figures as the central grounding feature, see Goldman (1986:44-5) presents a reliabilist theory of warrant, on which warrant requires that an agent’s belief-forming process have both “local” and “global” reliability.

3 Richard Feldman (1985) and Goldman (1979) are key figures who highlighted this important distinction for making sense of reliabilism. Most philosophers agree that only types are the sorts of things that can be measured for reliability. Although, recently Juan Comesaña (2006) articulates a way in which tokens could be evaluated for reliability with respect to some space of possible worlds as a reference class. Even if this way of
have noticed, a given process token exemplifies many types. A given case of seeing a red ball at close distance—and then coming to believe that there’s a red ball on the basis of visual experience—exemplifies the following types: [belief formation], [perception], [vision], [vision in good light], [vision in good light at a close distance], [vision in good light at a close distance with a medium sized object], etc. Clearly, not all of these types are evaluated for reliability in determining whether the target belief has warrant. Indeed, perhaps only one of them is relevant for determining warrant. The generality problem gets going by asking the following question:

What is it that makes a given process type relevant or irrelevant for epistemic warrant?

In other words, what makes it the case that certain kinds of properties of a token are “held-fixed” in the relevant type description, while certain other kinds of properties of the token aren’t? Here, it’s important to point out that this very question is the starting point of the famed generality problem for reliabilism. According to the literature, the generality problem question turns into the generality problem objection to PR upon realizing that humans—including the best philosophers—do (and have done) a poor job answering the generality problem

conceiving measuring a token for reliability makes sense, Comesaña, correctly, recognizes that framing reliabilism like this doesn’t get the reliabilist out of the generality problem. The reliabilist still would have to provide an account of which possible worlds were contained within the reference class used to evaluate the token’s reliability. As I argue in 5.2.1, providing an account of the reference class suffices for answering the generality problem as normally construed (as a problem of determining a relevant type).

For example, presumably, [belief formation] isn’t the type that gets evaluated for reliability. Consider someone who, sadly, is a BIV, such that their perceptual beliefs are all systematically false. Such a person, presumably, can still have reliable and justified intuition beliefs in a priori claims. But, if the relevant type for these intuition belief tokens was just [belief formation], then the reliability measurement would be significantly decreased from all of the false perceptual beliefs formed by the agent (since perceptual belief formation is contained within the category of [belief formation])
question. It’s the nature of our lack of understanding into the notion of type relevance that constitutes the key premise in the generality problem objection.\(^5\)

Here, it’s important to point out that very little work has been done to formulate exactly how our lack of ability to answer the generality problem question generates a defeater for reliabilism. It’s beyond the scope of this chapter to attempt to formulate an argument that represents the strongest version of the generality problem objection.\(^6\) However, responding to the generality problem has motivated many attempts to formulate a relevance theory over the course of the past three decades.\(^7\) A relevance theory presents conditions or principles that determine which features of tokens are held fixed in their corresponding relevant type descriptions. It’s also the case that over the course of the past three decades, critics of PR have pointed out crippling difficulties with many of the extant relevance theory proposals.\(^8\)

As I mention, it’s still rather unclear how the absence of a correct and informative relevance theory generates a defeater for PR. But whether or not the reliabilist bears such a burden of providing a relevance theory, there are at least two good reasons for reliabilists to work on a relevance theory. First, presenting a working theory of relevance would potentially make more philosophers sympathetic to PR, seeing as how there’s at least some sizable

---

\(^5\) Earl Conee and Richard Feldman are the arch defenders of the generality problem objection (Feldman (1985); Conee and Feldman (1998) and (2002)), and it turns out that others are inclined to agree with them. See Plantinga (1993: 28-9) and Matheson (2015).

\(^6\) Indeed, in chapter 1, I argue that successfully formulating an argument against PR from a premise involving our current lack of understanding of type relevance doesn’t look promising. It seems like any plausible approach to formulating this argument requires one to either invoke principles that would generate a defeater for any theory (thus proving too much), or to invoke premises that we have little to no reason to believe.


\(^8\) For criticisms of Heller and Alston, see Conee and Feldman (1998). For a criticism of Adler and Levin, see Conee and Feldman (2002). For a criticism of Becker, see Brueckner and Buford (2013), For a criticism of Comesaña, see Matheson (2015). For a criticism of Bishop, see Conee (2013), for a criticism of Beebe, see Dutant and Olsson (2013).
population who, as a matter of fact, are hesitant to accept PR because of generality problem considerations. Secondly, and most importantly, making progress on a relevance theory would not only give us greater understanding of the nature of knowledge (supposing PR is true), but would also give us more guidelines in applying PR to particular cases—allowing us to see how and if PR delivers the intuitive warrant verdicts on those cases. In other words, developing a theory of relevance would give us the tools to see just how much explanatory power PR has.

Making significant progress on a relevance theory is the point of this chapter. I make this progress in two ways. First, I’ll present relevance principles for belief-forming cases of arithmetical intuition. Insofar as I only specifically address this one kind of belief formation, my principles don’t come close to constituting a complete relevance theory. But much more importantly, this chapter makes progress on a relevance theory by pointing out two neglected aspects of epistemic reliability that most relevance theories have failed to take account of.

First, most of the generality problem literature proceeds under the assumption that for any given token, there’s just one type that’s relevant for measuring reliability. I’ll argue that this is a mistake, and that in fact there are multiple relevant types for a given token. The second neglected idea has to do with relevance principles that posit that in order to have warrant, an agent must be reliable with respect to a certain field, or set, of propositions. What others haven’t noticed is that there are importantly different ways that one can measure reliability with respect to a particular field. In order to explain warrant in the cases of arithmetical intuition I present below, I argue that we must invoke relevance principles that specifically make use of different forms of measuring reliability within a field. Interestingly, as I explain below, if different reliability measurements are needed to explain whether a token confers warrant on a target belief, then it follows that there are multiple types that are relevant for a given token.
In §5.2, I’ll present some relevant background regarding the metaphysics of relevant types, and discuss the question of whether there could be multiple relevant types for a given token. I explain how this question has largely gone unaddressed in the literature. In §5.3, I present Ernest Sosa’s account of intellectual virtue as a helpful structure for thinking about epistemic reliability more broadly, and for formulating relevance principles in particular. Here, I point out how relevance theorists have given no attention to explicating or invoking the different ways to measure reliability within a field. In §5.4, I present cases of intuitive belief formation, arguing that a complete explanation of reliability for arithmetical intuition will require invoking different reliability measurements for a field, and multiple relevant types for a token. Here, I also argue that the various forms of measuring reliability in a particular field are particularly probative for our understanding the differences between intuitive warrant and perceptual warrant. In §5.5, I present some closing comments and suggestions for further relevance theory development that incorporate the two neglected aspects of epistemic reliability presented in this chapter.

