THIS PAPER SURELY CONTAINS SOME ERRORS

Brian Kim*

Abstract   The preface paradox can be motivated by appealing to a plausible inference from an author's reasonable assertion that her book is bound to contain errors to the author's rational belief that her book contains errors. By evaluating and undermining the validity of this inference, I offer a resolution of the paradox. Discussions of the preface paradox have surprisingly failed to note that expressions of fallibility made in prefaces typically employ terms such as *surely*, *undoubtedly*, and *bound to be*. After considering what these terms mean, I show that the motivating inference is invalid. Moreover, I argue that a closer consideration of our expressions of fallibility suggest that epistemically responsible authors would not be rational to believe that their books contains errors. I conclude by considering alternative expressions of fallibility that employ terms such as *possible* and *probable*, and discuss the role that expressions of fallibility play in conversation.

Academic authors customarily write in the prefaces of their books that there are bound to be errors for which the author alone is responsible. While these *preface claims* are commonplace expressions of the author's fallibility, they raise difficulties for our understanding of rational belief.

Consider a historian who has made such a preface claim. Being an expert in her field, she possesses good evidence for each claim that she has made and is rational to believe each of them.[1] If, on the basis of these beliefs, the author competently deduces that the conjunction of the book's claims is true, she would clearly be rational to believe this conjunction. Unfortunately, by adopting this latter belief, the author appears left with an inconsistent set of beliefs, all of which we have deemed to be rational. Her remark that there are bound to be errors is a perfectly reasonable one and indicates that she believes that one of the book's claims is false. So the author possesses a rational belief that all of the book's claims are true and a rational belief that not all the book's claims are true. But intuitively, rationality demands consistency so one cannot possess an inconsistent set of rational beliefs. Furthermore, the author would now be a position to deduce a contradiction on the basis of fully rational beliefs. And since the conclusions of

---

[1]We should consider the subset of a book's claims which are clearly categorical claims. This deflects the reply that even though authors assert that P, they really mean that probably P (c.f. Lacey 1970, Hoffman 1973). In addition, we should assume that the book is one in which each categorical claim that p transparently reflects the author's belief that p. In many cases, the author's attitudes towards the books' claims are more complicated. For example, in (Sorensen 1988, 13), the author remarks that the proffered philosophical accounts are proposed as "fertile [accounts]". This suggests a more complicated relationship between the author and the book's claims. Thanks to a reviewer for raising these cases.

competent deductions that rely on rational beliefs result in rational beliefs, we are left with the paradoxical conclusion that the author could rationally believe an explicit contradiction.

This paradox of the preface has typically been thought to raise problems for the traditional view that a set of rational beliefs is deductively cogent (i.e. consistent and deductively closed). For example, (Makinson 1965), (Klein 1985), (Kyburg 1997), and (Foley 1993) each resolve the paradox by rejecting the principle that rational belief is closed under conjunction. (Christensen 2004) goes further and argues that a subject's set of beliefs need not be deductively cogent because deductive logic has little relevance for the logic of belief. The aim of this paper is to propose an alternative resolution of the preface paradox, a resolution that arises through a careful analysis of the ways in which authors express their fallibility in prefaces. By considering what preface claims express, we will find good reasons to conclude that authors, like our historian, would not be rational to believe that there are errors in their books. Thus, preface claims pose no problems for the deductive cogency of rational belief.

I will begin by articulating the intuitions that motivate the preface paradox, and in Section 1, we will see that preface claims typically employ terms such as *surely* and *bound to be*. The use of these terms in preface claims will raise questions about these motivating intuitions. In Section 2, I offer an analysis of these preface claims and use the analysis to reject the claim that authors are rational to believe that there are errors in the book. While the bulk of the paper will be devoted to preface claims that appeal to adverbs like *surely*, I also consider, in Section 3, preface claims that appeal to terms such as *possible* and *probable*, and argue that these claims pose no problems as well. To conclude, I propose, in Section 4, that expressions of fallibility can play an important conversational role in articulating the type of interaction that speakers would like to have with their interlocutors. And in the special case of authors and readers, preface claims articulate the interaction that authors would like to have with their readers.

1   Preface Claims and the Preface Paradox

The preface paradox arises because our intuitions about what it is rational to believe conflict when applied, on the one hand, to what an author may believe about the book's claims, and on the other hand, to the beliefs that an author expresses in the book's preface. Applied to the book's claims, we conclude that the author may rationally believe the conjunction of the book's claims. Applied to what is expressed in the preface, we conclude that the author may rationally believe the negation of this conjunction. The former and latter intuitions thereby come into conflict with view that rational beliefs must be consistent.

As I noted above, most discussions of the preface paradox have focused on the intuitions used to support the claim that the author may rationally believe the conjunction of the book's claims. In opposition, I will argue that the paradox can be resolved by considering the intuitions that support the claim that the author may rationally believe that her book contains falsehoods. Thus, for the sake of our discussion, let us take the first set of intuitions for granted and assume that with respect to each of the book's claims, the author satisfies the norms that govern the fixation of belief, whatever they may be.

Let us now turn our attention to the intuitions used to support the claim that authors are rational to believe that their books contain errors. Two types of intuitions are typically offered

in support. The first type of intuition concerns the epistemic significance of recognizing one's fallibility. For example, in the original presentation of the paradox, Makinson noted that "the discovery of errors among statements which previously [the author] believed to be true gives him good ground for believing that there are undetected errors in his latest book."(Makinson 1965, 206) While the paradox is sometimes motivated by these epistemic intuitions, I will, for the moment, set them aside. The main reason for doing so is that by considering another set of intuitions that drive the paradox, we will be able to return to these epistemic intuitions and raise some clear doubts about them. Another reason to set them aside is that it is not at all clear that the author's recognition of her own fallibility provides sufficient evidence to rationally believe that there are errors in her book.[2] After all, if this intuition were a clear one, it would be hard to understand why the paradox really comes to light and gets its grip when we consider what authors assert in prefaces. Something is clearly lost if we start with the intuition that authors are rational to believe that their books contain falsehoods. More importantly, as (New 1978) and (Ryan 1991) have argued, in cases where the recognition of fallibility does provide sufficient evidence for this belief, an author may be irrational to believe each of the book's claims.

