Rational Understanding: Toward a Probabilistic Epistemology of Acceptability

Finnur Dellsén
University of Iceland
Inland Norway University of Applied Sciences

Forthcoming in *Synthese* (special issue: "Themes from Elgin");
please cite published version when available

Abstract

To understand something involves some sort of commitment to a set of propositions comprising an account of the understood phenomenon. Some take this commitment to be a species of belief; others, such as Elgin and I, take it to be a kind of cognitive policy. This paper takes a step back from debates about the nature of understanding and asks when this commitment involved in understanding is epistemically appropriate, or 'acceptable' in Elgin's terminology. In particular, appealing to lessons from the lottery and preface paradoxes, it is argued that this type of commitment is sometimes acceptable even when it would be rational to assign arbitrarily low probabilities to the relevant propositions. This strongly suggests that the relevant type of commitment is sometimes acceptable in the absence of epistemic justification for belief, which in turn implies that understanding does not require justification in the traditional sense. The paper goes on to develop a new probabilistic model of acceptability, based on the idea that the maximally informative accounts of the understood phenomenon should be optimally probable. Interestingly, this probabilistic model ends up being similar in important ways to Elgin's proposal to analyze the acceptability of such commitments in terms of 'reflective equilibrium'.

1 Introduction

Barbara is a historian who just finished a research project on European emigration to North America in the late 19th and early 20th century. She now understands why Europeans came to America in such numbers, why their arrivals peaked when it did, why the settled down in the specific locations they did, and so forth. To gain

understanding of these facts, Barbara has had to learn a great deal about the harsh living conditions in Europe at the time, the perceived promise of moving over the Atlantic, the ways in which European immigrants were encouraged to settle down in some places rather than others, and so forth. So Barbara's understanding is based on a number of propositions to which she is in some sense 'committed' for the purposes of understanding.[1] For example, if understanding involves grasping or otherwise representing an explanation of the understood phenomenon (a common view that I will not contest here),[2] then Barbara must in some sense *commit* (or, if you prefer, *appeal*)[3] to the propositions that jointly constitute the explanans of the explanations that Barbara grasps or represents.

What, exactly, is the nature of this commitment involved in understanding? A standard view identifies it with belief – or, at least, a species of belief (e.g. a belief about what would cause or explain the phenomenon). This type of view is not only accepted by those who take understanding to be a species of propositional knowledge (Grimm, 2006, 2014; Khalifa, 2013b, 2017; Greco, 2014; Kelp, 2015, 2017; Sliwa, 2015), but also by some who explicitly reject a knowledge-based analysis of understanding (Hills, 2016; Pritchard, 2009; Lawler, 2016). By contrast, Elgin (2004, 2017) and I (Dellsén, 2017, 2018b) have argued that understanding instead involves a kind of acceptance in Cohen's sense of the term, i.e. having a policy of treating the propositions in question as given in relevant contexts (Cohen, 1992, 4-5).[4] For Elgin, the move from belief to acceptance is necessary to account for the role of idealizations in understanding, as when one understands why heating a closed container increases the pressure inside by appealing to the ideal gas law, since such idealizations are clearly not believed in an ordinary sense of the term.[5] By contrast, my motivation is to allow for understanding agents to be skeptical towards the often speculative

[1]This is true for 'objectual' as well as 'explanatory' understanding, so my discussion in this paper applies to both kinds of understanding – if indeed they really are distinct (see Khalifa, 2013a).

[2]Although see Dellsén (2018a) and Wilkenfeld (2018).

[3]I will use 'commit' in what follows, although this can be replaced with 'appeal' throughout for those who prefer that terminology.

[4]Of course, to say that understanding involves acceptance rather than belief is not to say that belief could not play any role at all in how or why an agent comes to understand something. Rather, it means that belief is not *necessary* for understanding – that one *could* understand something without believing the propositions on which the understanding is based.

[5]My own view is that this is not a convincing reason to move from belief to acceptance, since idealizations function as a tool for gaining understanding rather than as part of the content of what is committed to in understanding something. See Lawler (2018) for a proposal along these lines.

propositions on the basis of which they understand, as when a string theorist has a low degree of confidence that string theory will turn out to be true (Dellsén, 2017).[6]

The arguments and results to be discussed in this paper will be relevant to this issue of what type of commitment is involved in understanding,[7] but my more immediate goal is to argue for a model of when such a commitment is epistemically appropriate, or 'acceptable' in Elgin's terminology (Elgin, 2017, 63-90). I first argue that the commitment involved in understanding is sometimes acceptable even when it would not be rational to assign a probability above any given threshold to the relevant propositions. My argument will appeal to the idea that acceptable commitments of the type involved in understanding must be logically coherent, i.e. deductively consistent and closed under deductive consequence (§2). This requirement, I will argue, shows that there can be no justification requirement of the traditional sort on understanding, roughly since such a justification requirement would have to be compatible with it being rational to assign an arbitrarily low degree of belief in justified propositions (§3). This leads me to present a model of acceptable commitments that replaces the 'satisficing' probability requirement that is implicitly at work in competing theories with an 'optimizing' requirement. Very roughly, my suggestion is that acceptable commitments are those that follow from the *most probable* account of the understood phenomenon (§4). I go on to prove that this model validates the idea that understanding should be logically coherent in the relevant sense (§5). I conclude by briefly comparing this model with Elgin's theory of acceptability in terms of 'reflective equilibrium', and indicate how Elgin and I have arrived at similar and even complementary conclusions by different routes (§6).

2 Understanding and Logical Coherence

This section introduces and precisifies the idea that logical coherence is a constraint on when the commitment involved in understanding is epistemically appropriate. But I start with a couple of preliminary points. As I have indicated, I am following Elgin in using the term 'acceptable' as a term of epistemic appraisal for commitments of the kind involved in understanding. This should not be taken to imply that we have already settled on an answer to the question of whether understanding involves

[6] See also Wilkenfeld (2016) and Baumberger (2018) for arguments and endorsements of the idea that understanding does not involve belief.

[7] See footnote 21 below.

acceptance as opposed to belief, for that would beg one of the questions that I take my arguments below to have bearing on. Rather, I will use 'acceptable' as a term of epistemic appraisal of the commitment involved in understanding, whatever the nature of this commitment turns out to be (i.e. whether it is a kind of belief, or acceptance, or something else entirely). When I think the reader might need to be reminded of the neutral sense in which I intend to be using the term, I will instead refer to such commitments as 'epistemically appropriate'.[8]

Another preliminary point concerns the role of idealizations and approximations in understanding, especially in science.[9] Oftentimes, the road to understanding in science is paved with felicitous falsehoods – claims we know or have every reason to suspect are false if interpreted literally, but that somehow provide us with understanding nonetheless. In this paper, I will take for granted that felicitous falsehoods of this kind are never strictly speaking acceptable.[10] Instead, what is strictly speaking acceptable is some suitably qualified statement, e.g. that the friction of a given plane is sufficiently small in comparison to other forces acting on the object in question to have negligible effects on the object's movement. Of course, exactly which qualified statement will count as acceptable for each felicitous falsehood will vary depending on the nature of the felicitous falsehood in question, so I doubt there is much to say in general about how to qualify statements in this way to convert a felicitous falsehood into something that could be acceptable. Nevertheless, I see no reason to think this conversion strategy would fail in any plausible case of idealizations or approximations that we would be inclined to associate with genuine understanding. After all, as Bokulich (2012) points out, there must be some way for scientists to translate fictitious elements of their theories into something that could be taken as true,[11] and a qualified version of a claim containing felicitous falsehoods

[8]Another terminological option would be to use terms that are traditionally associated with epistemically appropriate beliefs, e.g. 'rational', 'reasonable' or 'warranted', but this would have the opposite effect of misleadingly indicating that the commitment involved in understanding is assumed to be a species of belief.