5.2 Type Relevance and Relevance Principles

5.2.1 The Metaphysics of Type Relevance

A type is a universal. This means that a process type description has, as its extension, all the possible particular token cases that exemplify the type. I’d like to connect the concept of a process type’s extension with the concept of a reference class for a reliability measurement. Intuitively, the degree of reliability for a process depends on an important sort of ratio: the ratio of instances in which some outcome x occurs to instances in which outcome y occurs, taken across some class of possible cases. This “class of possible cases” used for the reliability ratio measurement is what philosophers call the reference class. With respect to the concept of
reliable belief-formation, the pertinent outcome x is true belief formation, outcome y is false belief formation, and the reference class of possible cases is constituted by particular belief-forming process token events.

For example, suppose that the reliability measurement, corresponding to a token’s relevant type, is 80%. This means that the ratio of true beliefs formed to false beliefs formed, taken across all of the particular cases of the token’s reference class, is 4:1.

At this point, one might wonder, what is the relationship between the extension of the relevant type description for a belief-forming process token and the reference class for a belief-forming process token? Both are classes of possible belief-forming event particulars that all share some set of properties in common with the token process being evaluated. One might think that these two classes are identical. While I’m sympathetic to this view, it’s beyond the scope of this chapter to argue for this identity. This much seems rather straightforward: the reference class for a token is either identical to the relevant type’s extension, or the reference class is a subset of the relevant type’s extension. One of these two options must be correct, because every possible belief-forming event particular that’s a member of the reference class must exemplify the relevant type description. The reliability measurement for a given token cannot depend on what occurs in possible belief-forming events that don’t exemplify the relevant type description.\(^9\)

Regardless of whether the subset or identity view is correct, we can conclude the following upshots for our engagement with the generality problem and various relevance theories. First,

\(^9\) Here’s one possible reason to hold that the reference class is just a subset of the relevant type extension: while the extension of the type contains all possible belief-forming event particulars that satisfy the type description, perhaps only modally nearby possible belief-forming events (satisfying the type description) are determinative of the reliability measurement. Notice, that if relevant type descriptions themselves contain modal distance qualifications, then it looks more reasonable to hold that relevant type extensions are just identical to reference classes.
offering an account of the reference class suffices to answer the questions that get the

generality problem going to begin with. Remember, the generality problem starts with the
observation that merely being acquainted with the process token doesn’t acquaint us with the
entity being measured for reliability.10 The entity over which reliability is measured just is the
token’s reference class. Secondly, given the relationship between the reference class and the
relevant type, any principle that places constraints on the relevant type is also thereby placing
constraints on which features are held fixed across the reference class.11 All of the relevance
theories put forward in the generality problem literature are ultimately theories that place
constraints on the reference class descriptions for tokens. In what follows, I’ll continue to
invoke the concept of a relevance theory constituted by various relevance principles. Here,
I’ve simply pointed out that a given principle P is a relevance principle only if it places
constraints on the reference classes for tokens. I’ll write in terms of relevance principles
“determining the relevant type for a token” and “determining the reference class for a token”
more or less interchangeably. Given the close connection between the property relevant type
and the property reference class, my terminology still accurately reflects the actual nature of the
principles historically put forward by relevance theorists.12

10 Given the characterization of the generality problem according to Feldman (1985) and Conee and Feldman
(1998), the problem for the reliabilist is that, although we can know the features of a given token, we don’t know
how to perform the reliability measurement for that token to determine whether that token confers justification
or warrant on the target belief. For the token case of seeing and coming to believe in the presence of a red ball
(discussed in the introduction), we don’t know whether to perform the reliability measurement with respect to
[belief-formation] in general, or with respect to [perceptual belief-formation], or with respect to [visual-belief
formation], etc. But the more precise way to formulate this puzzle is in terms of the reference class corresponding
to the token: we don’t know whether the token’s reference class is comprised of particular cases of all sorts of
belief formation, or just with particular cases of perceptual belief-formation, or just with cases of visual belief-
formation, etc. If we could completely determine the features of a token’s reference class, then this would
satisfactorily answer the questions Conee and Feldman raise to get the generality problem going.

11 Although, as I describe above, it’s not obviously the case that the converse of this relation holds.

12 Wallis (1994:251-262), argues that the relevant type and the reference class are distinct, and hence there
are actually two questions raised by the generality problem: First, what is the relevant type for a token? Secondly,
what is the reference class for measuring reliability for a token? He claims that making progress on the generality
As I mention above, relevance theories state principles regarding which types are relevant for tokens. It’s important to remember that the sort of type relevance we’re concerned with here is type relevance for *epistemic warrant* (as opposed to some other epistemic property, like justification).\(^{13}\) It’s reasonable that the demand for an informative theory of type relevance is ultimately a demand for a deeper explanation of what warrant amounts to on a reliabilist picture. Given that this is the case, it’s appropriate to formulate relevance principles as necessary or sufficient conditions on warrant. The relevance principles I present and defend in this chapter are necessary conditions that specify the kinds of reference classes exemplified by tokens of arithmetical intuition that must have high reliability measurements in order for that token to confer warrant on the target belief.

5.2.2 Multiple Relevant Types, or Only One?

Throughout the vast majority of the generality problem literature, it’s been assumed that the correct way to formulate the notion of epistemic reliability is in terms of the token exemplifying some *one* relevant type that’s sufficiently reliable.\(^{14}\) But it’s at least conceptually open that epistemic reliability (with respect to warrant) entails that the token has high reliability problem requires keeping these two questions (and their answers) distinct in our theorizing. Here, however, he doesn’t specifically argue why, methodologically, it’s important for addressing these two questions separately. As I’ve articulated above, if one offers an account of the reference class, one thereby also succeeds in offering an account of the relevant type—given that all of the reference class cases share the relevant type description in common.

\(^{13}\) It’s at least conceptually open that the relevant types corresponding to reliability measurements for warrant are different than the relevant types corresponding to reliability measurements for justification. Goldman proposes a theory of warrant and justification that seems to indicate that the relevant types for warrant are different than the relevant types for justification (1986: 44-5). He claims that justification only requires “global reliability” but warrant requires both “local” and “global” reliability. According to Goldman, the difference between these two kinds of reliability is “the range of uses for which the process is reliable” (44-5 emphasis mine).