The second type of intuition concerns what authors are reasonable to assert in prefaces and what beliefs these assertions express. After all, authors appear perfectly reasonable in asserting that there are bound to be errors. And these assertions not only seem to express that the author has low confidence in the conjunction of the book's claims but also that she believes the negation of this conjunction. She believes that at least one of the book's claims is false. Moreover, since authors appear rational in asserting preface claims, this would imply that they are rational in believing what these claims express. While the validity of this inference is typically taken for granted, a simple argument can be given if we adopt the widespread view that there are epistemic norms that govern assertion.[3] The most plausible norms of assertion entail that I reasonably assert that P only if I rationally believe that P. For example, this inference is valid if a knowledge or rational belief norm is the epistemic norm governing assertion. In addition, since the cases of interest are ones in which the author reasonably believes that she satisfies the epistemic norm governing assertion, then the inference is also valid if a truth norm governs assertions.[4] Thus, according to most of the proposed epistemic norms of assertion, if the speaker reasonably asserts that P, then the speaker is rational to believe that P. Call this the *reasonable assertion-to-belief inference*.

While the reasonable assertion-to-belief inference offers a clear and intuitive way of motivating the preface paradox, it is important to note that the paradox requires an additional assumption. For if we are to infer from the reasonableness of asserting preface claims to the reasonableness of the belief that there are errors, we must assume that by asserting a preface claim, an author is asserting that there are errors.[5] From what I can tell, discussions of the preface paradox have

---

[2]This discussion, as exemplified in (Pollock 1986), (Hawthorne and Bovens 1999) and (Douven 2003), is complicated by questions about probabilistic rules of acceptance. I will set these issues aside since I am not using the preface paradox as a puzzle about probability and rules of rational acceptance.

[3](Williamson 1996) and (DeRose 2002) are prominent supporters of this view.

[4]These secondary norms are those that govern whether the subject reasonably believes that some act conforms to a norm. And if truth is the epistemic norm governing assertion (c.f. Weiner 2005) and the speaker reasonably believes that her assertion satisfies the epistemic norms governing assertion, then she reasonably believes that P is true. See (DeRose 2002) for a discussion of these secondary norms.

[5]It may be worth noting here that we should ignore preface claims that acknowledge the existence of typo-

3

found this assumption so innocuous that it is not even worth mentioning.[6] However, once we consider how preface claims are actually made, we should have questions about what preface claims express, and once we clarify what preface claims express, we will have good reasons to doubt that the reasonable assertion-to-belief inference can be used to motivate the paradox.

The most straightforward way for an author to express that the conjunction of a book's claims is false would be to assert that there are errors in the book. It is extremely peculiar then that preface claims never take this form. For example, Makinson originally wrote, "there will inevitably be some mistakes and [the author] is responsible for them." (Makinson 1965, 205) In fact, most preface claims are stated in a similar way (italics are mine):[7]

> In his preface to his historical narrative, the author admits that he has *undoubtedly* made some mistakes. (Stalnaker 1984, 92)

> This is the paradox of the preface (so named because in the original version the author confesses in the preface that his book *probably* contains a falsehood). (Pollock 1986, 246)

> Subsequently, as you draft the preface to your work you add the usual disclaimer: 'I wish to express my gratitude to my colleagues for their many thoughtful comments and suggestions on earlier drafts; any remaining errors - and there are *bound to be* some - are entirely my own.' (Hawthorne and Bovens 1999, 242)

> In the preface of the book you are now reading, I apologize for the errors that are *sure to be* in the text. (Sorensen 2003, 103)

The fact that preface claims are not expressed as straightforward claims of error raises the immediate worry that we've been too quick in our interpretations of them. Makinson assumed that preface claims were straightforward claims of error when he noted that "[modal] concepts do not even enter into [the] problem." (Makinson 1965, 206) And he argued for the uniqueness of the preface paradox on the basis of this assumption. This historical fact makes it all the more curious that preface claims typically come in the following form: There are

---

graphical errors. These are errors at the level of communication, not at the level of the author's beliefs. A nice example of this is the preface of (Suppes 1960) where the author thanks his editors for their help in removing errors and goes on to say that errors will still be found. In the context, it seems clear that Suppes is referring to typographical errors.

[6](Evnine 1999) is a notable exception. Evnine thinks that the reasonable assertion-to-belief inference does not apply because we need an additional inference from a second-order belief to a first-order belief. He proposes that preface claims express the author's rational belief that something she wrote was false. However, if we are to infer from this that the author ought to believe that one of her book's claims is false, we must assume that authors are required to possess an accurate inventory of all the claims made in the book. Evnine argues that since rationality does not make this demand, this inference is not always valid. While Evnine's observation is right, I do not think it solves the paradox. Authors who know exactly what they have written still appear reasonable in asserting preface claims. This situation does not depend upon idealized agents with perfect memory for the author may use various tools like a computer to keep track of what he has written and whether he has accepted each written claim. And if these authors may assert preface claims, the paradox still arises.