[9]In idealizations, there is some claim or part of a claim that is radically false, as when we suppose that a population is infinite or that an object is moving on a frictionless plane. In approximations, there is some claim or part of a claim that is 'close to the truth', as when we round up or down to the nearest integer for a given quantity or suppose that a nearly spherical object (such as a planet) is in fact a perfect sphere.

[10]This is not the place to argue for that claim, although see Lawler (2018) for an argument that I endorse (similar arguments are made by, e.g., Mizrahi, 2012; Strevens, 2017; Sullivan and Khalifa, 2019).

[11]Bokulich (2012, 735) calls this a 'translation key' (see also Frigg and Nguyen, 2016, 228).

would simply be making this explicit.[12]

With these preliminary points in place, let us consider what sort of epistemic constraints there might be on the commitment involved in understanding. Several authors, including Elgin (2007, 2009, 2017), have suggested that understanding requires some type of *coherence* among the propositions to which one is committed (see also, e.g., Kvanvig, 2003; Riggs, 2009; Carter and Gordon, 2014; Gijsbers, 2015).[13] Now, 'coherence' is notoriously hard to define or measure with any precision, in part because it seems to involve defining and measuring probabilistic and explanatory relations between propositions (see, e.g., Shogenji, 1999, 2001; Akiba, 2000; Fitelson, 2003; Olsson, 2005; Schupbach, 2011). However, there is a quite modest notion of coherence – *logical coherence* – that can simply be characterized in terms of whether a body of propositions obeys ordinary deductive logic. Plausibly, any stronger notion of coherence will need to include logical coherence as a necessary but insufficient requirement.[14] My contention in this section is that the commitments involved in understanding should be coherent in this modest sense, i.e. logically coherent.

To flesh out this requirement of logical coherence for understanding, let us return to our historian. Take any two of the propositions Barbara is committed to in understanding various aspects of European emigration to America; call them P_1 and P_2. Supposing that Barbara's commitments to P_1 and P_2 are acceptable (i.e. epistemically appropriate), would it also be acceptable for her to commit to their conjunction, $(P_1 \wedge P_2)$, in understanding the same phenomenon? I think it's fairly clear that a plausible account of the normative requirements on understanding should say that it would be. To see why, consider what is supposed to distinguish understanding from other ways of representing the world, such as knowledge or true belief. As Jonathan Kvanvig notes, "what is distinctive about understanding has to do with the way in which an individual combines pieces of information into a unified body" (Kvanvig, 2003, 197).[15] So, when you understand a phenomenon, you do not merely know an

[12]Indeed, if it were impossible to qualify a given idealization/approximation in this way – if there was no 'translation key' which could be used to convert the statement into something that isn't known to be false – then the falsehood would not really be felicitous in the first place; it would just be an ordinary (known) falsehood (Bokulich, 2012, 731-736).

[13]For dissent, see Khalifa (2016).

[14]This seems to be in line with Elgin (2017), who explicitly requires logical consistency as a minimal condition (Elgin, 2017, 71), and also seems to endorse a conjunction rule for understanding (Elgin, 2017, 63-4).

[15]Similar views are expressed by Cooper (1994), Zagzebski (2001), Pritchard (2009), Grimm (2011), Bengson (2015), and of course Elgin (2009, 2017).

assorted series of facts about it; rather, you somehow grasp how different pieces of information about the phenomenon are related, or perhaps how those pieces of information are related to other things, such as its causes or effects. Thus it seems that one's understanding would be defective, or perhaps fail to qualify as understanding at all, if one were committed to two separate propositions about some phenomenon but not to their conjunction. Doing so would involve a kind of compartmentalization of information that is at odds with what understanding is supposed to be, and thus could hardly be the epistemically appropriate course of action in situations like that of Barbara.

To really see the force of this argument we can make the example of Barbara a bit more concrete. Let L_N and L_S describe the harsh living conditions in Northern and Southern Europe respectively, and suppose committing to each one is acceptable. Now, on the compartmentalization picture that I think we should reject, it could turn out not to be acceptable to commit to $(L_N \wedge L_S)$, which simply describes the harsh living conditions in Europe generally. But since what Barbara is seeking to understand is precisely why Europeans generally emigrated to America (rather than why Northern or Southern Europeans emigrated specifically), it seems that $(L_N \wedge L_S)$ could only provide a greater understanding of the relevant phenomenon that each of L_N and L_S would do separately. Indeed, it is doubtful that merely committing to L_N and L_S without also committing to their conjunction – assuming this is even possible – can be said to provide any understanding at all of why Europeans generally emigrated to America, since someone who accepts L_N and L_S separately but not their conjunction would seem to be missing a crucial component of understanding why Europeans generally – as opposed to Southern or Northern Europeans – emigrated to America.

The upshot of these considerations is that it seems that the commitment involved in understanding is normatively closed under conjunction: if it would be acceptable to be committed to two separate propositions P_1 and P_2 in understanding some

[16]I am not quite suggesting that any two acceptable commitments can be 'conjoined' into an acceptable conjunctive commitment; rather, I am suggesting that this type of conjunction principle holds when the commitments in question are made in the context of understanding the very same phenomenon. To see why this matters, note that there might very well be cases in which it would be acceptable for an agent to commit to P_1 in understanding X_1, and also acceptable for the agent to commit to P_2 in understanding X_2, but it wouldn't be acceptable to commit to $(P_1 \wedge P_2)$ in understanding anything. For example, perhaps it would be acceptable for current physicists to commit to general relativity in understanding gravitational lensing, and to quantum mechanics in understanding the emission spectrum of hydrogen, but not acceptable to commit to the conjunction of general relativity and quantum mechanics, e.g. because the two theories are seemingly inconsistent.

phenomenon X, then it would be acceptable to be committed to their conjunction $(P_1 \wedge P_2)$ in understanding X.[16] This principle quickly generalizes from two to any finite number of propositions n, since the conjunction $(P_1 \wedge ... \wedge P_n)$ can be obtained by iterated applications of this two-proposition principle.[17] More precisely:

($\wedge$) If it would be acceptable to commit to each proposition $P_1, ..., P_n$ in understanding some phenomenon X, then it also would be acceptable to commit to their conjunction $(P_1 \wedge ... \wedge P_n)$ in understanding X.

There is a natural but non-trivial way to generalize even this principle. The conjunction $(P_1 \wedge ... \wedge P_n)$ is the strongest possible logical consequence of $\{P_1, ..., P_n\}$ – any other logical consequence of $\{P_1, ..., P_n\}$ is strictly weaker than $(P_1 \wedge ... \wedge P_n)$, in the sense that the latter entails it but is not itself entailed by it. Thus, if it is acceptable to commit to $(P_1 \wedge ... \wedge P_n)$ then it is surely also acceptable to commit to any other logical consequence of $\{P_1, ..., P_n\}$. We thus generalize ($\wedge$) as follows:

($\Rightarrow$) If it would be acceptable to commit to each proposition $P_1, ..., P_n$ in understanding some phenomenon X, then it would also be acceptable to commit to any deductive consequence of $\{P_1, ..., P_n\}$ in understanding X.