\(^{14}\) Both Conee and Feldman both write in terms of some singular relevant process type. Feldman writes that process reliabilism posits there being “some ‘relevant’ type.” (1985:160). See Conee and Feldman (1998:2); (2002:98); Conee (2013:753); Schmitt (1992: 140); Plantinga (1993:28). Alston (1995: 2-4). All these thinkers, and many more, articulate the generality problem in terms of some *one* type that the process reliabilist must identify as the relevant one.
measurements with respect to *multiple types*. This possibility was flagged by Bishop (2010), but the only philosopher who’s explored this at length is Mark Wunderlich (2003). I say “explore” because he doesn’t specifically defend a set of relevance principles. As he states, there’s much work to be done in finding “selection principles” that, roughly, determine which types to consider when determining a belief’s epistemic status. Ultimately, I think Wunderlich is correct about the relevance of multiple types, and hence, the relevance of reliability measurements across multiple reference classes. Therefore, philosophers should craft relevance principles that accommodate this fact. Basically, I’ll argue in favor of a multiple-types relevance theory. To make this notion a bit more precise, multiple-types relevance theories claim that there are at least some tokens for which, whether that token confers warrant on the target belief is determined by multiple distinct reliability measurements corresponding to multiple distinct reference classes. Below I’ll present which sorts of reference classes are relevant for evaluating cases of arithmetical intuition.

5.3 The Structure of a Relevance Theory

I contend that there is a two-component structure to any principle of relevance. This structure can helpfully be read off of Ernest Sosa’s theory of intellectual virtue. Ernest Sosa is certainly an externalist reliabilist of sorts. As a necessary condition on warrant, Sosa makes clear that the agent must be reliable in some reference class of possible belief-forming situations. He captures this notion in terms intellectual virtue (or competence). Knowledge for Sosa amounts to a certain sort of success from ability. He claims that knowledge is apt

---

15 (2003: 249,257) Wunderlich, specifically writes in terms of justification, but we can apply his ideas to a reliabilist theory of warrant. Wunderlich thinks that the degree of justification a belief gets from its token is a function of the overall reliability score of what he calls a reliability vector, which has as its vector components the reliability scores of many of the types exemplified by the token. (246)
belief—belief that is accurate because it is adroit.\textsuperscript{16} Accuracy amounts to forming a true belief. Adroitness here refers to an agent’s use of a competence or virtue she possesses in the course of undergoing a belief-forming process. Aptness occurs when an agent arrives at a true belief because she used a competence to form the belief. In other places, Sosa describes apt belief as believing out of intellectual virtue. Sosa’s more in depth description of what it is to believe out of intellectual virtue reveals the reliabilist character of his overall theory.

Subject S believes proposition P at time t out of intellectual virtue only if there is a field of Propositions F, and there are conditions C, such that: (a) P is in F; (b) S is in C with respect to P; and (c) S would most likely be right if S believed a proposition X in field F when in conditions C with respect to X. \textit{(1991: 138)}

Given that Sosa holds that believing out of intellectual virtue is at least partially constituted by being disposed to most likely be right about some sort of claim in some sort of circumstance, it’s rather natural to see Sosa as offering an account of warrant that can be rephrased in terms of reliable processes.\textsuperscript{17} Here, we can read Sosa as claiming that, for a given token t, we must hold fixed some set of conditions C and a field F—corresponding to t—in t’s relevant type description. Insofar as that type is sufficiently reliable, that token can confer warrant on the target belief.

Just looking at the structure of Sosa’s theory of intellectual virtue, we can discern two important components of a relevance principle—where relevance principles are construed as necessary conditions on warrant:

1. The Conditions Component
   This component describes which states (not having to do with belief content) of the token are held fixed in the cases constituting the reference class.

\textsuperscript{16} Sosa (2007: 22-4).
\textsuperscript{17} Sosa, in the text, (1991:138) seems reticent to categorize his view as a “historically reliabilist” view, which is how he views Goldman’s standard process reliabilist framework. This is because he wants his theory to be able to account for the warrant conferred by instantaneous (and hence, “non-historical”) mental processes, like grasping one’s own existence in the cogito. Here, in how I’m using the term “reliabilism,” I’m taking this to possibly include instantaneous mental belief-forming process events too. So, Sosa’s theory fits within the broader reliabilist camp.
2. The Field Component
   This component states the field of propositions—that contains the target proposition from the token—on which the subject makes judgments in the cases constituting the reference class.

Reflecting on particular cases reveals Sosa’s two-component framework to be particularly helpful in explicating all of the aspects of epistemic reliability that are intuitively constitutive of having warrant. Consider the following case:

RED

Sam, with a normal human visual faculty, under normal lighting conditions, and at a good range, looks at a red ball, has visual experience that represents to him the presence of a red ball, and on the basis of this experience comes to believe that object $x$ is red. Sam has no defeaters for this proposition.

I’ll discuss the conditions component first. Intuitively, RED is a paradigmatic case of warranted belief formation. The reliability measurement for this belief is quite high. However, suppose for the sake of argument, that the relevant type description was just Sam forming judgments on whether some object is red on the basis of visual experience. If this were the case, then the extension of this type description will at least include possible belief-forming events in which Sam forms judgments on whether some object is red on the basis of visual experiences caused by evil demons, or experiences caused by brain-in-vat computer systems. Importantly, Sam forms false judgments in all of these bizarre possible cases. Were these cases of false belief formation included in the reference class for Sam’s token, the reliability measurement for the token would take a hit.

But supposing that RED takes place in a normal world like ours, these bizarre possibilities are modally very far away. They’re modally far off in the sense that, structurally and causally, they’re so much different than the actual world. Intuitively—and perhaps this is one of the intuitions at the heart of the general externalist-reliabilist approach to epistemology—what
happens in these modally far-off and bizarre belief-forming situations is not epistemically relevant to whether a token confers warrant on a target belief.

Hence, reliabilists are keen to invoke principles that screen off these far-off and bizarre belief-forming situations from reference classes. The way in which these possibilities are screened off is captured in the conditions C component of a relevance principle. One popular way of screening out such bizarre possibilities is to adopt relevance principles specifying that certain causal features exemplified by the token must be held fixed throughout the reference class. Call these causal condition relevance principles. Here, a causal feature of a token is any fact that's a constituent of the token's complete causal explanation for how the token judgment comes about. To account for a case like RED, one could adopt causal condition relevance principles that hold fixed (in the reference class) causal features like *S’s use of eyes in a well-lit environment, the operation of a functional nervous system*, etc.\(^{18}\) Seeing as how the subject’s eyes are not used in BIV belief-forming scenarios, this sort of relevance principle wouldn’t allow this sort of bizarre scenario into the reference class for a token like RED.