[7]Here are some additional examples (italics are mine). "Though I've done everything possible to assure that the information which follows is correct, *undoubtedly* it still contains a few errors, and for this I apologise." (Pinto 2001, 53) "I realize that, because of the complex nature of the issues involved, the text of the book is *bound* to contain some errors. For these I now apologize in advance." (Rescher 2001, 213) "I have to acknowledge my own responsibility for the errors this book will *inevitably* contain." (Clark 2002, xiii)

surely/undoubtedly/inevitably errors in this book. Let's call these *surely claims of fallibility* or *surely claims* for short. Surely claims are natural and commonplace expressions of fallibility but despite their ubiquity, it is not obvious how we should interpret them. The most literal interpretation, that the author is sure there are errors, is clearly inappropriate. As (Christensen 2004) notes, it would be inappropriate for the author to straightforwardly assert that there are errors.[8] And if the straightforward claim of error is inappropriate, it would be even more inappropriate to make the stronger assertion that one is certain that there are errors in the book. What evidence would allow the author to be certain? And could the author possess such evidence while nevertheless being rational in believing each of the book's claims? After all, if the author is certain that there are errors, she comes off as dishonest or irresponsible by asserting each claim without qualification. The oddity of expressing assuredness raises a question about preface claims that will guide our entire discussion. If it is inappropriate for the author to assert her assuredness or certainty, then why is it perfectly reasonable for her to assert the surely claim? This calls out for a more careful analysis of surely claims. Of course, preface claims do take other forms and we will consider preface claims of the form, there are probably/possibly errors, in Section 3.

2   Surely Claims

The most common preface claims involve epistemic adverbials such as *surely*, *certainly*, and *undoubtedly* as well as adverb phrases that function similarly like *bound to be*. Call these *surely adverbials* and the assertions that employ them *surely assertions*.[9] To avoid confusion, "surely claim" refers a special class of surely assertions that express one's fallibility. I will argue that in the cases of interest to the preface paradox, surely adverbials function as evidentials. So before we turn to *surely*, let me briefly describe what evidentials are.[10]

2.1   Evidentials

Evidentiality is concerned with the linguistic marking of the sources of knowledge or belief. So evidentials identify sources of information. For some natural languages, evidentials have their own grammatical category. In the following example from Wintu, a language spoken by the Wintu people of northern California, the copula 'ke' functions as an evidential (EV) and identifies hearsay as the evidential source of one's assertion.[11]

Coyi:la-   ke: -ni
Drunk   EV   I
I am drunk (So they tell me)

---

[8] "Admittedly, there would be something odd about a preface that baldly stated 'This book contains errors!'" (Christensen 2004, 35-36) This also places some constraints on what variants of the preface paradox we must consider. Many resolutions of the preface paradox fall prey to reformulations of preface claims. However, it is important that we remember that such reformulations cannot amount to straightforward claims of errors or to claims that are stronger than these straightforward claims.

[9] I borrow the term from (Biber and Finegan 1988).

[10] While I will focus on the adverb *surely*, my analysis applies to all surely adverbials.

[11] (Fox 2001)

Since the copula marks the evidential source as coming from hearsay, the speaker is asserting that he has heard from others that he is drunk. Alternatively, he could have marked a direct first-person source in which he would have been asserting that he has first-person awareness of his drunken state. Evidentials can mark a wide variety of sources from direct sensory sources to indirect reports such as testimony and folklore.[12] For our discussion, any term that marks the evidential source of what is asserted will count as an evidential. Thus, all languages have markers of evidentiality even if they are not built into the language's grammar.

2.2 'Surely'

I propose that in the cases that are relevant to the preface paradox, *surely* marks the evidential source of an author's beliefs and thereby functions as an evidential.[13] This function will help us to understand what surely claims express and why they are particularly suitable as preface claims. To defend this proposal, consider some uses of *surely*. The first thing to note is that *surely* does not always express the speaker's assuredness and may be used when the speaker has some degree of doubt. For example, I may come across a familiar song and assert, "Surely I've heard that somewhere." Though I may believe that I've heard the song before, I am not expressing that I am absolutely sure. Therefore, *surely* does not always express the speaker's assuredness.

The evidential role of *surely* becomes particularly clear in situations where it is appropriate to assert that surely P, but inappropriate to assert that P.[14] These are the cases of interest because in prefaces, straightforward assertions of error are inappropriate but surely claims are not. Consider the following case.

> Sue and Bill are college physics majors discussing the current state of the science. Sue expresses her optimism, noting the many successful predictions arising from quantum mechanics. Bill replies, "But surely quantum mechanics is false. In the very least, it is not a fully adequate theory. There is bound to be some phenomenon that the theory cannot account for."

Though Bill's surely assertion sounds very natural, the situation would be different if he had omitted the surely adverbial. It would be very odd for Bill to straightforwardly assert that quantum mechanics is false or that it is not a fully adequate theory. What evidence does Bill have to make this claim? He is no expert, and there is no evidence that he can point to that can support this claim. Bill's weak epistemic state with respect to this claim explains the oddity of the straightforward assertion. Given the epistemic norms governing assertion, a speaker's epistemic state must at least be strong enough to license belief in the asserted proposition.

---

[12]See (Willett 1988) for a taxonomy of evidential markers found in natural languages.

[13](Downing 2001) argues more generally for this claim. The argument found in (Von Fintel and Gillies 2010) that the epistemic must is an evidential marker signaling an indirect inference may be thought to apply to *surely*. My claim that *surely* is an evidential marker of potential, unspecified evidence is meant to be compatible with the fact that it also signals an indirect inference.

[14]Even in cases where the straightforward assertion seems appropriate, *surely* appears to function as an evidential. In a math textbook, the author may write that the conclusion surely follows. Here, the author appears to be identifying an unarticulated argument as the source of the author's belief in the asserted proposition. I discuss cases like these below.