To be sure, while ($\wedge$) follows from ($\Rightarrow$), the converse does not hold (since not all logical consequences of some set of hypotheses are conjunctions of those hypotheses). Nevertheless, I think the generalization is well motivated, and I'll assume it holds in what follows.[18]

Now, suppose a set of propositions $\{P_1, ..., P_n\}$ is inconsistent. Could it be acceptable to commit to all the propositions in such a set? Well, given ($\Rightarrow$) – or, indeed, given only ($\wedge$) – this entails that it would be acceptable to commit to a contradictory proposition. By deductive explosion, it would then be acceptable to commit to any proposition whatsoever. What we have said so far does not rule this out as unacceptable, but it's fair to say that a model of when the commitment involved in understanding is acceptable ought to rule it out. We therefore add that it cannot be acceptable to commit to an inconsistent set of propositions in understanding something:

---

[17] That is, we can use mathematical induction in which the inductive step consists in using the two-proposition principle to obtain that it would be acceptable to commit to $(P_1 \wedge ... \wedge P_{i+1})$ if it would be acceptable to commit to $(P_1 \wedge ... \wedge P_i)$ and P_{i+1} respectively.

[18] However, not much in what follows turns on making this generalization however. In particular, the arguments for separating acceptability and justification in the next section use only the restricted principle ($\wedge$) and not also the more general ($\Rightarrow$).

($\neg\bot$) If it would be acceptable to commit to each proposition $P_1, ..., P_n$ in understanding some phenomenon X, then $\{P_1, ..., P_n\}$ is deductively consistent.

Note that given ($\Rightarrow$), this requirement is equivalent to the claim that it would not be acceptable to commit to a contradiction.

The three principles we have now discussed – ($\wedge$), ($\Rightarrow$), and ($\neg\bot$) – constitute a way of spelling out the idea that understanding should be logically coherent. For convenience, we can state these principles as a single requirement as follows:

Deductive Cogency for Understanding: The set of propositions to which it would be acceptable to commit in understanding some phenomenon is deductively consistent and closed under deductive consequence.

There is no doubt more to be said about how this requirement should best be understood, defended, and distinguished from similar requirements on belief (see Dellsén, 2018b). In this paper, however, I am concerned with exploring the philosophical consequences of adopting it; what, in particular, does it imply for the epistemology of understanding? As we shall see in the next section, this requirement plausibly rules out the idea that understanding requires epistemic justification in any traditional sense of the term. Furthermore, as I go on to show, the requirement also motivates a particular 'optimizing' model of the conditions under which commitments involved in understanding are acceptable.

3 Enter the Paradoxes

As I have noted, a common view of the commitment involved in understanding identifies it with belief or some species thereof. If this view is correct, it would seem to follow that the commitment involved in understanding is acceptable – i.e., epistemically appropriate – only when such a belief is epistemically justified. Indeed, many authors explicitly endorse a justification requirement on understanding in addition to identifying the relevant commitment with belief (e.g., Pritchard, 2009; Grimm, 2006, 2014; Khalifa, 2013b, 2017; Greco, 2014; Kelp, 2015, 2017).[19] In this section, I argue that such justification requirements on understanding conflict with the idea that understanding must be logically coherent, in the sense spelled out in the previous section. This conflict can be brought to light by suitably modified versions

[19] Although see Hills (2016) and Dellsén (2017) for arguments against such justification requirements on understanding.

of the lottery and preface paradoxes (Kyburg, 1961; Makinson, 1965), which will be illustrated here by extending our example of Barbara the historian.[20]

For a version of the preface paradox, suppose Barbara wrote a thick and informative book about her research into European emigration to North America. In the book, Barbara appeals to various distinct claims, each of which she takes to be well supported by the data that she has gathered in her research. Thus let us suppose that Barbara would be justified in believing each of these claims. Now consider the conjunction of all these claims in the book. If Barbara's book contains sufficiently many distinct claims, it seems that Barbara would not be justified in believing the conjunction. After all, it is overwhelmingly probable that Barbara has made at least one error somewhere in her book, so the probability that such a conjunction is true would be miniscule. (This is not just 'intuitively' so; it follows from the probability axioms that the probability of a conjunction of non-entailing claims decreases for each added conjunct. So for a sufficiently lengthy book filled with at least somewhat independent claims, the conjunction of the claims in the book will be as improbable as you like.) The upshot is that Barbara would be justified in believing each of the claims in her book while she would not be justified in believing their conjunction. By the same token, since the conjunction of any set of claims is a logical consequence of those claims, we have that Barbara would be justified in believing a set of claims while she would not be justified in believing one of their logical consequences.

In order to extend this to a version of the lottery paradox, let us suppose that Barbara is also justified in believing the negation of the conjunction of the claims in her book, i.e. that at least one of the claims in her book is false. This might be because Barbara realizes how unlikely it is that this conjunction is true, from which she infers that the negation is very probably true (sufficiently so for her to be justified in believing it). Alternatively, we can imagine that a highly trustworthy colleague of Barbara's told her that the book contains an error, although the colleague refuses to reveal which claim is in error. In any case, if Barbara's book contains sufficiently many distinct claims, we can suppose as before that Barbara still has very good reasons to think that each individual claim she made in her book is true – reasons

[20] Although I use the traditional terms for the two paradoxes, the cases that I use to illustrate the two paradoxes will not mention either lotteries or prefaces. In my view, these features of Kyburg's and Makinson's original examples distract from the relevant logical points about the relationship between probability and justified belief. For example, it seriously misleading in my view to focus on the fact that the lottery paradox involves making a 'statistical inference' (Nelkin, 2000), since the logical structure of the lottery paradox can be manifested by bodies of propositions that involve no statistical information at all (see my version of the case below).

that make each such claim extremely probable by Barbara's lights (even though their conjunction is improbable). And so it seems clear that Barbara would be justified in believing each such claim. But then Barbara would be justified in believing a set of claims that are jointly inconsistent, viz. the set consisting of each of the claims made in her book in addition to the negation of the conjunction of those claims.

These cases illustrate that the set of propositions that one is justified in believing need not be logically coherent. In particular, the first (preface-style) case illustrates that the set of propositions one is justified in believing need not be closed under deductive consequences, while the second (lottery-style) case illustrates that this set need not be consistent. By contrast, I have argued in the previous section that these same logical coherence requirements do apply to what it would be acceptable to commit to in understanding something, roughly because it is a distinctive feature of what it is to understand a phenomenon that one's representation of the phenomenon not be compartmentalized or contradictory. For example, although Barbara would not necessarily be justified in believing a particular conjunction of claims in her book, it could well be acceptable for her to commit to that conjunction in the context of understanding something (assuming it was acceptable for her to commit to each conjunct), roughly since the conjunction might provide a deeper understanding than each individual conjunct. From this it immediately follows that the conditions for acceptable commitment of the sort involved in understanding are not identical to the conditions for justified belief, i.e. that acceptability is not justification. So understanding, even when epistemically appropriate, does not require justification.[21]

I see two ways of resisting this argument. First, it might be objected that these arguments rely on a conception of justified belief that some would deny, viz. the 'Lockean' thesis that belief in a proposition is justified just in case it is rational to have credence in the proposition that exceeds a given threshold (Foley, 1992, 2009). There is some truth to this allegation. The first (preface-style) argument relies on the premise that Barbara would not be justified in believing something that is exceedingly improbably from her own rational point of view – or, more precisely, that there is some threshold t_a such that Barbara would not be justified in believing

[21]I note in passing that this argument also puts pressure on the idea that understanding requires belief, i.e. that the commitment involved in understanding is belief or a species thereof. For if the commitment involved in understanding were a species of belief, then epistemic justification would presumably at least be a necessarily condition on epistemically appropriate commitment of this kind. By contrast, if the commitment involved in understanding is not a species of belief, then we should expect it to be possible for the epistemic conditions for each type of attitude to come apart in some cases (in accordance with the current argument).

a proposition P unless it is rational for her to assign a probability to P that exceeds t_a. Similarly, the second (lottery-style) argument relies on the premise that Barbara would be justified in believing a proposition that is exceedingly probable from her own rational point of view – or, more precisely, that there is some threshold t_b such that Barbara would be justified in believing P if it is rational for her to assign a probability to P that exceeds t_b.

