

SAFETY AND THE PREFACE PARADOX

Michael J. SHAFFER

ABSTRACT: In the preface paradox the posited author is supposed to know both that every sentence in a book is true and that not every sentence in that book is true. But, this result is paradoxically contradictory. The paradoxicality exhibited in such cases arises chiefly out of the recognition that large-scale and difficult tasks like verifying the truth of large sets of sentences typically involve errors even given our best efforts to be epistemically diligent. This paper introduces an argument designed to resolve the preface paradox so understood by appeal to the safety condition on knowledge

KEYWORDS: safety, the Preface Paradox, truth, knowledge

1. Introduction

David Makinson discovered the preface paradox and it arises out of a story of the following sort, although there are some variations in the details.¹ Suppose there is an author of a significantly long non-fiction book and that that author is especially diligent in having carefully attempted to establish the truth of every sentence in the book in question. So, on this basis, the author claims to know that every individual sentence in the book is true. The author then reasons, by agglomeration, that she knows that the conjunction of every sentence in the book is true. Suppose further, however, that, based on past experience of error involving non-fiction books composed of large sets of sentences, the author knows also that it is overwhelmingly likely that she has made a mistake somewhere in the book. So, it seems to be the case that the author knows that at least one of the sentences in the book is false. She knows that the disjunction of the denials of every sentence in the book is true. As a result, the author is supposed to know both that every sentence in the book is true

¹ See David Makinson, "The Paradox of the Preface," *Analysis* 25 (1965): 205-207, Doris Olin, *Paradox* (Montreal: Queens-McGill University Press, 2003), ch. 4 and Roy Sorensen, "Epistemic Paradoxes," *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2017/entries/epistemic-paradoxes/>> (2011).

Michael J. Shaffer

and that not every sentence in the book is true. But, this result is paradoxically contradictory. The paradoxicality exhibited in such cases arises chiefly out of the recognition that large-scale and difficult tasks like verifying the truth of large sets of sentences typically involve errors even given our best efforts to be epistemically diligent. This paper introduces an argument designed to resolve the preface paradox so understood by appeal to the safety condition on knowledge.²

2. Knowledge and Safety

The safety condition on knowledge is a necessary condition for knowing that has been most systematically defended by Williamson, Sosa and Pritchard.³ It is supposed to reflect the basic idea of the sort of reliability associated with bona fide knowledge that notably distinguishes it from accidentally true belief. The safety condition can be understood simply as follows:

If A knows that p , then A could not easily have falsely believed that p .

This relatively non-technical gloss on safety and it can be made more precise as follows:

(Safety) $(w_i \models K_A p) \rightarrow \neg [\langle w_i \rangle \models (B_A p \& \neg p)]$.

Here ' $\langle w_i \rangle$ ' is the set of world sufficiently close to w_i and ' $B_A p$ ' represents that A believes that p . So understood, the safety condition is the claim that if A knows that p at w_i , then A does not believe that p when p is false in worlds sufficiently

² This is the so-called "knowledge" version of the paradox, but there are other related paradoxes involving states other than outright knowledge. The most prominent of these other versions involve rational belief that does not rise to the level of knowledge, where rational belief is defined in a qualitative way (as opposed to involving a fixed probability threshold). Clearly then, the knowledge version of the paradox is a restricted case of the rational belief version where the rationality of belief rises to the level of knowledge. This is important because if safety or some weekend correlate of safety applies to rational belief in addition to knowledge, then the solution offered here (or something very like it) may have broader application to the mere rational belief versions of the paradox.

³ Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000), Ernest Sosa, "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999): 141-154, Duncan Pritchard, "Anti-Luck Epistemology," *Synthese* 158 (2007): 277-298, Duncan Pritchard, "Knowledge, Luck, and Lotteries," in *New Waves in Epistemology*, eds. Vincent Hendricks and Duncan Pritchard (London: Palgrave Macmillan, 2008), 28-51, Duncan Pritchard, "Safety-Based Epistemology: Whither Now?" *Journal of Philosophical Research* 34 (2009): 33-45, and Duncan Pritchard, *Knowledge* (London: Palgrave Macmillan, 2009).

similar to w_i . This regimentation captures the core idea of the safety condition well. What is useful here is the contrapositive of safety:

$$\text{(Contrapositive Safety)} [\langle w_i \rangle \models (B_A p \ \& \ \neg p)] \rightarrow \neg(w_i \models K_A p).$$