Therefore, the inappropriateness of making the straightforward assertion is explained by the fact that Bill is not rational in believing that quantum mechanics is false. This raises the question, why is it appropriate for Bill to assert the corresponding surely assertion?

The appropriateness of Bill's surely assertion can be explained by *surely*'s function as a very unique evidential. What is particular about Bill's case is that he is appealing to evidence that he does not yet possess. Bill is indicating that there is some evidence out there that would support his assertion. This potential, unspecified evidence is the evidential source of his assertion. The evidence is potential because Bill does not possess it, and it is unspecified because he cannot tell Sue what the evidence would be or what could she could do to find it. Therefore, Bill is indicating that by the letter of the law, he is not in a position to make the assertion. However, by appealing to potential, unspecified evidence, he is attempting to satisfy the norms of assertion by proxy.[15] And Bill is hoping that Sue will cooperate by agreeing that such evidence exists and accepting the assertion on the basis of this appeal.

To clarify the nature of this potential, unspecified evidence, we should ask a further question. Is Bill rational in believing that such evidence exists? That is, does Bill have evidence that such evidence exists? As Richard Feldman has argued, "evidence of evidence is evidence."[16] So potential, unspecified evidence cannot be second-order evidence. If it was, then according to Feldman's slogan, Bill would thereby possess evidence that quantum mechanics is false and would be warranted in making the straightforward assertion. So we would be left with no explanation of why the surely assertion is appropriate. While there are problems with Feldman's slogan, we may assume it for the sake of our discussion.[17] After all, Bill does not even possess evidence that there is evidence. As Bill's final assertion indicates, he can only assert that there is bound to be such evidence and it would have been inappropriate for Bill to straightforwardly assert that such evidence does exist. So, when we consider Bill's epistemic state with respect to the evidence that is supposed to support the claim that quantum mechanics is false, we find that Bill can only appeal to potential, unspecified evidence that such evidence exists. In fact, the evidential support for Bill's surely claim appears to be unspecified all the way down. There is no stage at which he can point to concrete evidence.[18] If higher-order evidence for p can be described as evidence that is removed to some degree from p, then potential, unspecified evidence is evidence that is removed to the limit from p. And I am proposing that in the limiting case where evidence is far enough removed, Bill possesses no evidential support for the asserted claim.[19]

Given the evidential meaning of *surely*, what role do surely assertions play in the cases of interest? Typically, assertions are proposals to update what is commonly presupposed amongst conversational participants and in order to update this common ground, speakers should know

---

[15]One might think that if inquiry went on long enough, Bill would eventually possess the evidence and he is thereby appealing to the strength of his future inevitable epistemic state.

[16](Feldman 2007: 151) Thanks to a reviewer for raising this point.

[17](Fitelson 2012) has shown there are counter-examples to Feldman's principle. Nevertheless, Fitelson agrees that the principle holds when we have conclusive evidence that there is conclusive evidence that quantum mechanics is false. This restricted version of Feldman's principle seems to apply here since Bill may express that there will surely be conclusive evidence that quantum mechanics is false.

[18]One may object that Bill knows that every prior scientific theory has been inadequate. As I discuss below, if this type of evidence does count as genuine evidence, then our form of the preface paradox does not arise and we are left with a paradox that is much like the lottery paradox.

[19](Hall 1999) offers a resolution of the surprise exam paradox and appeals to the intuition that if evidence is far enough removed, it weakens the evidential support provided by that evidence.

or rationally believe what they are asserting.[20] However, there are situations in which speakers would like to update the common ground without being in a strong enough epistemic position to assert that P. The evidential meaning of *surely* in cases like Bill's allows speakers to use a surely assertion to achieve this aim when the interlocutors are sufficiently cooperative.[21]

Though I have considered just a special set of cases, surely assertions play a similar role when we consider a more relaxed set of cases. In these cases, the speaker may possess evidence or arguments for the asserted proposition but may have no desire or intention to defend challenges to the assertion. These cases differ from Bill's assertion since the speaker may be reasonable in making the straightforward assertion. This type of case often arises when one is asserting his or her intuitions. For example, while discussing a case like the original trolley case, Peter Singer writes, "Surely we should rescue the larger number of people."(Singer 2010, 59) Though Singer may possess arguments for the asserted proposition, he is hoping that the reader will, for the sake of efficiency, grant that the unspecified argument or evidence exists and accept the assertion without challenge. Surely assertions are also commonplace in discussions of rational preference.[22] It may be asserted that one should surely prefer the choice with better consequences. These assertions seem to indicate that the author has no intention of providing an argument or evidence for what is asserted. It is important to note that surely assertions are not suppositions. They propose to update the common ground and do not merely ask one's interlocutors to suppose something for the sake of the argument.

To summarize, *surely* functions as an evidential, marking the evidential source of one's assertion. For our discussion of preface claims, I will focus just on those cases where the assertion that surely P is appropriate but assertions omitting the surely adverbial are not. In these special cases, *surely* marks the evidential source as unspecified and not possessed by the speaker, and surely assertions are used by these speakers to satisfy the norms of assertion by proxy.

2.3 Surely Preface Claims

Now that we have an account of what *surely* expresses in the appropriate cases, let's return to the preface paradox. In prefaces, it is inappropriate for authors to make straightforward assertions of error but appropriate to assert that there are surely errors in the book. We can now explain this discrepancy. When authors assert surely claims, they admit that they do not possess sufficient evidence that there are errors. Rather, they are appealing to potential, unspecified evidence, on the basis of which they would, if they possessed it, be rational to believe that there are errors. In section 4, I discuss what conversational function surely assertions play in prefaces, but for the moment, let us consider how this analysis resolves the preface paradox.