Another approach to blocking out bizarre and modally far off possibilities is to adopt purely modal condition relevance principles. Purely modal condition principles claim that the only possible belief-forming events that make it in to the reference class are possibilities that lie within some specific modal distance from the token case.\(^{19}\)

In this chapter, I’m focusing mostly on the field component, and how there are different ways to measure reliability with respect to a field. So for simplicity’s sake, in this chapter I’ll


\(^{19}\) In chapter 4 of my dissertation, I argue that causal condition relevance principles are rather limited in their ability to account for all of the important features of a reference class.
present relevance principles that take a purely modal approach to the conditions component just so we can get some relevance principles on the table for investigating the field component. This is a natural choice for the following reason: given how the motivation behind causal condition relevance principles is to block out modally far-off situations from the reference class, there seems to be something more fundamental about the purely modal approach. The condition components of the relevance principles I’ll invoke will place the following constraint on the reference class: for a given token, a possible belief-forming event can be present within the reference class only if that possible belief-forming event is within modal distance $D$ from the token case.\textsuperscript{20}

While Sosa’s conditions component is particularly helpful for formulating the specific circumstantial features of the token case that are relevant to the reliability measurement, the conditions component doesn’t capture every epistemically important detail. We see the importance of a field $F$ component to a relevance principle by considering the following sorts of questions about RED. “In order for Sam to have warrant in this token case, does Sam need to be generally reliable at making judgments ascribing redness to objects, or does he need to be reliable at making judgments that ascribe any color to objects? Or, rather, must he be reliable in his judgments ascribing any perceivable property to objects? These are questions about the relevant field corresponding to the reliability measurement for Sam’s token.

Here, I want to draw attention to Sosa’s condition that an agent must “most likely be right” about claims in $F$ in $C$. Up through this point, my discussion of reliability measurements has

\textsuperscript{20}The notions of “nearby” and “far off” possible worlds got their popularity from the debates about the truth conditions of counterfactual sentences. One model that emerged was the Lewis/Stalnaker semantics model, that says, roughly, that sentences of the form \textit{If event X had occurred, then event Y would have occurred}, are true iff the nearest possible world in which X occurs is also a world in which Y occurs. (Stalnaker (1968), Lewis (1973)) Of course, the notion of modal distance is actually figurative, corresponding to some measure of modal similarity or difference between two possible situations/worlds.
been rather simple. You look at a reference class of situations in which S makes a judgment on whether p and you take the ratio of the measure of cases in which S judges correctly against the measure of cases in which S judges incorrectly. But notice, in this rather easy construal, the only thing we’re measuring is reliability with respect to judging one proposition p. What no one working on the generality problem in particular—or reliabilism in general—has discussed is that the notion of a reliability measurement gets much more complicated once reliability in a field of propositions is raised as a requirement for warrant, where these fields contain many propositions. It turns out that there are three different ways of understanding what S’s being reliable in a field of propositions F might amount to.

Max Positive Reliability: S is such that, for all of the claims in F, were S to make a judgment on any of them, S would most likely be correct on all of those judgments.

Min Positive Reliability: S is such that, for the vast majority of claims in F (constituting some subclass of F that’s nearly identical to F), were S to make a judgment on any of the claims in that subclass, S would most likely be correct on all of those judgments.

Negative Reliability: S is such that, for all the claims in F, were S to make a judgment on a claim in F, most likely that judgment would be correct. The sense of “a claim in F” here is unique and important. S’s being negatively reliable in F is consistent with S being such that she would form incorrect judgments on most of the claims in F were she to form judgments on them at all. All negative reliability in F requires is that, for the subclass of claims in F that S is disposed to make judgments on, the vast majority of them are claims S would most likely get right. S would cease to be negatively reliable in F were she to gain dispositions to make judgments on the other claims in F she’d most likely get wrong.

Negative reliability is less demanding than positive reliability. Either sort of positive reliability entails negative reliability in that field F, but the converse doesn’t hold. Here, it’s
important to remember that the “likeliness” claims within these three measurement descriptions are relative to a reference class comprised of belief-forming events within modal distance D from the token. For example: Suppose F has 100 propositions in it. Suppose, for example, S is such that if she made a judgment on any of the claims 1-5, she’d most likely get them right (within modal distance D), but if she made a judgment on any of the claims 6-100, she’d most likely get it wrong. But, suppose also that S is constituted such that it’s most likely that she will withhold judgment on all claims 6-100 (within D). Such an agent is still negatively reliable within F. But such an agent wouldn’t be min or max positively reliable in F. In order for S to have min positive reliability within F, she must be constituted such that were she to make a judgment on the vast majority of claims within 1-100 (within modal distance D), she’d most likely get them right. For max positive reliability, she must have this sort of reliability for all of the claims 1-100. Note, that if the relevant field F only includes 1 claim (the target proposition from the belief-forming token), then being positively reliable and negatively reliable with respect to F amounts to the same thing. The notions only pull apart if F contains multiple propositions.

There’s another important way to conceive of positive reliability. Suppose F contains propositions \( p_1 - p_n \). To say that S has max positive reliability within F, with respect to belief-forming possibilities within modal distance D, is just to say that the reliability measurements for the following distinct reference classes are all high:

- the reference class of cases within D of S forming judgments on \( p_1 \)
- the reference class of cases within D of S forming judgments on \( p_n \)

... 
- the reference class of cases within D of S forming judgments on \( p_n \)
For min positive reliability, it would just be the case that the reliability measurements for most of these reference classes was high.\footnote{If S has max positive reliability within F (p_1-p_n), then it follows that, first, there’s a high reliability measurement across the reference class constituted by particular possibilities that satisfy the type description \textit{S making a judgment on p}. Secondly, it follows that there’s a high reliability measurement across the reference class constituted by particular possibilities that all satisfy the type description \textit{S making a judgment on p_2}…and so on up through the high reliability measurement for the reference class comprised of possible particulars that all satisfy the type description \textit{S making a judgment on p_n}.}

5.4 Arithmetical Intuition

5.4.1 Negative Reliability and Multiple Relevant Types

It’s beyond the scope of this chapter to offer a rigorous theory or analysis of intuition. Here, I use the notion of intuition in a way that’s right in line with the common view accepted within epistemology. According to this common view, intuition is a faculty that generates judgments on metaphysically necessary truths. In this chapter, I’m going to focus on just one form of intuitional belief formation: arithmetical intuition—intuition with respect to the metaphysically necessary claims of arithmetic. While intuition, along with memory, can be used in inferential belief formation (e.g., running through a long math proof in one’s head), here I’m just going to focus on intuition as a source of non-inferential belief.