However, note that these arguments do not require any specific values to be assigned the thresholds t_a and t_b, nor do they require that $t_a = t_b$. So t_a can be as low a threshold as you like, while t_b can be as high as you like – provided only that both are strictly between 0 and 1. Indeed, these arguments allow that the distinction between justified and unjustified belief is to some extent vague or indeterminate. This idea could be accommodated by formalizing the probability threshold not as a number t but as an interval, $I_t \subseteq (0, 1)$, where a belief counts as justified just in case its probability exceeds the upper bound of the probability-interval I_t, not-rational just in case its rationally-assigned probability falls below the lower bound of I_t, and indeterminate otherwise. Furthermore, each threshold can as far as these arguments go be allowed to fluctuate, e.g. depending on the interests or stakes of the subject S or of those who attribute a justified belief to S.[22]

Given these qualifications, it seems to me that the version of the Lockean thesis required by the two arguments given above is exceedingly plausible. To deny the Lockean thesis in this form is to say, first, that someone could be justified in believing a proposition that is, by her own rational lights, as improbable as you like (provided its probability is not zero); and, second, that someone could fail to be justified in believing a proposition that is, by her own rational lights, as probable as you like (provided its probability is not one). It seems to me that this would do much more violence to our pretheoretical conceptions about justification and rationality than rejecting anything about the connection between understanding and justification. In any case, those who do deny this (qualified) Lockean thesis should feel free to read the argument of this section as demonstrating only that understanding does

[22]That said, the threshold t cannot depend on the probability assigned to the proposition P itself or logically related propositions, since one could then simply 'solve' the problem by appropriate stipulation of thresholds for each proposition. This is roughly what Leitgeb (2014, 2017) proposes to do in a systematic way with his 'stability theory' of belief. I lack the space here to consider Leitgeb's proposal in detail, but suffice it to say that Leitgeb's theory makes the choice of threshold depend on factors about the agent that seem to me to be entirely irrelevant to whether she would be justified in believing the propositions in question (in any recognizable sense of 'justified'), such as which partition the agent happens to consider (see Leitgeb, 2014, 152-160).

not require justification *in this Lockean sense*. For them, the rest of the paper can be seen as spelling out a non-Lockean conception of justification that is suitable for an explication of acceptability. For the rest of us, the remainder of this paper can instead be viewed as spelling out a model of the conditions under which the commitment involved in understanding is normatively appropriate, i.e. acceptable, in a way that separates it from epistemic justification.

The second objection I will consider rejects Deductive Cogency for Understanding because it implies that it could be acceptable to commit to an extremely improbable conjunction of propositions, provided that each conjunct is acceptable. My response to this objection is that the implication that extremely improbable conjuncts could be acceptable should be viewed as a feature rather than a bug. This is so for two related reasons. First, there are independent reasons to think that acceptability comes apart from probability in some cases, such as in cases of understanding without justification presented by Hills (2016) and me (2017) . For a case of this sort, consider a theoretical physicist who appeals to string theory (a speculative and extremely ambitious 'theory of everything') in her understanding of natural phenomena. Given the dearth of evidence in favor of string theory, it would seem rational for her to assign an extremely low probability to the theory. And yet it is hard to see why it wouldn't be acceptable for our physicist to commit to the theory in understanding aspects of the natural world. After all, string theory could be, and plausibly is, better and more probable than any competing theory that could help us understand the same range of phenomena.

A second reason to embrace the implication that extremely improbable conjunctions could be acceptable is that the contrary view assumes that there is some probability threshold for acceptability, much like the Lockean thesis takes to hold for belief. However, such a probability threshold would effectively compartmentalize understanding in an arbitrary and implausible way. To see why, suppose it would be acceptable to commit to P_1 and P_2 separately in understanding X, but that the probabilities of both P_1 and P_2 are just above the threshold t_a, while the probability of their conjunction $(P_1 \wedge P_2)$ is on or below t_a. In that case, it would not be acceptable to commit to $(P_1 \wedge P_2)$ in understanding X, no matter how unified or systematic an account this would provide of X. For example, returning to Barbara's understanding of European emigration to America, let P_1 be a proposition describing the lack of employment opportunities in Europe in the late 19th century and let P_2 be a proposition that describing the abundance of such opportunities in America at the same time. By fixing the relevant probabilities we could then construct a

case where it would be acceptable to commit to each of these propositions but not to their conjunction – even though it seems to be precisely the conjunction, rather than each conjunct, that provides the deepest understanding of the phenomenon. In sum, then, a probability threshold on acceptability would force us to keep commitments fragmented and disunified in a way that seems arbitrary and downright implausible.

Let us take stock. I have argued that understanding – or, more precisely, the commitment involved in understanding – should be consistent and closed under deductive consequence in order to be acceptable. But I have also argued that the preface and lottery paradoxes show that this is not true of justified belief, at least not on a Lockean conception on which justification is tied to rational probability assignments above some threshold. This means that an epistemology of understanding – a philosophical theory of what makes the commitments involved in understanding epistemically appropriate, i.e. acceptable – cannot simply be an application of traditional epistemology of belief. In particular, it might seem hopeless to develop a *probabilistic* model of acceptability, given how central a version of the Lockean thesis has been to the argument given in this section. In the next section, however, I show that it's possible to develop a probabilistic model of acceptance, although it will require a change in perspective regarding *how* probability works as a standard for acceptability.

4 Understanding and Probabilistic Optimality

The model that I will develop is inspired by a distinction between two structurally different ways of evaluating something with reference to some standard, e.g. as a proposition may be evaluated with reference to its probability. On the one hand, the object can be evaluated with regard to whether it is *sufficiently good* relative to the standard – e.g. as a proposition may be taken to be sufficiently probable for some purpose. On the other hand, the object may be evaluated with regard to whether it is the *best* among some set of alternatives relative to the standard, e.g. as a proposition may be taken to be the most probable among some set of competing propositions. Roughly following Simon (1956), we can say that the former is a *satisficing* evaluation, while the latter is an *optimizing* evaluation. Note that although these ways of evaluating something are clearly distinct – something can be sufficiently good without being the best among some set of alternatives, and vice versa – the standard or standards against which one is evaluating the thing in question could be identical. So, in this sense, what separates satisficing and

optimizing evaluations is the *structure*, rather than the *standards*, of the evaluation.

Inspired by this distinction, I will now articulate a model of when it is acceptable for someone to commit to a proposition in the context of understanding something. The account is *probability-based* in the sense that the standard of evaluation will be taken to be probability and probability alone. In this respect, the account resembles the aforementioned Lockean thesis regarding justified belief. However, whereas the Lockean thesis identifies justification with a satisficing evaluation – probability above a threshold – I propose that acceptability can be identified with a type of optimizing evaluation – probability greater than alternatives. To a first approximation, the idea is that it is acceptable for S to commit to a proposition P in understanding X just in case it is rational for S to assign a (significantly) higher probability to an account of X that entails P than to any alternative account of X that entails $\neg P$, where each such account is a set of propositions that would provide maximal understanding of X if its propositions were true. Call this *the Optimality Model*. The remainder of this section fleshes out this model.