The preface paradox is motivated by the appropriateness of asserting preface claims. As we have already noted, this intuition along with the epistemic norms governing assertion is thought to entail that an author is thereby rational to believe that her book contains falsehoods. Our

---

[20] "An assertion should be understood as a proposal to change the context by adding the content to the information presupposed." (Stalnaker 1999, 10)

[21] Of course, *surely* can also play a rhetorical function by expressing incredulity at believing the negation of what is asserted. When used in this way, utterances using *surely* may be illocutionary acts that attempt to persuade one's interlocuters. Thanks to Don Hubin for making this point.

[22] See p. 35 of (Kaplan 1983) for some examples.

analysis of surely adverbials undermines the argument from the reasonableness of preface claims to the reasonableness of the author's belief. Authors utter surely claims when they are not in a strong enough epistemic state to make the more straightforward assertion of error. And in these cases, *surely* is an evidential that identifies a set of potential, unspecified evidence as the evidential source of the author's assertion. By appealing to potential, unspecified evidence, the author attempts to satisfy the norms of assertion by proxy. Therefore, if a speaker reasonably asserts that surely P, this does not entail that she is rational to believe that P.

Even though this particular inference is invalid, might the reasonableness of preface claims offer any reason to think that an author is rational to believe that one of her book's claims is false? In the cases we are considering, the author's only evidence for believing that she is in error is potential, unspecified evidence. So we can replace this question with another one. Is a subject rational to believe a proposition if the only evidence she can appeal to in support is not strictly speaking possessed by the subject? In other words, does potential, unspecified evidence offer good reasons to believe a proposition?

A brief consideration of the epistemic properties of potential, unspecified evidence will suggest that this evidence does not make one's belief rational. Normally, the amount of evidence a subject possesses in favor of some proposition P is in direct proportion to the amount of trust or reliability we will attribute to a subject as an informant about whether or not P. However, when a subject can appeal only to potential, unspecified evidence, we intuitively judge her to be an unreliable informant, not to be trusted. Furthermore, if one is rational to believe P on the basis of testimony, then the testimony of subjects who possess only potential, unspecified evidence should be sufficient for justifying my own beliefs. This would entail that Bill's testimony can justify the beliefs of others.

Next, potential, unspecified evidence does not always respond to new evidence in the right way. The support that a set of evidence offers for the acceptance of a hypothesis can typically be undermined or rebutted. Having seen a die land with four pips facing up for twenty consecutive throws, I may conclude that the die is not fair and that it may be loaded. My beliefs should change once I see a thousand additional throws that suggest otherwise. Though the original set of evidence offered support for the hypothesis that the die was not fair, the additional evidence undermines this support. In contrast, potential and unspecified evidence does not respond to new evidence in the right way. After some reading, Bill may come to believe that quantum mechanics has been unparalleled in its successes. Despite his improved epistemic state, Bill may nevertheless insist that the theory is surely false. Since Bill has not specified what type of evidence he is referring to or how it might be discovered, he can hold onto this belief despite the fact that he has gained evidence to the contrary. His evidence, since it is potential evidence, cannot be undermined.

Of course, it may be objected that when it comes to academic books, historical books in particular, authors do possess evidence that their books contain errors. Authors typically know that they have made errors in the past and that errors have been found in most every book in the field. This objection brings us back to the epistemic intuitions that we had previously set aside. These were intuitions about the epistemic significance of recognizing one's fallibility. After all, doesn't a recognition of one's fallibility provide strong evidence to believe that something one believes is false?

While a full discussion of the significance of meta-inductive evidence is beyond the scope of this paper, let me offer a few replies that stays within the spirit of our discussion. If the recognition of fallibility provides strong support for the claim that one of the book's claims is false, then why is it inappropriate for authors to assert that there are errors? The simplest explanation is that the author lacks sufficient evidence to assert the straightforward claim of error, and the evidential meaning of *surely* explains why surely claims are, in contrast, appropriate. If the meta-inductive objection is to have any bite, we need an alternative explanation of the discrepancy in the appropriateness of the two types of assertions. Without such an explanation, our linguistic intuitions strongly suggest that responsible authors simply do not possess sufficient evidence to rationally believe that there are errors in their books.

Of course, there may be a pragmatic explanation of this discrepancy. Straightforward assertions of error may be inappropriate because they pragmatically implicate the false claim that the author knows where the error is. And this can be supported by the fact that an author could cancel this implicature by asserting that there are errors but she does not know where they are. While such assertions are reasonable, they offer no support for the claim that authors are rational to believe that there are errors. To show why, let's return to Bill who believes that quantum mechanics is descriptively inadequate. He too may assert that there is phenomenon that quantum mechanics cannot explain but he does not know what they are. This assertion reveals exactly why Bill's belief is irrational. In canceling the implicature, Bill straightforwardly asserts that he lacks the relevant evidence. As far as I see, Bill and our authors appeal to the same type of evidence, and Bill's example just highlights how odd it would be to think that this type of evidence is genuine, justifying evidence.