It’s rather straightforward that in order to know p—whether it’s from perception, intuition, or some other faculty—S must be negatively reliable with respect to a field containing just p itself.\footnote{And given that negative reliability is the same as positive reliability when the field contains just one proposition, it follows that necessarily, in order to warrantedly believe p, S must be positively reliable with respect to p itself.} To have warrant, it can’t be the case that the odds are high (within D) that, if S were to make a judgment on whether p, S would get it wrong. But what I’m interested in here is whether there are interesting and important necessary conditions on
arithmetical warrant that invoke negative or positive reliability for fields that include more than
the target proposition.

To begin, there are cases that suggest that there’s a true relevance principle for arithmetical
intuition that posits the need for negative reliability over some field of claims. Note, in what
follows, I take verb phrase intuiting $p$ to be factive, but the phrases basing $p$ on intuition and using
intuition to form $p$ and basing $p$ on an intuitive seeming to be non-factive.

PETER

Peter is a normal adult human, except for one unfortunate feature. Three
months ago, he drank some contaminated water. The harmful chemicals
ended up scrambling his brain with the following (and only) effect: His brain
is now constituted such that whenever he attempts to intuit arithmetical claims
in his head, he almost always does this incorrectly—Nowadays, he’s wrong
55% of the times he uses intuition to non-inferentially come to believe
arithmetical claims of one sort or another. Unfortunately for Peter, he loves
doing arithmetic in his head, so he’s always running numbers and calculations
through his mind. Hence, every hour of every day, he’ll run numbers through
his head and proceed to form many false intuitive judgments, like misjudging
the quotients of $9/3$, $12/4$, $458/17$, $56/9$, $144/12$, $357/33$, misjudging sums
like $2+5$, $23+58$, $59+46$, misjudging products like $44\times4$, $5\times3$, and so on (at a
rate of 55% incorrectness). Although, there is one arithmetical claim that has
been shielded from the effects of his unfortunate drinking incident: whenever
he considers what the quotient of $24/8$ is, he’ll always intuit that the answer is
3. Peter has no idea that this terrible accident has happened to his brain. He’s
also in an odd isolated environment, so that he’s not in a position at all to
discover that he has such a cognitive malfunction regarding his non-inferential
intuitive belief formation. He himself has no inkling that there might be
something wrong with him. One day, Peter considers what the quotient of
$24/8$ is. He then has an intuitive seeming that the answer is three, and bases
his belief (that the answer is 3) on that seeming.

Here, it doesn’t seem like he knows the answer is 3. This is the case even though, with
respect to the target claim $p$, he’s quite reliable (both positively and negatively). Also, given
the last few details in the case, it doesn’t seem like his lack of warrant can be accounted for by
his possessing some sort of a rationality defeater or counter-evidence for his belief. It’s
stipulated that he’s in no position at all to discover that anything odd is going on with his
arithmetical belief formation. The best explanation for his lack of warrant is grounded in the abundance of modally nearby belief-forming events in which Peter forms false intuitive arithmetical judgments—where these are judgments regarding claims other than $24/8 = 3$.

In particular, we see that Peter in fact makes many false arithmetical intuitive judgments, and is disposed (with respect to nearby worlds) to make many false arithmetical intuitive judgments. It’s this dispositional fact that seems to undermine his warrant. Given that negative reliability with respect to some field amounts to the absence of a disposition for lots of error within that field, I think we can infer that PETER is a great illustrative example showing that there is a negative reliability necessary condition on arithmetical intuition. But what is the field with respect to which negative reliability is required? Given that Peter is making mistakes across all sorts of arithmetical claims, and that all of these mistakes seem to be relevant to undermining his warrant in the case, I think there’s good reason to hold that the negative reliability requirement applies to a field containing all arithmetical claims. Hence, we should adopt the following relevance principle:

\[ \text{NA} \quad \text{Necessarily, if } S \text{ has warrant for her intuitive judgment whether } p, \text{ where } p \text{ is an arithmetical claim, then } S \text{ is negatively reliable, with respect to possible intuitive belief-forming events, within modal distance } D \text{ of the token, in which } S \text{ makes intuitive judgments on claims present in the field comprised of all arithmetical propositions.} \]

While the reader might be rather confident that PETER shows that there’s some negative reliability condition on arithmetical intuition, she might be less convinced that relevant field contains all arithmetical propositions. I’ll delay my explanation for this choice until after I’ve explained the positive reliability condition on arithmetical intuition. Suffice it to say, there’s good reason at this point to accept a negative reliability condition in a field containing many arithmetical propositions.
Surprisingly, this result also entails that there are *multiple* reliability measurements across distinct reference classes that determine whether one has warrant for an intuitive arithmetical belief. As I mention above, in addition to satisfying NA, an agent must also be negatively (and thus, positively) reliable with respect to just the target claim itself. Peter highlights how one can have negative reliability with respect to the target claim while lacking negative reliability with respect to a larger field of many arithmetical propositions. But one can also satisfy NA while failing to be negatively reliable with respect to the target judgment itself. Consider Grant, who is negatively reliable with respect to all arithmetical claims, but, due to some neurophysiological glitch, is unreliable with respect to judging whether $24/8=3$. One day, he makes the intuitive judgment that $24/8$ in fact equals $3$, luckily coming to the correct answer. In most nearby possible belief forming situations in which he forms a judgment on this quotient, he judges incorrectly. Even though Grant satisfies NA, he fails to have warrant in this particular case. It follows that there’s a reliability measurement across a reference class that only includes possible belief-forming events in which Grant makes judgments on the specific target claim, that partially determines whether Grant has warrant. This reference class is distinct from the reference class corresponding to the type invoked by NA, which is comprised by cases of Grant forming judgments on *many* different arithmetical claims. Importantly, claiming that both of these distinct reference classes corresponding to a given token of arithmetical intuition are determinative of whether that token confers warrant on the resultant belief commits one to a multiple types relevance theory. Reasonably, adopting a multiple-types relevance theory is the right way to accommodate our intuitions on the cases of Peter and Grant.
5.4.2 Positive Reliability and a Key Difference from Perception

There are cases that motivate an additional positive reliability necessary condition on intuitive arithmetical warrant. Consider the following claim:

\[ \text{GB} \quad \text{The peano axioms of arithmetic entail that Goldbach’s conjecture is true.} \]

