Erkenn (2017) 82:169–183
DOI 10.1007/s10670-016-9811-0

ORIGINAL ARTICLE

Fallibilism, Verisimilitude, and the Preface Paradox

Gustavo Cevolani[1]

Received: 23 July 2015 / Accepted: 8 April 2016 / Published online: 20 April 2016

Abstract The Preface Paradox apparently shows that it is sometimes rational to believe logically incompatible propositions. In this paper, I propose a way out of the paradox based on the ideas of fallibilism and verisimilitude (or truthlikeness). More precisely, I defend the view that a rational inquirer can fallibly believe or accept a proposition which is false, or likely false, but verisimilar; and I argue that this view makes the Preface Paradox disappear. Some possible objections to my proposal, and an alternative view of fallible belief, are briefly discussed in the final part of the paper.

1 Introduction

The author of a book apologizes in the preface for the errors that doubtless will be found in his work, and yet remains committed to all the assertions made in the volume. On the one hand, it seems that the author holds logically incompatible beliefs, and hence that he is irrational. On the other hand, the author seems rationally entitled to hold the specific incompatible beliefs under consideration: thus, the so called "Preface Paradox" arises. In this paper, I propose a way out of the paradox based on the ideas of fallibilism and verisimilitude (or truthlikeness). More precisely, I defend the view that a rational inquirer can fallibly believe or accept a proposition which is false, or likely false, but verisimilar; and I argue that this view makes the Preface Paradox disappear.

In Sect. 2 I present the paradox and outline my proposal. The details are elaborated in Sect. 3 and Sect. 4, where the ideas of fallible belief and of (expected)

✉ Gustavo Cevolani
g.cevolani@gmail.com; gustavo.cevolani@unito.it

1 Department of Philosophy and Education and Center for Logic, Language, and Cognition, University of Torino, via Sant'Ottavio 20, 10124 Turin, Italy

verisimilitude are presented, respectively. Following theorists like Kuipers (2000), Schurz and Weingartner (2010), Oddie (2014), and especially Niiniluoto (1987, 1999), I introduce a notion of fallible belief according to which the author in the Preface Paradox is rationally entitled to believe the contents of his book; I also prove a formal result corroborating this conclusion. Section 5 concludes the paper with a discussion of some possible objections to my proposal. The connections among the notions of belief, acceptance, and probability are here taken into account, and a recent solution to the paradox advanced by Leitgeb (2014a) is also considered. Finally, the "Appendix" contains a formal presentation of the concepts employed through the paper, and a proof of its main result.

2 The Preface Paradox: Taking Fallibilism Seriously

The Preface Paradox is usually traced back to a short 1965 paper by David Makinson, who presents it as follows (Makinson 1965, p. 205, notation modified):

> Suppose that in the course of his book a writer makes a great many assertions, which we shall call $b_1, b_2, \ldots, b_m$. Given each one of these, he believes that it is true. If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true. His approach is eminently rational; he has learnt from experience. […]
>
> However, to say that not everything I assert in this book is true, is to say that at least one statement in this book is false. That is to say that at least one of $b_1, b_2, \ldots, b_m$ is false, where $b_1, b_2, \ldots, b_m$ are the statements in the book; that $(b_1 \wedge b_2 \wedge \ldots \wedge b_m)$ is false; that $\neg(b_1 \wedge b_2 \wedge \ldots \wedge b_m)$ is true. The author who writes and believes each of $b_1, b_2, \ldots, b_m$, and yet in a preface asserts and believes $\neg(b_1 \wedge b_2 \wedge \ldots \wedge b_m)$ is, it appears, behaving very rationally. Yet clearly he is holding logically incompatible beliefs: he believes each of $b_1, b_2, \ldots, b_m, \neg(b_1 \wedge b_2 \wedge \ldots \wedge b_m)$, which form an inconsistent set. The man is being rational though inconsistent.