Finally, if we did assume that the recognition of one's fallibility amounted to the possession of statistical evidence of error, the preface paradox may simply disappear or become a version of the lottery paradox. On this assumption, the truth or falsity of any particular claim in the book would appear analogous to possessing the losing or winning ticket in a very large and fair lottery. For we would think that the reliability of the author's fallible belief forming processes was akin to the likelihood of one's lottery ticket failing to be picked. Just as one is confident that each ticket will lose but certain that some ticket will win, the author is confident that each claim is true but certain that some claim is false. Once we view the uncertainty of the author's beliefs as akin to the riskiness of lottery beliefs, then the paradox seems to change form. Here, I would agree with (Ryan 1996) that one is not rational to believe that any single ticket will lose. By analogy, an author would not be rational to believe each of the book's claims. So if we assume that the recognition of fallibility counts as statistical evidence, then we may have simply rejected some of the intuitions – ones we have taken for granted – that motivate the preface paradox. Whether one agrees with these remarks, the main point is that it is not clear that the lottery paradox and the preface paradox should be viewed as the same paradox.[23]

---

[23]One key difference between the lottery paradox and the preface paradox concerns the aggregation of risk or uncertainty. In the lottery, we know exactly how to measure the risk we would undertake if we were to believe that some ticket will lose. Furthermore, we also know how to aggregate these risks if we were to conjoin our beliefs. We also possess conclusive evidence that some ticket will win. Our author is in a very different situation. Not only do we not know how to measure the uncertainty the author embraces if she were to believe each well-researched historical claim, we also do not know how to aggregate this uncertainty. Finally, the author does not possess conclusive evidence that some claim is false.

To summarize, reasonably asserting that surely P does not imply that the speaker is rational to believe that P. Therefore, while authors are reasonable to assert that surely there are errors, the reasonableness of this assertion does not imply that they are rational to believe that there are errors. As a result, it does not matter whether we interpret the author as expressing a belief that there are errors. In fact, the author may be absolutely certain that there are errors but since she does not possess sufficient evidence, her belief is not rational.

3   Alternative Expressions of Fallibility

Though preface claims are typically expressed as surely claims, authors sometimes express their fallibility in other ways. But before we consider these alternatives, let's briefly reconsider why straightforward claims of errors are inappropriate.[24] I proposed that the simple reason for the infelicity was that authors lack sufficient evidence to assert them. However, another reason for their infelicity is that when we make straightforward claims of error about previously asserted claims, we often express a change of mind. If I straightforwardly assert that I was wrong, I am indicating that my past beliefs were mistaken and that I no longer possess that belief. So by asserting that there are errors in the book, the author would be retracting some of the claims in the book. Therefore, straightforward claims of error can be inappropriate if they attempt to retract claims before they have been made. In contrast, if an author makes such an assertion at a later period of time, it would be a clear indication that she has changed her mind.

Though straightforward claims of error are inappropriate, authors do appear reasonable in asserting that it is possible or probable that there are errors.[25] There is currently a lively debate about the relationship between qualitative all-or-nothing belief, degrees of belief, and epistemic possibility.[26] Since it is beyond the scope of our discussion to consider all of these accounts, let me simply consider one account of belief for which assertions of possible error would be appropriate.[27] The notion of *expectation* as used by formal epistemologists identifies one type of qualitative belief on which one may concurrently believe that p and judge not-p to be epistemically possible. When a subject expects that p, p is judged to be maximally probable (i.e., probability 1). Nevertheless, the subject may still coherently judge not-P to be epistemically possible. For example, Bill may assign probability 1 to every proposition of the form, Bill does not weigh x pounds, where x is some real-value between 0 and 500. In this case, Bill expects that he does not weigh x pounds and also judges that it is epistemically possible for him to weigh x pounds.

---

[24]See pages 33-34 of (Christensen 2004) for discussion

[25]This suggests that the preface claim expressing possible error should be treated differently than the Moorean claim that p but p might not be the case. The infelicity of asserting the latter can be explained pragmatically. One may not assert that p but p might not be the case because one would be proposing an incoherent update to the common ground such that every world is a p world and that some world is a not-p world. In contrast, if one were to preface a set of assertions by stating that what I assert might be wrong, then one is simply weakening the assertoric force of the subsequent utterances. In §4, I propose that these remarks can be used by speakers to allow for and countenance potential disagreements. Thanks to a reviewer for pointing out the connection to Moorean sentences.

[26]See (Arló-Costa and Pedersen 2012) for a summary of the debate.

[27]Some types of qualitative all-or-nothing beliefs are incompatible with the possibility of error. A *full belief*, which is characterized in (Levi 1980), offers one example. The incompatibility of full belief and the possibility of error is a sufficiently good reason not to view our author's beliefs as full beliefs.

Expressions of probable error are more problematic since it seems incoherent to believe that P while also judging not-P to be probable. It should first be noted that there may be cases in which it would be inappropriate for authors to assert claims of probable error. After all, just like straightforward expressions of error, expressions of probable error (e.g. There are probably errors in the book) often express a change of mind. Suppose I believe my dog to be in the backyard, but as I walk home, I see the side gate open and proclaim, "Oh no! The dog probably got out." My assertion indicates that I no longer believe that my dog is in the backyard. Therefore, in some cases, the assertion that there are probably errors in this book sounds just as strange as the straightforward expression of error. Analogously, it is just as strange to say as I turn in a job application, "There are probably errors in the application" as it is to say "There are errors in the application." If it is probable that there are errors, then shouldn't I go over the application to uncover and fix those errors? To make unqualified assertions that are prefaced by an assertion of probable error makes one's assertions appear disingenuous.

Nevertheless, authors sometimes appear reasonable in asserting that there are probably errors, and we can offer an interpretation of these assertions that is compatible with the deductive cogency of rational belief. Probability claims can encode at least two different types of epistemic evaluations.[28] One dimension of evaluation specifies the strength of one's beliefs, which may be represented as a quantitative or qualitative degree of belief.[29] The second dimension specifies the reliability or stability of one's beliefs or degrees of belief. For example, Bill may possess no empirical information about a coin and judge that it is equally likely that it land heads or tails when tossed. Alternatively, he may possess statistical information about previous tosses and evidence about the physical symmetries exhibited by the coin on the basis of which he judges that it is equally likely that the coin land on one side or the other. Intuitively, Bill's epistemic state in the first case is different than the second. The difference can be made out by differentiating the reliability that identical probability judgments can have. After all, we find Bill's probability judgment in the latter case to be much more reliable and supported by more evidence than his probability judgment in the former case.