Goldbach’s conjecture is the arithmetical thesis that \textit{every even natural number greater than 2 is the sum of two prime numbers}. The peano axioms of arithmetic are the formal axioms governing natural number arithmetic. Interestingly enough, Goldbach’s conjecture hasn’t been \textit{proven} to be true, even though it’s been shown to hold for a great many particular cases of even numbers. Being a thesis of arithmetic, its truth or falsity follows from the peano axioms. Suppose that GB is true. It’s clear that GB is not an arithmetical claim that normal human beings have the ability to reliably (and warrantedly) intuit, unlike claims like \((11/2)>(9/2)\), which we can reliably intuit. But, it seems metaphysically possible for there to be creatures with remarkably reliable intuitive abilities who can reliably intuit \textit{and thereby come to know} claims like GB. Call such a species of creature \textit{species Z}. For the purposes of the rest of this chapter, let’s just assume that Goldbach’s conjecture and GB are true. Members of species Z can simply reflect on upon GB, reliably obtain an intuitive seeming that GB is true, and base a belief in GB on that seeming. Members of species Z are also reliably disposed to correctly intuit all sorts of other remarkably complicated arithmetical theorems as well. But what exactly does having a reliable warrant-conferring intuitive capacity amount to? Considering the following case provides some insight.

\[ \text{WES} \]

Wes is a human adult who’s taking an introductory college survey course called \textit{basic puzzles in mathematics}. Wes is in this basic course because he was never really that great at high school math. He has the basic arithmetical abilities (addition, division, subtraction, etc.) that one learns in elementary school, but he never excelled in algebra or pre-calculus. He knows he’s bad at those subjects, so he never tries algebra or pre-calculus anymore (although, he still
performs basic arithmetic all the time). One day in his *basic puzzles* class, Wes learns about the Peano axioms and Goldbach’s conjecture. He’s fascinated that no proof of the conjecture has been found. Wes then begins to zone out of the lecture and consider whether the thesis GB is true. Interestingly enough—due to some odd growth of neurons in his brain from birth—he is stably disposed such that, were he to ever consider whether GB is true, he would receive an intuitive seeming that GB is true with the characteristic sort of intuitive phenomenology that members of species Z have when they intuit that GB is true. In fact, this is exactly what happens to Wes when he considers GB, and on the basis of his intuitive seeming he comes to believe GB. The odd brain growth gives Wes no other dispositions to form any other mathematical beliefs. Wes is also rather unreflective, in how he doesn’t feel as if there’s anything odd or misplaced about his intuitive seeming that GB is true. To him, it just feels like the sorts of intuitive seemings he gets when he intuits basic addition claims. It’s also the case that Wes lives a pretty solitary and ‘aloof’ life. He never really talks to others about his thoughts, so he’s in no position to learn that there’s anything odd or uncommon about the intuitive seeming he just had.

Here, I contend that Wes doesn’t warrantedly believe GB on the basis of intuition. While it is the case that his odd brain growth gives him some sort of a cognitive ability with respect to GB, it doesn’t give him the same sort of warrant conferring competence that a member of species Z has. This is true, even though the case highlights how Wes is negatively reliable with respect to GB itself and with respect to all arithmetical propositions. He’s not disposed to make many erroneous arithmetical judgments, unlike Peter. But this just goes to illustrate that there must be other necessary conditions on intuitive arithmetical warrant other than NA. As I’ll argue for below, arithmetical intuitive warrant requires a more nuanced *positive* reliability condition.

What is the relevant field with respect to which an agent must be positively reliable in order to have arithmetical intuitive warrant? Reflecting on particular cases of arithmetical intuitive warrant suggests a reasonable answer to this question. But before reflecting on these cases, I’ll present an important concept for understanding positive reliability for arithmetical intuition: *grounding structures*. 
A grounding structure for a class of claims U is a set of principles that serve as *truth makers* for the following two kinds of facts:

i. The necessary truths in U.

ii. The claims of U standing in particular relationships of metaphysical complexity, similarity, and ordering.

So, take a class comprised of all the arithmetical claims. It’s reasonable to think that the peano axioms of arithmetic comprise at least part of the grounding structure of all arithmetical claims—perhaps along with the principles specifying the workings of the various arithmetical operators (+, -, \, \times, etc.). The complete grounding structure of arithmetic grounds all the necessary truths of arithmetic, e.g., 2+2=4, 5x3=15, the truth of GB (if it’s true), etc. But it’s also the case that the arithmetical grounding structure grounds important ordering relations among arithmetical properties. For instance, the number 34 is greater than the number 12. This is grounded by the peano axioms and the definitions of 34 and 12. Regarding relations of metaphysical similarity and difference, the grounding structure of arithmetic also grounds the fact that the property of *summing to 34* is metaphysically more similar to the property of *summing to 33* than it is to the property *having the product of 4,571*.

It also seems as if the claim

4>2

is metaphysically more complex than the claim

1>0

Given the grounding structure for arithmetic, 4>2 is constituted by many more basic arithmetical properties than the latter claim is. This is because the numbers 4 and 2 are built up out of more iterations of applying the successor principle (from the peano axioms) than 1
is.\textsuperscript{23} Arithmetical claims stand in these interesting relationships in virtue of their sharing the same grounding structure.

I want to return to the idea that all arithmetical (necessary) truths are derived from and grounded in the grounding structure for arithmetic. This seems right. Take any somewhat complex arithmetical claim like \( \frac{450}{25} = \frac{36}{2} \). There are many ways to derive \( q \) by simply starting with the definitions of the numbers and operators involved, and the peano axioms, and using arithmetical claims of lesser metaphysical complexity than \( q \) to \textit{work up} to the truth of \( q \). Such a chain of derivation steps is what I call a \textit{grounding derivation}.

Regarding the relevant field for the positive reliability condition, one important piece of data that readily suggests itself is that there’s a crucial dis-analogy between what a member of species Z can do, and what Wes can do. A species Z member has an ability such that were he to form intuitive judgments on arithmetical claims \textit{of equal or lesser} metaphysical complexity than GB, he’d probably get them right. Wes lacks this disposition. Because of this, a species Z member has the ability to intuit, correctly, all the steps that would be needed to formulate a \textit{complete} grounding derivation of GB down to the grounding structure of arithmetic. Again, Wes certainly lacks this ability.