This version of safety essentially is the assertion that if A could easily have falsely believed that p , then A does not know that p . More technically, it is the claim that if in worlds sufficiently similar to w_i A believes that p and p is false, then A does not know that p at w_i . As noted above, safety has independent merit as a condition on knowledge as it reflects a primitive notion of reliability. One compelling way to deal with what is going on in preface cases is then to appeal to the safety condition on knowledge and to argue that—despite appearances to the contrary—the author in preface case does not, in point of fact, know that the conjunction of every sentence in the book she authored is true. What will ultimately be shown here is that if safety is a necessary condition on knowledge, then there is a clear way out of the preface paradox that does not involve anything radical at all (e.g. subscriptions to dialethism and the like). That is to say, it offers a short path to a straight solution to the paradox by showing that one of the paradox constituting propositions that seems to be true, is, in fact, false.

3. The Safety Solution to the Preface Paradox

Let us then make the presentation of the knowledge version of the preface paradox more precise and see how attention to the safety condition results in this sort of straight resolution of that paradox. Where b_1, b_2, \dots, b_n are the sentences that constitute a non-fiction book authored by A and where n is sufficiently large we can generate the preface paradox as follows:

$$\text{(PP1) for all } n, K_A(b_n).^4$$

This is simply the claim that A knows that every sentences in the book is true. This amounts to the following claim for a book with n sentences:

$$\text{(PP2) } K_A b_1 \ \& \ K_A b_2 \ \& \dots \ \& \ K_A b_n.$$

By agglomeration this implies:

$$\text{(PP3) } K_A(b_1 \ \& \ b_2 \ \& \dots \ \& \ b_n).$$

⁴ Of course, what is really known are the propositions expressed by these sentences. For the purposes of fidelity to typical presentations of the paradox we can ignore this little complication.

Michael J. Shaffer

In other words, A knows that every individual sentence in the book is true and so knows that the conjunction of sentences constituting the book is true. On the other hand, based on good evidence about our fallibility in general and specifically about our fallibility in preface-like cases involving such complex tasks, we also have the following principle:

(PP4) $K_A(\neg b_1 \vee \neg b_2 \vee \dots \vee \neg b_n)$.

In other words, based on A 's past performance with respect to tasks like the one in question, A knows that there is at least one false sentence in the book. But, PP3 and PP4 are contradictory and so we have a paradox.

As Olin points out, the basic nature of a paradox is that it involves a set of propositions A each of which is *prima facie* reasonable to endorse, but where (in the context of background knowledge Σ) the set A appears to imply a contradiction.⁵ So paradoxes are essentially sets of propositions that appear to be individually rationally endorsable but which cannot collectively be endorsed. This can be because the set A is itself internally inconsistent or because A appears to imply some proposition p and Σ implies $\neg p$. Let us refer to a given set A_i as the *paradox constituting propositions* of paradox i . We can then also present paradoxes as deductive arguments where the members $\lambda_1, \lambda_2, \dots, \lambda_n$ of a given set A are the premises and where they either appear to directly imply $(p \ \& \ \neg p)$ or where A appears to imply p and Σ implies $\neg p$. So in this case {PP3, PP4} is a paradox generated by the preface paradox story. Safety is part of our background theory of knowledge and in order to resolve the paradox in a straight manner one or more of PP3 and PP4 has to be given up.

The safety solution to the knowledge version of the preface paradox then involves the recognition that we ought to accept safety and PP4 but reject PP3, thus resolving the paradox. Safety helps to explain why it is not the case that $K_A(b_1 \ \& \ b_2 \ \& \dots \ \& \ b_n)$ even where the author has been diligent in checking each sentence in the book. This is because A could easily have falsely believed that $(b_1 \ \& \ b_2 \ \& \dots \ \& \ b_n)$ where n is large, and so A does not really know that $(b_1 \ \& \ b_2 \ \& \dots \ \& \ b_n)$. This is easily seen by noting that there are clearly many conceivable close possible worlds where the author believes $(b_1 \ \& \ b_2 \ \& \dots \ \& \ b_n)$ on the basis of her careful and diligent attempts to verify each b_n but where, nevertheless, $(b_1 \ \& \ b_2 \ \& \dots \ \& \ b_n)$ is false because one or more of the sentences in the book is false as per PP4. This is simply because we are fallible knowers, especially in the case of complex tasks like verifying

⁵ See Olin, *Paradox*.

the truth of large bodies of sentences. But that means that the author *A* does not in fact know the conjunction of the set of sentences that constitute the book in question despite her best efforts to verify every sentence individually. The belief that (b_1 & b_2 & ... & b_n) is unsafe and, again, this will be true for every such preface case where n is sufficiently large.