Our full beliefs can also be more or less "reliable". This notion of reliability can be made out by appealing to the notion of entrenchment.[30] I fully believe that I am not a brain in a vat and that I had a birthday dinner this past year in Los Angeles. However, the former belief is more entrenched than the latter. It is more entrenched because it is more likely that I gain evidence that undermines my belief about the location of my birthday dinner than I gain evidence that undermines my belief about not being brain in a vat. Only in very far-fetched possibilities would my belief that I am not a brain in a vat be false. There are however some nearby possibilities, where I possess faulty memory and geographical knowledge, in which my

---

[28]See Chapter 1 in (Gardenfors and Sahlin 1988).

[29]It would be sufficient for our purposes to interpret strength of belief as epistemic probability. What is important is that this notion be distinguished from the credences that are derived from fair betting rates. If we did interpret strength of belief in terms of the credences derived from betting behavior then the resulting paradox or problem would be quite different. For example, one preface claim that appeals to these credences is found in (Sorensen 2001: 100-1). However, it is unclear what the relationship is between rational belief and rational betting behavior. For example, on the view of belief found in (Ross and Schroeder 2012), one may rationally believe that p even though one is not willing to bet at any odds on p. On such a view, it is not clear that any paradox arises. Thanks to a reviewer for raising this type of case.

[30]See (Gardenfors 1988)

belief about the location of my birthday dinner would be false. This suggests that my brain in a vat belief is more entrenched and thereby more reliable than my birthday dinner belief. I may express this difference by asserting that I could potentially be wrong about the location of my birthday dinner but I am certainly right about not being a brain in a vat. Moreover, with regard to the least entrenched beliefs that I have, it seems reasonable to assert that I am probably wrong about those beliefs. While I believe them to an equal degree as my brain in a vat belief, they are the most likely to change and thus the most likely to be judged false as I gain more information.

When it comes to the author's beliefs, the least entrenched belief is about the conjunction of the book's claims. This is the strongest belief, the belief that entails all the others, and thus the belief that is most susceptible to revision. So when an author admits that she is probably or certainly wrong, she may simply be expressing a fact about the reliability of this strongest belief. However, nothing we assert about the stability or reliability of our beliefs undermines the strength of our beliefs. Even if one's probability judgments and beliefs are very unreliable and highly susceptible to revision, this would not undermine the .5 degree of belief that the coin lands heads nor would it change the fact that I believe my last birthday dinner was in Los Angeles. These expressions of unreliability convey that one is not an omniscient creature and does not possess all the relevant evidence. Our beliefs are corrigible, and for any extremely informative belief, it is likely that we will change our mind.

Lastly, authors could reasonably assert that errors will be found.[31] These assertions of future error can be accommodated by appealing to our interpretation of surely claims. In the cases of interest, authors lack sufficient evidence to assert that errors exist. So when authors make expressions of future error, the author is most plausibly interpreted as asserting that from the perspective of an agent (perhaps a future agent) with more information, some of the book's claims will be correctly judged to be false. As in the case of surely claims, the author is appealing to a set of evidence that she does not yet possess. And we have concluded that in these cases, even if the author believes that errors will be found, entailing that there are errors, she is not rational to believe that there are errors.

4 Prefaces and Our Expressions of Fallibility

I have argued that the reasonableness of preface claims offers no reason to think that authors are rational to believe that their books contain errors. Furthermore, I have argued that a careful consideration of preface claims suggests that authors are not rational to believe such a claim. So if preface claims do not express the author's belief that there are errors in the book, what purpose do they serve? To conclude our discussion, I'd like to offer a suggestion and also propose a general observation about the role that expressions of fallibility play. We can consider one role of preface claims by considering the types of books that do not and should not contain prefatory remarks that express the author's fallibility. With respect to these books, what is the relationship between the author and reader? Furthermore, if certain authors, given their aims, should not, pragmatically speaking, express their epistemic modesty, what does this reveal about the task that the author expects of her readers? By considering the contrasting case, we will

---

[31](Christensen 2004) uses future-tensed expressions of error as examples of appropriate preface claims.

gain insight into how to answer these questions when we consider the prefaces of books where authors do express their fallibility.

Religious and short mathematical texts typically do not contain prefatory acknowledgments of epistemic modesty. In some Judeo-Christian traditions, the New and Old Testaments are considered to be infallible sources of information. And for short mathematical texts, the author can be sure that no typos exist and that every claim is true. In both cases, it is natural to think that no prefatory expressions of fallibility ought to be made. Given the author's epistemic superiority, the relationship between the author and reader in these cases is unique. The book's claims are not to be doubted, and this places normative constraints on the reader's task. The reader should not debate with the text, question its claims, or criticize its arguments. Rather, the burden is placed on the reader to interpret the text appropriately with truth being a condition for adequate interpretation. If the reader finds errors, the fault lies in her own interpretation or lack of evidence. This may have been Wittgenstein's intention when he wrote in the preface to the Tractatus that his aims "would be attained if there were one person who read it with understanding." (Wittgenstein 1922, 27) For rather than express his fallibility in the preface, he writes that "the truth of the thoughts communicated here seems to [him] unassailable and definitive. [He is], therefore, of the opinion that the problems have in essentials been finally solved." (Wittgenstein 1922, 29) Wittgenstein's readers are not asked to read the book as a conversation between epistemic peers but to approach the book as an oracle.