Let's start by defining the type of 'accounts' that are being compared in an optimizing evaluation. Note that some propositions would not provide any understanding of a given phenomenon, even when the propositions are true and one commits to them. For example, no amount of information about the extinction of dinosaurs will help you understand the random motion of particles known as Brownian motion.[23] But now consider the various propositions that *would* provide us with understanding of a given phenomenon, provided they be true and we committed to them. These propositions can be thought of as different answers to questions about the phenomenon we seek to understand, e.g. about what caused the phenomenon, how the phenomenon is related to other things, and so forth.[24] Call such questions about the phenomenon 'understanding-seeking questions'.[25] So a question is an understanding-seeking question about some phenomenon X when committing to a correct answer to that question provides some understanding of X. Roughly, the idea then is to define

[23]A parellel point is true of explanation: Some propositions are not even potential explanantia of a given explanandum.

[24]How to finish this list will clearly depend on the specifics of one's theory of understanding, e.g. on whether one takes understanding to require explanation – an issue we leave open here.

[25]This is meant to mirror van Fraassen's (1980) framework for thinking about scientific explanations, where a potential explanation is identified with an answer to a specific kind of question, viz. an explanation-seeking why-question. Indeed, again mirroring van Fraassen, an 'understanding-seeking question' can simply be identified with its potential answers in the manner of a Hamblian account of questions Hamblin (1973).

the kind of accounts we are interested in – what I will refer to as 'noetic accounts' – as sets of maximally informative answers to understanding-seeking questions about the phenomenon in question.

To define this notion of a 'noetic account' more precisely, I first define a notion of a 'complete answer' to a single question Q:

Definition (Complete Answers). Given a set of possible answers $\{A_1, \ldots, A_n\}$ to a question Q, a *complete answer* to Q is a conjunction of n propositions in which the i-th member is either A_i or $\neg A_i$.

Thus, for $n = 2$ for example, the set of complete answers would be:

$$\{A_1 \wedge A_2, \neg A_1 \wedge A_2, A_1 \wedge \neg A_2, \neg A_1 \wedge \neg A_2\}$$

Using this notion, we then define 'noetic account' of a given phenomenon as a proposition (or, if you like, conjunction of propositions) that entails a complete answer to any understanding-seeking question about the phenomenon that we have or might have:[26]

Definition (Noetic Accounts). A proposition N_X is a *noetic account* of X iff, for any understanding-seeking question Q_i about X, N_X entails a complete answer to Q_i.

In what follows, we will use this notion to build a model of acceptability that compares competing noetic accounts relative to the same set of questions about a given phenomenon.[27]

---

[26]Here and in what follows, I mean to be deliberately non-committal regarding whether a noetic account needs to entail an answer to all *actual* understanding-seeking questions (i.e. those that we are actually interested in, or that are actually salient for us at a given time) or all *possible* understanding-seeking questions (i.e. all understanding-seeking questions in logical space). I don't know of any convincing reason to prefer either version of the definition, so I prefer not to be committed either way on this issue.

[27]One might object that in most cases when we attribute understanding to someone, the understanding agent will never have thought of many of the noetic accounts of the relevant phenomenon. After all, noetic accounts will in many cases be enormously complicated constructions – much more complicated than we can reasonably take ordinary agents to cognize – so my definition of a 'noetic account' may seem to make it impossible for ordinary agents to satisfy the demands set by the model. This objection is misplaced, for reasons that will become clearer below. For now, suffice it to say that the role of the concept of a 'noetic account' is not to be the content of the type of commitment that's involved in understanding, but rather to be a piece of theoretical machinery that determines – in conjunction with other pieces of the machinery – when such commitments are normatively appropriate. Thus while a 'noetic account' must satisfy certain normative conditions in order for an agent's commitments to be acceptable, the agent need never have actually committed to the noetic account itself in order for her commitments to be acceptable.

Now, recall that the basic idea I am attempting to spell out is that it would be acceptable to commit to a proposition just in case a noetic account that entails it is (significantly) more probable than any noetic account that entails its negation. Having spelled out what it means for something to be a 'noetic account', I will now spell out what it means for a noetic account to be *(significantly) more probable* than another. One obvious issue concerns which interpretation of probability is at play here. For my purposes, the probability of P for an agent S, $Pr_S(P)$, will be taken to be the degree of confidence, i.e. credence, that it is rational for S to assign to P.[28] (However, since it will often be obvious or unimportant whose probabilities are being referred to, I shall often write $Pr(P)$ instead of $Pr_S(P)$.) This means that I am assuming the 'probabilist' thesis that rational credences must at least be probabilistically coherent at a time, i.e. that they satisfy the standard Kolmogorov axioms of probability.[29] Of course, this is in some respects an extremely demanding requirement – something that most, if not all, ordinary agents will fail to satisfy – and should be seen as a useful idealization to help us focus on the central issue at hand. Indeed, I shall throughout make a related idealization, viz. that rational agents assign probabilities to every hypothesis in logical space. I believe that this requirement can be relaxed, but in order to keep the discussion simple and focused I shall not attempt to do so here.

Given this common and fairly uncontroversial notion of probability as rational degree of confidence, the meaning of '(significantly) more probable' would be straight-

---

A related point worth making here is that the Optimality Model is an 'externalist' theory of acceptability in the sense that it will often be difficult, and sometimes even impossible, for an agent to tell whether a given proposition is acceptable for her. That said, the agent may well make educated guesses about acceptability by a number of routes, e.g. by asking herself whether any remotely probable noetic account would entail the proposition (if not, she has some grounds for taking the proposition not to be acceptable). Note also that the Optimality Model is perfectly compatible with the idea that acceptability supervenes on the relevant agent's mental or evidential state. (Whether the model entails such a supervenience will depend on what interpretation of probability one chooses to adopt – see below.) So while the Optimality Model is 'accessibility externalist', it is at least compatible with 'mentalist internalism'. (Thanks to Christoph Baumberger for discussion of this point.)

[28]Note that, on this interpretation, the probability of P for an agent S does not represent the *actual* credence that S assigns to P. Even if 0.7, say, is the credence that it is rational for S to assign to P, S may have some credence other than 0.7 in P, or indeed no (precise) credence at all.

[29]I leave open the possibility that rational credences must conform to other norms as well, e.g. norms connecting credences to physical chances such as the Principal Principle (Lewis, 1980), norms connecting credences to symmetry considerations such as the Principle of Indifference (Laplace, 1951; Keynes, 1921) and the Maximum Entropy Principle (Jaynes, 2003; Rosenkrantz, 1977), or even norms that constrain rational credences based on explanatory considerations (Huemer, 2009; Weisberg, 2009).

forward if it were not for the parenthetical modifier, 'significantly'. That is, we could simply say that N_X^i is more probable than any alternative noetic account N_X^j just in case $Pr(N_X^i) > Pr(N_X^j)$ for all $j \neq i$. Call that *weak probabilistic optimality*. However, we may also want to capture something stronger in the same vicinity, viz. the condition that N_X^i be *much* more probable than any such N_X^j – what we may call *strong probabilistic optimality*.[30] A clue for how to define this stronger requirement is provided by the standard mathematical interpretation of the much-greater sign, '$\gg$'. A claim of the form $a \gg b$ is standardly interpreted as claiming that a is larger than b by some factor r, e.g. 10, 10^3 or 10^6 (often a power of ten). So on this interpretation $a \gg b$ is equivalent to $a > r \times b$ given such a factor r. This suggests a ratio-based approach to 'significantly more probable' that suits our purposes, viz. that N_X^i is significantly more probable than N_X^j just in case $Pr(N_X^i) > r \times Pr(N_X^j)$ for some factor $r > 1$. Here the factor r may be taken to vary with context or stakes, much like the thresholds in the Lockean thesis discussed in section 3. Moreover, if one thinks 'acceptable to commit' is vague, then r may be replaced by an interval I_r so that P is neither rational nor not-rational when $Pr(N_X^i)/Pr(N_X^j) \in I_r$ for the most probable noetic accounts that entail P and $\neg P$ respectively.[31]

We now have all the theoretical machinery we need to state the model in its definitive form. Informally, the model holds that whether it would be acceptable to commit to a proposition is determined by a probabilistic comparison between the most probable ways of completely answering all of one's understanding-seeking questions that affirm and deny the hypothesis, respectively. Specifically, acceptable commitments are entailed by *optimally probable* ways of answering such questions, where something is 'optimally probable' just when it is (perhaps significantly) more

---

[30] As I note below, I do not argue here for requiring strong as opposed to weak probabilistic optimality for acceptability. Rather, I leave this as an open question to be answered by considerations that fall outside the scope of this paper.