Consider a more common example of arithmetical intuition. Small children in elementary school can certainly get to the point, with enough practice, where they can competently intuit (and come to warrantedly believe) truths like \( w: (45/10)<(39/5) \). But in order to warrantedly intuit a claim like \( w \), it’s not the case that these children must be disposed to reliably intuit

\[ \left( \frac{4 \times 4}{2 \times 2} \right) > \left( \frac{4}{2} \right) \]
\[ 4 \geq 2 \]

The former claim is also more complex than the latter because it’s constituted by more arithmetical operators than the latter.

\textsuperscript{23} Also, consider the following arithmetical claims:

\[ \left( \frac{4 \times 4}{2 \times 2} \right) > \left( \frac{4}{2} \right) \]
\[ 4 \geq 2 \]
even more complex arithmetical claims, like \((394/27)<(278/19)\). Once again, the sorts of children (and people in general) that are able to warrantly intuit claims like \(w\) are constituted such that, were they to make intuitive judgments on arithmetical claims of equal or lesser complexity as \(w\), they’d most likely judge correctly. Indeed, these individuals have the ability to intuit the steps of a grounding derivation for \(w\).

This suggests that there’s the following positive reliability requirement for arithmetical intuition.

\[
\text{PA} \quad \text{Necessarily, If S has warrant for her intuitive judgment whether } p, \text{ where } p \text{ is an arithmetical claim, then S is positively reliable in possible intuitive belief-forming events, within modal distance } D \text{ from the token, in which S makes judgments on claims present in a field of the arithmetical propositions that are of equal or lesser metaphysical complexity as } p.
\]

I’ll note here, that PA, as a necessary condition on arithmetical intuitive warrant, elegantly fits with the sort of phenomenology that’s constitutive of having an intuitive seeming that \(p\). Intuitive phenomenology is \textit{impulsional} in nature.\(^{24}\) Having an intuitive seeming that \(p\) doesn’t just communicate \(p\)’s truth to the subject. Moreover, it seems to communicate that \(p\) \textit{must} be true to the subject. While this next comment is somewhat speculative, upon introspection, the relevant sense of “must” here seems to have the following character: \(p\) must be the case, given the kind of claim \(p\) is. Where \(p\) is an arithmetical claim, the phenomenology of an intuitive seeming that \(p\) communicates that \(p \textit{must be true}, given the nature of arithmetic.\(^{25}\) But this is just a reference to \(p\)’s being derived from, and grounded in, the grounding structure of

\(^{24}\) See Plantinga (1993: 192-193) for a discussion of the “impulsional evidence” that seems to accompany the use of intuition in a priori belief formation.

\(^{25}\) It’s beyond the scope of this chapter to argue it here, but I do think that introspection reveals the character of this “must” to be different than the “must” of mere metaphysical necessity. It seems to be much more content/domain specific. Also, I’m not claiming that this specific sense of “must” is always transparent to the subject having an intuitive seeming. Often times the phenomenology of intuition is fleeting, and our grasp of the content associated with intuitive phenomenology is implicit rather than explicit.
arithmetic. This is why requiring a subject to be in a position to intuit the grounding derivation for some arithmetical claim p seems to fit nicely with the nature of intuitive phenomenology. Seeing as how PA articulates just such a requirement, I take this elegant fit with intuitive phenomenology to be a strong point in PA’s favor.\textsuperscript{26}

Regarding PA, what is the relevant sense of positive reliability? Min or max? There’s good reason to believe that PA only invokes min positive reliability. Consider the case of Caroline who uses rational intuition to form the belief that q: (475/20)>(312/19). Suppose that Caroline is disposed such that, were she to consider almost any arithmetical claim of equal or lesser metaphysical complexity to q, she’d reliably come to form the correct intuitive judgment. I say almost here, because Caroline has one odd cognitive glitch that gives her one small pocket of unreliability. With respect to the claim r, (59/4)>(55/4), Caroline is reliably disposed to

\textsuperscript{26} Here, some might see some similarities between my account and Peter Markie’s (2013) account of intuitive knowledge. Markie defends a view he sees as more “traditional” and in the spirit of Leibniz and Descartes. He writes,

Let us say that some of the claims we rationally intuit are “first principles;” these propositions are so simple it is impossible for an inquirer to understand them without rationally intuiting them. All our other rational intuitions are of claims we deduce, with no reliance on memory, from currently intuited first-principles…This return to the past has significant drawbacks. It severely restricts the scope of rational intuition to first principles, which must be intuited if understood, and what can be deduced from them with no reliance on memory. (287-8)

Here, by “first principles,” he means something very similar to the basic grounding structural claims of a particular domain. I agree with Markie that his account does have significant drawbacks, not the least of which is the commitment to the claim that one can only intuit (non-inferentially) the basic grounding structure principles of a domain. This doesn’t seem to psychologically capture how intuition works in normal everyday life. For instance, when I use intuition to see that (11/2)>(9/2), it doesn’t feel like I’m performing a discursive inference. But even if I am performing some sort of quick and short inference, it’s even more unlikely that I’m performing an inference that invokes—as steps—the basic grounding structure facts, like the nature of the successor relation. Secondly, Markie invokes some notion of deduction without reliance on memory in his account of intuition. Here I’ll just raise the concern that I can’t see any way to make sense of this notion. Deduction (and any inference for that matter) is not an instantaneous mental event, but rather something that’s diachronic for humans. But if this is the case, then it’s hard to see how there could be competent inference without the use of memory. Thirdly, even if Markie is right and arithmetical intuition is necessarily much more inferential than I’m describing it, it’s still the case that a complete theory of intuitional warrant needs to posit relevance principles and reliability necessary conditions. This is because, for each warrant conferring inference one makes, it must be the case that either the agent is reliably disposed to infer the correct conclusions from certain premises, or that the agent can reliably intuit conditional premises linking the premises with the conclusion. In either case, the agent needs to have some sort of reliability in making the right sorts of connections between premises and conclusions, and I think this reliability could be modelled in a way that’s analogous to NA and PA (but just directed towards the conditional steps linking premises and conclusions in an inference.)
form the intuitive judgment that \( \sim r \) were she to consider whether \( r \) is true. Notice, \( r \) is just one claim within the vast class of propositions that are of equal or lesser metaphysical complexity to \( q \). Such a small pocket of unreliability doesn’t seem to undermine Caroline’s warrant for believing \( q \). But such a pocket would undermine warrant if we interpret PA in terms of \( \max \) positive reliability. It’s clear that we can multiply cases like Caroline’s, making it highly reasonable that PA only invokes min-positive reliability.