In contrast, most authors do not represent themselves as infallible sources of information, and the expressions of epistemic modesty found in prefaces are conventional ways of acknowledging this fact. This acknowledgment allows the author to make her relationship with the reader explicit. She is hoping to engage in a cooperative conversation with her reader and is thereby inviting the reader to adopt a critical attitude and to engage in debate.[32] Our analysis of *surely* helps to identify at least one function that preface claims have. Since the author acknowledges that there is evidence that could potentially undermine the book's claims, she is thereby acknowledging that the reader's epistemic state may be different and better than her own. This acknowledgment leaves open the possibility that opposing views are reasonable and correct. While expressions of fallibility in prefaces may achieve a multitude of aims, my proposal is that they at least play this one role.[33]

From the special case of authors and readers, we can make a more general conclusion. When speakers preface their assertions by expressing their fallibility, they do so to acknowledge their corrigibility and articulate the type of interaction they want to have with their interlocutors. Despite this, the reasonableness of these assertions does not imply that speakers are thereby rational to believe that one of their beliefs is false.

---

[32]The pragmatic role of preface claims explains why it would be inappropriate to assert that the book is error-free. This expression of immodesty would misrepresent the type of interaction that the author would like to have with her readers.

[33]Thanks to a reviewer for the observation that this role is just one of many.

References

Arló-Costa, Horacio and Arthur Paul Pedersen (2012). "Belief and probability: A general theory of probability cores." *International Journal of Approximate Reasoning* 53.3, pp. 293–315.

Biber, Douglas and Edward Finegan (1988). "Adverbial Stance Types in English." *Discourse Processes* 11.1, pp. 1–34.

Christensen, David (2004). *Putting Logic in Its Place: Formal Constraints on Rational Belief*. Oxford: Clarendon Press.

Clark, Michael (2002). *Paradoxes from A to Z*. New York: Routledge.

DeRose, K. (2002). "Assertion, knowledge, and context." *The Philosophical Review* 111.2, p. 167.

Douven, Igor (2003). "The preface paradox revisited." *Erkenntnis* 59.3, pp. 389–420.

Downing, Angela (2001). ""Surely you knew!": Surely as a marker of evidentiality and stance." *Functions of Language* 8.2, pp. 251–282.

Evnine, Simon J (1999). "Believing conjunctions." *Synthese* 118.2, pp. 201–227.

Feldman, Richard (2007). "Reasonable Religious Disagreements." In: *Philosophers Without Gods: Meditations on Atheism and the Secular*. Oup, pp. 194–214.

Fitelson, Branden (2012). "Evidence of evidence is not (necessarily) evidence." *Analysis* 72.1, pp. 85–88.

Foley, R. (1993). *Working without a Net*. Oxford University Press.

Fox, Barbara (2001). "Evidentiality: Authority, Responsibility, Entitlement in English Conversation." *Journal of Linguistic Anthropology* 11.2, pp. 167–192.

Gardenfors, P. and N.E. Sahlin (1988). *Decision, probability, and utility: Selected readings*. Cambridge Univ Pr.

Gardenfors, Peter (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge: Mit Pr.

Hall, Ned (1999). "How to set a surprise exam." *Mind* 108.432, pp. 647–703.

Hawthorne, James and Luc Bovens (1999). "The Preface, the Lottery, and the Logic of Belief." *Mind* 108.430, pp. 241–264.

Hoffman, Robert (1973). "The Paradox of the Preface Again." *Mind* 82.327, pp. 441–441.

Kaplan, M. (1983). "Decision Theory as Philosophy." *Philosophy of Science* 50, p. 549.

Klein, Peter (1985). "The virtues of inconsistency." *The Monist* 68.1, pp. 105–135.

Kyburg, Jr Henry E. (1997). "The Rule of Adjunction and Reasonable Inference." *Journal of Philosophy* 94.3, pp. 109–125.

Lacey, AR (1970). "The Paradox of the Preface." *Mind* 79.316, pp. 614–615.

Levi, Issac (1980). *The Enterprise of Knowledge*. Cambridge, MA: MIT Press.

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

New, Christopher (1978). "A Note on the Paradox of the Preface." *Philosophical Quarterly* 28.113, pp. 341–344.

Pinto, Robert (2001). *Argument, Inference, and Dialectic*. Springer.

Pollock, John L. (1986). "The Paradox of the Preface." *Philosophy of Science* 53, pp. 246–258.

Rescher, Nicholas (2001). *Paradoxes: Their Roots, Range, and Resolution*. Chicago: Open Court.

Ross, Jacon and Mark Schroeder (2012). "Belief, Credence, and Pragmatic Encroachment." *Philosophy and Phenomenological Research*.

Ryan, Sharon (1991). "The preface paradox." *Philosophical studies* 64.3, pp. 293–307.

Ryan, Sharon (1996). "The epistemic virtues of consistency." *Synthese* 109.2, pp. 121–141.

Singer, Peter (2010). *The life you can save*. Pan Macmillan.

Sorensen, Roy (1988). *Blindspots*. Oxford: Oxford University Press.

— (2001). *Vagueness and contradiction*. Oxford University Press.

— (2003). *A Brief History of Paradox : Philosophy and the Labyrinths of the Mind*. Oxford, New York: Oxford University Press.

Stalnaker, Robert (1984). *Inquiry*. Cambridge: MIT Press.

— (1999). *Context and Content: Essays on Intentionality in Speech and Thought*. New York: Oxford University Press.

Suppes, Patrick (1960). *Axiomatic set theory*. Princeton, N.J., Van Nostrand.

Weiner, M. (2005). "Must we know what we say?" *The Philosophical Review* 114.

Willett, Thomas (1988). "A crosslinguistic survey of the grammaticalization of evidentiality." *Studies in Language* 12, pp. 51–97.

Williamson, T. (1996). "Knowing and asserting." *Philosophical Review* 105.4, pp. 489–523.

Wittgenstein, Ludwig (1922). *Tractatus Logico-Philosophicus*.