[31] It is perhaps worth contrasting the ratio-based interpretation suggested here with another salient possibility for interpreting 'significantly more probable'. Consider a difference-based approach, according to which N_X^i is taken to be significantly more probable than N_X^j just in case $Pr(N_X^i) > d + Pr(N_X^j)$ for some $0 < d < 1$. Now, this does not accord with the standard mathematical interpretation of '$\gg$', but then again we shouldn't let mathematical practice dictate our model of rational understanding. The real problem with a difference-based approach is that it smuggles in a probability-threshold of the kind we have already rejected (see §3). For note that on this approach the probability of any noetic account N must exceed d (whatever it is) in order for a proposition that follows from it to be acceptable, so d effectively becomes a probability-threshold. (Yet other interpretations of 'significantly greater' are of course possible, e.g. in terms of exponential functions rather than ratios or differences. I leave it to future work to explore which (if any) of these other interpretations would result in notions of strong probabilistic optimality that are better or worse suited for our purposes here.)

probable than all of its alternatives. This can now be made precise as follows:

> **The Optimality Model.** It is acceptable for an agent S to commit to a proposition P for the purposes of understanding a phenomenon X just in case the probability S would be rational in assigning to the most probable noetic account of X that entails P is r times greater than the probability S would be rational in assigning to the most probable noetic account that entails $\neg P$, i.e. just in case:
>
> $$\max\{Pr_S(N_X) : N_X \Rightarrow P\} > r \times \max\{Pr_S(N_X) : N_X \Rightarrow \neg P\},$$
>
> where $r \geq 1$.

If $r = 1$, this amounts to requiring that P be entailed by a weakly probabilistically optimal noetic account; if $r > 1$, then P must be entailed by a strongly probabilistically optimal noetic account (where the required 'strength' of the account increases with the value of r).

I do not wish to take a stand here on whether to set r at 1 or higher, and all results below are obtained independently of what value is given to r within this range. In this sense, r will be treated as a free parameter for the purposes of this paper. With that said, however, it is worth commenting on the skeptical consequences of setting r at significantly higher than 1. In that case, we can easily have situations in which very few propositions will count as acceptable, due to there being a number of different noetic accounts which are sufficiently close (probability-wise) to the probabilistically optimal account. For example,[32] suppose Barbara makes 100 claims in her book, and that the most probable noetic account incorporates all of these. However, suppose also that the 100 other noetic accounts which incorporate only 99 of these claims and deny the remaining claim are sufficiently probable for it to be false that these noetic accounts are r times less probable than the first account (for a suitably chosen r), so none of the 100 claims would count as acceptable according to the Optimality Model. In so far as we are troubled by implications of this sort, we should set r closer (or identical) to 1, since the closer r is to 1 there will be fewer noetic accounts whose probability are within a factor of r of the probability of the maximally probable account.

[32] I am indebted to an anonymous reviewer for this example.

5 Escaping Paradox

We have seen that it is possible to provide a precise statement of a probabilistic model of acceptability based on the idea that acceptable commitments are optimally rather than satisficingly probable – i.e., more probable than alternatives, as opposed to more probable than a given threshold. But is this a plausible model of acceptability? Well, I have suggested that the commitments involved in understanding should, minimally, be *logically coherent* – a requirement expressed more precisely in the form of three formal principles, ($\wedge$), ($\Rightarrow$), and ($\neg\bot$) (jointly labelled Deductive Cogency for Understanding; see §2). Recall also that we rejected what is perhaps the most popular theory of acceptability – that acceptability is epistemic justification – because the set of propositions one would be justified in believing need not be logically coherent in this sense (see §3). So the first test of the Optimality Model is whether the model is compatible with these principles of logical coherence. As it turns out, the Optimality Model is not just compatible with these principles; it validates them.

A simple and illuminating approach to proving that these principles hold in the Optimality Model proceeds from considering the set of all noetic accounts of some phenomenon X whose probability, when multiplied with r, is at least as high as the probability of one of the maximally probable noetic accounts of X.[33] Call this set of noetic accounts $\mathbf{N}^{+}_{\mathbf{X}}$. Since a given proposition P_i will *not* be acceptable by the Optimality Model just in case there is a noetic account in $\mathbf{N}^{+}_{\mathbf{X}}$ which does *not* entail it, the set of all acceptable propositions $\mathbf{P}$ will be all and only those proposition that are entailed by every member of $\mathbf{N}^{+}_{\mathbf{X}}$, i.e. by every noetic account whose probability multiplied by r is at least as high as that of the maximally probable noetic account. With this in hand, it now becomes a rather simple matter to prove that the Optimality Model validates ($\wedge$), ($\Rightarrow$), and ($\neg\bot$):

To show that the Optimality Model validates ($\wedge$) – roughly, the requirement that acceptability be closed under conjunction – let $\{P_1, \ldots, P_n\}$ be a set of propositions each of which it would be acceptable to commit to in understanding X. From what has been said above, it follows that each of these propositions is entailed by every noetic account in $\mathbf{N}^{+}_{\mathbf{X}}$. Now, any noetic account of X that entails each of these propositions also entails their conjunction $(P_1 \wedge \cdots \wedge P_n)$. So, since $P_1, \ldots, P_n$ are all

[33] I am deeply indebted and extremely grateful to an anonymous *Synthese* reviewer for suggesting this approach to proving the following results. My original proof strategy was considerably more cumbersome and less illuminating.

entailed by every noetic account in $\mathbf{N}^+_{\mathbf{X}}$, the same is true for $(P_1 \wedge \cdots \wedge P_n)$ as well. Since the set of all acceptable propositions $\mathbf{P}$ is (according to the Optimality Model) all and only those propositions that are entailed by every member of $\mathbf{N}^+_{\mathbf{X}}$, it follows that $(P_1 \wedge \cdots \wedge P_n)$ is in $\mathbf{P}$, i.e. that it is among the propositions to which it would be acceptable to commit in understanding X. Hence $(\wedge)$ holds on the Optimality Model.

A similar argument shows that the Optimality Model validates $(\Rightarrow)$ – roughly, the requirement that acceptability be closed under deductive consequence. Let $\{P_1, \ldots, P_n\}$ be a set of propositions each of which it would be acceptable to commit to in understanding X; thus each such proposition is entailed by every noetic account in $\mathbf{N}^+_{\mathbf{X}}$. Now let P_c be some deductive consequence of this set. Since every member of $\mathbf{N}^+_{\mathbf{X}}$ entails every member of $\{P_1, \ldots, P_n\}$, and since the latter jointly entail P_c, it follows from the transitivity of entailment that every member of $\mathbf{N}^+_{\mathbf{X}}$ entails P_c. So P_c is in $\mathbf{P}$, i.e. is among the propositions to which it would be acceptable to commit in understanding X. Hence $(\Rightarrow)$ holds on the Optimality Model.