Here, I return briefly to explain why NA requires negative reliability with respect to all arithmetical claims, while PA only requires positive reliability with respect to claims with metaphysical complexity to the same degree as the target proposition. Consider once again Peter’s odd case. If we altered the case to stipulate that Peter’s negative unreliability only applied to claims more metaphysically complex than \( 24/8=3 \), but still held fixed that 55% of all the arithmetical intuitive judgments Peter makes day to day are false, it still seems as if Peter’s intuitive judgments that \( 24/8=3 \) fall short of warrant. The problem with Peter is that his intuition faculty—characterized by the unique phenomenology accompanying instances of intuitive belief formation—leads him astray too much of the time. A negative reliability condition like NA simply requires that agents aren’t disposed to use faculties that consistently (in nearby possibilities) cause them to make false judgments when used, regardless of the specific propositions the faculty is delivering judgments on.

NA and PA are correct in how they don’t require an agent to be negatively or positively reliable with respect to intuiting other sorts of a priori necessary truths, e.g., ethical truths, ontological truths, action-theory truths, epistemological truths, etc. Consider the example of a brilliant mathematician who’s also terrible in philosophical ethics class. His intuitive unreliability (either negative or positive) with respect to ethical truths doesn’t seem to undermine the reliability (and warrant-conferring nature) of his mathematical intuitive processes.
Presumably, other kinds of necessary truths, like ethical truths and epistemological truths, have their own respective grounding structures that are distinct from the arithmetical grounding structure. While it’s beyond the scope of this chapter to fully argue this point, it’s reasonable to hold that having intuitive warrant for these other kinds of necessary truths will require both positive and negative reliability in an analogous way to NA and PA, but relativized to their own respective grounding structures.

Employing the positive and negative reliability distinction for helping us understand arithmetical intuition is particularly probative in contrasting intuitive warrant with how warrant works for other sorts of non-inferential belief formation. Consider non-inferential color-vision belief formation. In this chapter, I won’t offer a complete set of relevance principles for color vision warrant. But, I’ll at least flag that it’s quite reasonable to think that there’s some negative reliability requirement for color vision that holds across some field of color propositions (of the form, *object x is color y*) that contains more color claims than the target proposition itself. For example, suppose someone is reliably disposed to always make correct color judgments for the particular color orange, but also disposed to misrepresent, and thus, falsely believe, color propositions ascribing *any other* color to various objects. This person is constantly making false visual color judgments, except when she happens to be looking at something that’s orange. It seems like such a person, at the very least, cannot come to know that some object x is orange, in virtue of being negatively unreliable with respect to some larger domain of color propositions that goes beyond ascriptions of orange. However, unlike arithmetical intuition, there doesn’t seem to be some interesting positive reliability requirement on color vision warrant corresponding to a field that’s greater than the target color proposition itself. Consider this case:
Jessica lives in a world in which the vast majority of objects are orange. Although, in her environment, there is a very small population of creatures that are either brown, purple, or blue. However, these creatures often stay remarkably well hidden, and barely ever show themselves to Jessica (or anyone else for that matter). Every now and then, (and it is quite rare), Jessica spots one of them in her visual field. Jessica has a normally functioning color perception faculty except for one slight cognitive glitch: everything that’s actually blue, brown, or purple gets represented to her as orange. Given that this is the case, she’ll become disposed to form incorrect judgments about the color of these creatures when she sees them, thinking that they too are orange on the basis of her visual experiences. Once again, she barely ever comes across these creatures, so the vast majority of what she sees is in fact orange.

It seems as if Jessica can still come to know that an object is orange when she’s looking at an orange object in her world. This holds, even though it’s the case that in the majority of nearby (within D) color visual belief-forming events where she judges the color of objects that aren’t orange, she judges incorrectly. For vision, she still has negative reliability (and positive reliability) with respect the field that only contains the color orange, and she’s also negatively reliable with respect to the field of all colors—given that the vast majority of the possible visual color belief-forming events within D involve her correctly seeing orange objects to be orange. Given that she has no defeaters for her color beliefs that correctly ascribe orange to some object, there’s good reason to believe Jessica has warrant for her color beliefs ascribing orange, even though she lacks positive reliability for a field of color propositions that ascribe colors other than orange.

Color vision’s having a much more scaled down positive reliability requirement constitutes a big difference between color vision warrant and arithmetical intuition warrant. For future research, it would be interesting to consider whether this difference between arithmetical intuitive warrant and color vision generalizes in way that characterizes a difference between all a priori and a posteriori belief formation. If this were the case, then the positive
reliability/negative reliability distinction is particularly probative for building an explanatorily deep theory of warrant in general.

5.5 Conclusion

If NA and PA are correct, then it follows that different kinds of reliability measurements within multi-proposition fields are constitutive of the property of warrant. In other words, these different reliability measurements (positive and negative) are central to understanding the nature of warrant itself. It’s also the case that, for a given token of arithmetical intuition, there are multiple reference classes that are relevant for determining whether an agent has warrant. As I discuss above, this result follows simply from the nature of negative or positive reliability within a multi-proposition field as distinguished from positive reliability with respect to just the target proposition. Hence, relevance theories which reject multi-type relevance are mistaken. The important upshot here is that philosophers responding to the generality problem should develop relevance principles and overall relevance theories that invoke the two different kinds of reliability measurements within a field (positive and negative), and grant the relevance of multiple types for a given token.

For future exploration, as I suggest above, philosophers should examine whether the NA or PA results carry over into relevance principles for other sorts of belief formation. As I discuss above, the generality problem, in many ways, amounts to a demand for more explanation regarding the nature of warrant from a reliabilist perspective. Our understanding of warrant certainly would increase insofar as more general relevance principles—which explain the features of relevant types that hold across many kinds of belief formation—are presented and defended. These general principles could take the form of answering the following questions: What are the features that characterize the relevant fields for all sorts of a posteriori
belief formation—for either positive or negative reliability? What characterizes the relevant fields for all sorts of a priori belief formation—for either positive or negative reliability? As I suggested in my discussion of color vision and arithmetical intuition above, if there are important differences between positive reliability requirements for a priori belief formation and a posteriori belief formation, it would be great to develop some general explanation for why this might be the case. In addition to these, one might wonder what sorts of positive and negative reliability requirements hold for inferential warrant. These are all fruitful and important questions to answer in the further development of a complete relevance theory.
BIBLIOGRAPHY


— (2010), Achieving Knowledge. Cambridge: CUP.


Lackey, Jennifer (2008), Learning from Words. Oxford: OUP.


Roeber, Blake (MS), “Anti-Intellectualism.”