Finally, to show that the Optimality Model validates $(\neg\bot)$ – roughly, the requirement that acceptable commitments be consistent – assume for reductio that there is a set of *in*consistent propositions $\{P_1, \ldots, P_n\}$ each of which it would be acceptable to commit to in understanding X. Each of these propositions would then be entailed by every noetic account in $\mathbf{N}^+_{\mathbf{X}}$. However, since $P_1, \ldots, P_n$ are jointly inconsistent, every noetic account that entails them would be inconsistent, including every account in $\mathbf{N}^+_{\mathbf{X}}$. But then any member of $\mathbf{N}^+_{\mathbf{X}}$ would have probability 0, since any inconsistent proposition has probability 0. But since $\mathbf{N}^+_{\mathbf{X}}$ is by construction the set of the most probable noetic accounts of X within a certain range (determined by r), they cannot all have probability 0. We thus conclude, by reductio, that the jointly inconsistent propositions $P_1, \ldots, P_n$ cannot all be acceptable after all.

In sum, then, the Optimality Model validates the three principles which jointly capture the thought that understanding should be logically coherent. This means, of course, that the Optimality Model obeys the conjunction of these principles, Deductive Cogency for Understanding, which in turn guarantees that the model is not susceptible to analogues of the preface and lottery paradoxes. This is a significant result, not least because the Optimality Model was not 'rigged' to avoid these paradoxes – that it does so is rather something that falls out of a natural way of thinking about understanding, viz. as an optimizing evaluation of propositions in terms of the probabilities of different ways of completely answering one's understanding-seeking

questions about a given phenomenon.

Let me end this section by returning to an earlier point. In section 3, I argued that understanding does not require justification in the Lockean sense, and briefly suggested that understanding should therefore not be taken to require belief either. In the course of these arguments, I defended the implication of Deductive Cogency for Understanding that extremely improbable conjunctions of propositions could be acceptable. It should be clear that the Optimality Model agrees on this point, since the fact that a proposition's most probable noetic account is more probable than that of its negation is no guarantee that the proposition or its associated noetic account is probable in an absolute sense. Indeed, it is perhaps worth noting that the noetic accounts themselves will in general be so extremely informative propositions that even the most probable among them will inevitably have a very low probability, since the probability of any given propositions is (roughly speaking) inversely proportional to its informativeness. So the Optimality Account implies – perhaps surprisingly – that there are acceptable noetic accounts whose probability is exceedingly low. While this may seem counterintuitive at first, it should be clear from the considerations adduced in section 3 that it could not be otherwise – one simply cannot simultaneously hold on to a threshold requirement for acceptability and hope to validate Deductive Cogency for Understanding.

6 Concluding Remarks

To understand something in an epistemically appropriate manner, the propositions to which one commits for that purpose should, minimally, be logically coherent. That is, these propositions should not be inconsistent with one another, nor should they be such that it would be unacceptable to commit to the deductive consequences of propositions to which one commits. We have seen that these requirements of logical coherence conflict with the otherwise-plausible-seeming view that the propositions one commits to in understanding should be epistemically justified (at least if epistemic justification has any connection at all to rational probability above a threshold). But we have also seen that an alternative model can be developed, the Optimality Model, where the acceptability of a given proposition is determined by whether it is part of a maximally informative account of the understood phenomenon that is (perhaps significantly) more probable than any alternative such account. Although more work remains to be done in developing and defending this model, the first steps have been taken here towards a probabilistic epistemology of

understanding.

In closing, I wish to briefly compare this probabilistic model with Elgin's theory of acceptability as it appears in her recent book, *True Enough* (2017). Elgin's theory is that a claim is acceptable for the purposes of understanding just in case it is part of an account that is in *reflective equilibrium*. An account is in reflective equilibrium when (i) its elements are reasonable in light of one another, and (ii) the account as a whole is as reasonable as any available alternative in light of antecedent commitments (Elgin, 2017, 66). These 'antecedent commitments' are claims that "we have some inclination to accept" (Elgin, 2017, 64), or "comprise our current best take on the matter under investigation" (Elgin, 2017, 64). So Elgin's theory is that individual claims derive their normative status – understanding-wise – from being part of a holistic system of claims that are mutually supportive and as a whole as reasonable as any other such system, given these antecedent commitments.

I want to draw attention to one similarity between Elgin's theory and the Optimality Model, and one dissimilarity. The similarity is that both theories are *holistic*, in the sense that they take the normative status of a single proposition to be determined by the normative status of an entire system of claims about the understood phenomenon, where the individual proposition or claim to be evaluated is part of, or entailed by, this system. The dissimilarity is that the two theories propose different ways of evaluating these systems. In the Optimality Model, the normative status of such a system (i.e., of a noetic account) is determined by a probabilistic comparison of it with other equally informative accounts of the target phenomenon. In Elgin's theory, the normative status of such a system is determined by whether it is in reflective equilibrium, i.e. by whether its claims are mutually supportive and as a whole as reasonable as any other such system.

Given this dissimilarity, it might seem that the Optimality Model must be at odds with Elgin's theory. However, a less combative approach would be to argue that the two theories are broadly compatible, or even complementary. The version of this approach that I prefer views the Optimality Model as providing a probabilistic framework in which relatively vague formulations such as "as reasonable as any alternative" (Elgin, 2017, 66) can be given precise mathematical meanings, which in turn might enable us to *prove*, as opposed to merely *intuit*, that our philosophical theories of understanding have various desirable properties. Indeed, I take myself to have effectively already gone some way towards doing exactly that for Elgin's theory, in that I have (roughly and among other things) shown that replacing "as reasonable

as any alternative" with "more probable than any alternative" makes the resulting theory respect the thought that acceptability requires logical coherence.

Relatedly, providing a probabilistic framework for Elgin's theory might help to simplify the theory (or, equivalently for our purposes, unify its different parts). Consider, in particular, the first clause in Elgin's two-tiered definition of 'reflective equilibrium', which requires that the elements of an account are mutually supportive. Although the Optimality Model does not explicitly specify that acceptable accounts should have this feature, a preference for accounts comprised of mutually supporting elements falls naturally out of the model as a consequence of the probability calculus. That is, an account consisting of mutually supporting elements will have a higher probability – and thus be more prone to be (more strongly) probabilistically optimal – than an otherwise similar account consisting of elements which do not support one another to the same extent. This is a consequence of the fact that the probability of a conjunction $(P_1 \wedge \cdots \wedge P_n)$ is, all other things being equal, higher to the extent that each proposition P_i provides stronger support to each other proposition P_j, even when the probability of each conjunct remains fixed in the comparison.[34]

[34]I am very grateful to Christoph Baumberger, Chris Dorst, Insa Lawler, and two anonymous reviewers, who all provided extraordinarily insightful comments on previous drafts of this paper. I am also indepted to my colleagues at Inland Norway University of Applied Sciences for helpful comments and illuminating discussions about this paper. Finally, I am grateful to the audience at Innsbruck University's workshop on Elgin's *True Enough*, and to Kate herself for inspiration, discussion, and encouragement.

References

Akiba, K. (2000). Shogenji's probabilistic measure of coherence is incoherent. *Analysis*, 60:356–359.

Baumberger, C. (2018). Explicating Objectual Understanding: Taking Degrees Seriously. *Journal for General Philosophy of Science*, Forthcoming.

Bengson, J. (2015). A Noetic Theory of Understanding and Intuition as Sense-Maker. *Inquiry*, 58:633–668.

Bokulich, A. (2012). Distinguishing Explanatory from Nonexplanatory Fictions. *Philosophy of Science*, 79:725–737.

Carter, J. A. and Gordon, E. C. (2014). Objectual Understanding and the Value Problem. *American Philosophical Quarterly*, 51:1–13.

Cohen, L. J. (1992). *An Essay on Belief and Acceptance*. Clarendon Press, Oxford.

Cooper, N. (1994). Understanding. *Aristotelian Society Supplementary Volume*, 68:1–26.

Dellsén, F. (2017). Understanding Without Justification or Belief. *Ratio*, 30:239–254.

Dellsén, F. (2018a). Beyond Explanation: Understanding as Dependency Modeling. *The British Journal for the Philosophy of Science*. DOI: 10.1093/bjps/axy058.

Dellsén, F. (2018b). Deductive Cogency, Understanding, and Acceptance. *Synthese*, 195:3121–3141.

Elgin, C. Z. (2004). True Enough. *Philosophical Issues*, 14:113–131.

Elgin, C. Z. (2007). Understanding and the Facts. *Philosophical Studies*, 132:33–42.

Elgin, C. Z. (2009). Is Understanding Factive? In A. Haddock, A. M. and Pritchard, D., editors, *Epistemic Value*, pages 322–330. Oxford University Press, Oxford.

Elgin, C. Z. (2017). *True Enough*. MIT Press, Cambridge, MA.

Fitelson, B. (2003). A probabilistic theory of coherence. *Analysis*, 63:194–199.

Foley, R. (1992). The Epistemology of Belief and the Epistemology of Degrees of Belief. *American Philosophical Quarterly*, 29:111–124.

Foley, R. (2009). Belief, Degrees of Belief, and the Lockean Thesis. In Huber, F. and Schmidt-Petri, C., editors, *Degrees of Belief*, pages 37–47. Springer, Dordrecht.

Frigg, R. and Nguyen, J. (2016). The Fiction View of Models Reloaded. *The Monist*, 99:225–242.

Gijsbers, V. (2015). Can Probabilistic Coherence be a Measure of Understanding. *Theoria: Revista de Teoría, Historia y Fundamentos de la Ciencia*, 30:53–71.

Greco, J. (2014). Episteme: Knowledge and Understanding. In Timpe, K. and Boyd, C. A., editors, *Virtues and their Vices*, pages 285–302. Oxford University Press, Oxford.

Grimm, S. (2006). Is Understanding a Species of Knowledge? *British Journal for the Philosophy of Science*, 57:515–535.

Grimm, S. (2011). Understanding. In Bernecker, S. and Pritchard, D., editors, *Routledge Companion To Epistemology*, pages 84–94. Routledge, London.

Grimm, S. (2014). Understanding as Knowledge of Causes. In Fairweather, A., editor, *Virtue Epistemology Naturalized: Bridges Between Virtue Epistemology and Philosophy of Science*, pages 347–360. Springer, Dordrecht.

Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, 10:41–53.

Hills, A. (2016). Understanding Why. *Nous*, 50:661–688.

Huemer, M. (2009). Explanationist Aid for the Theory of Inductive Logic. *British Journal for the Philosophy of Science*, 60:345–375.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge.

Kelp, C. (2015). Understanding phenomena. *Synthese*, 192:3799–3816.

Kelp, C. (2017). Towards a Knowledge-Based Account of Understanding. In Grimm, S., Baumberger, C., and Ammon, S., editors, *Explaining Understanding: Perspectives from Epistemology and Philosophy of Science*, pages 251–271. Routledge, New York, NY.

Keynes, J. M. (1921). *A Treatise on Probability*. Macmillan, London.

Khalifa, K. (2013a). Is understanding explanatory or objectual? *Synthese*, 190:1153–1171.

Khalifa, K. (2013b). Understanding, grasping, and luck. *Episteme*, 10:1–17.

Khalifa, K. (2016). Must understanding be coherent? In Grimm, S., Baumberger, C., and Ammon, S., editors, *Explaining Understanding: Perspectives from Epistemology and Philosophy of Science*, pages 139–164. Routledge, London.

Khalifa, K. (2017). *Understanding, Explanation, and Scientific Knowledge*. Cambridge University Press, Cambridge.

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press, Cambridge.

Kyburg, H. E. (1961). *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, CT.

Laplace, P.-S. (1951). *A Philosophical Essay on Probabilities*. Dover, New York.

Lawler, I. (2016). Reductionism about understanding why. *Proceedings of the Aristotelian Society*, 116:229–236.

Lawler, I. (2018). Scientific Understanding and Felicitous Legitimate Falsehoods. Unpublished manuscript.

Leitgeb, H. (2014). The Stability Theory of Belief. *Philosophical Review*, 123:131–171.

Leitgeb, H. (2017). *The Stability Theory of Belief*. Oxford University Press, Oxford.

Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability*, volume 2, pages 263–93. University of California Press, Berkeley.

Makinson, D. C. (1965). The Paradox of the Preface. *Analysis*, 25:205–207.

Mizrahi, M. (2012). Idealizations and Scientific Understanding. *Philosophical Studies*, 160:237–252.

Nelkin, D. K. (2000). The lottery paradox, knowledge, and rationality. *The Philosophical Review*, 109:373–409.

Olsson, E. J. (2005). *Against Coherence: Truth, Probability, and Justification*. Oxford University Press, Oxford.

Pritchard, D. (2009). Knowledge, Understanding, and Epistemic Value. In O'Hear, A., editor, *Epistemology (Royal Institute of Philosophy Lectures)*, pages 19–43. Cambridge University Press, Cambridge.

Riggs, W. D. (2009). Understanding, Knowledge, and the Meno Requirement. In haddock, A., Millar, A., and Pritchard, D. H., editors, *Epistemic Value*. Oxford University Press, Oxford.

Rosenkrantz, R. D. (1977). *Inference, Model and Decision: Towards a Bayesian Philosophy of Science*. Dordrecht, Reidel.

Schupbach, J. N. (2011). New Hope for Shogenji's Coherence Measure. *British Journal for the Philosophy of Science*, 62:125–142.

Shogenji, T. (1999). Is coherence truth-conducive? *Analysis*, 59:338–345.

Shogenji, T. (2001). Reply to Akiba on the probabilistic measure of coherence. *Analysis*, 61:147–150.

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, 63:129–138.

Sliwa, P. (2015). Understanding and knowing. *Proceedings of the Aristotelian Society*, 115:57–74.

Strevens, M. (2017). How Idealizations Provide Understanding. In Grimm, S., Baumberger, C., and Ammon, S., editors, *Explaining Understanding: New Essays in Epistemology and Philosophy of Science*, pages 37–49. Routledge, New York.

Sullivan, E. and Khalifa, K. (2019). Idealizations and Understanding: Much Ado About Nothing? *Australasian Journal of Philosophy*, DOI: 10.1080/00048402.2018.1564337.

van Fraassen, B. C. (1980). *The Scientific Image*. Clarendon, Oxford.

Weisberg, J. (2009). Locating IBE in the Bayesian Framework. *Synthese*, 167:125–143.

Wilkenfeld, D. A. (2016). Understanding without Believing. In S. Grimm, C. B. and Ammon, S., editors, *Explaining Understanding: Perspectives from Epistemology and Philosophy of Science*, pages 318–334. Routledge, New York.

Wilkenfeld, D. A. (2018). Understanding as Compression. *Philosophical Studies*, DOI: 10.1007/s11098-018-1152-1.

Zagzebski, L. (2001). Recovering Understanding. In Steup, M., editor, *Knowledge, Truth, and Duty*, pages 235–252. Oxford University Press, Oxford.