Why is the author "eminently rational" in believing "that not everything he has written in his latest book is true"? And why do many writers actually insert in their books similar "prefatory statements" to the same effect? The reason, I submit, is that they are willing to plainly acknowledge their own fallibility, and the tentative nature of their conclusions. "It is, after all, our human fallibility that is to blame for the fact that we cannot expect to tell a story of any significant ambition (that purports to describe how the world is) without having told a story that is most probably false", as Kaplan (2013, p. 29) puts it. And this is so widely acknowledged indeed, that authors are not even required to explicitly mention this plain fact in the prefaces to their works and, accordingly, many do not do so (as noted by Leitgeb 2014a, p. 15).

In what follows, I propose a solution to the Preface Paradox based on the widely shared idea that all ordinary and scientific human knowledge is fallible. In a nutshell, I argue that the author of the book actually accepts b—as we shall call the conjunction of the statements $b_1, b_2, \ldots, b_m$ in the book.[1] Accordingly, he doesn't accept $\neg b$, as the prefatory statement may suggest. Instead, what the author is saying in the preface is just that b is his best attempt to approximate the truth about the relevant subject matter. In other words, the author fallibly accepts b as the most "verisimilar" or "truthlike" statement at his disposal given the available evidence; and this is compatible with also believing that b will turn out to be false, or that it has already been falsified.

Philosophers of science are familiar with this kind of epistemic situation from their discussions of progress, fallibilism and scientific realism. In fact, some of them have occasionally emphasized that the Preface Paradox illustrates a typical case of fallible scientific knowledge. For instance, in his entry on "Fallibilism" for the *Routledge Encyclopedia of Philosophy*, Nicholas Rescher (1998) writes:

> We learn by empirical inquiry about empirical inquiry, and one of the key things we learn is that at no actual stage does science yield a final and unchanging result. We have no responsible alternative to supposing the imperfection of what we take ourselves to know. […] We occupy the predicament of the "Preface Paradox" exemplified by the author who apologizes in his preface for those errors that have doubtless made their way into his work, and yet blithely remains committed to all those assertions he makes in the body of the work itself. We know or must presume that (at the synoptic level) there are errors, though we certainly cannot say where and how they arise.

A similar point is made by Kitcher (2001, pp. 170–171), who, while defending the realist viewpoint about scientific progress, notes in passing:

> If we are going to make an induction on the history of science, then we seem less warranted in antirealist pessimism than in the conclusion that we're approximately right about most of what we claim in our most successful theories; yet, since it's overwhelmingly likely that there are errors we've failed to detect, our acceptance of the whole should be tempered by consciousness of our own fallibility. Our predicament is like that of the author who confesses in her preface that she is individually confident about each main thesis contained in her book but equally sure that there's a mistake somewhere.

Being conscious of his own fallibility, a rational inquirer is virtually certain that some of his beliefs are false. Still, he is typically unable to correct them, at least for the time being, since usually he cannot say exactly what is wrong with them. Otherwise, he would entertain a different set of beliefs: if our author knew which statements b_i are wrong in his book, he would not accept b, but the conjunction of

[1] I'm using "belief" and "acceptance" as synonymous here, postponing to Sect. 5 a discussion of the purported distinction between these two notions.

the remaining claims.[2] In sum, once the fallibility of human knowledge is taken into account, the situation of the preface author is perhaps puzzling, but not paradoxical after all. To make sense of it, however, one needs to explain what fallible belief is, a task that I shall undertake in the next two sections.

3 Strong Fallibilism and Verisimilitude

According to Peirce, who first introduced the term into contemporary philosophical discussion, fallibilism is the thesis that "people cannot attain absolute certainty concerning questions of fact" (*CP* 1.149, in Hartshorne et al. 1931–1958). Accepted as it is by the vast majority of contemporary thinkers, fallibilism comes in various forms, with appreciably different implications (Hetherington 2005). In any case, all these variants constitute, as it were, a *continuum* between two extreme epistemological positions, which are clearly distinguishable from fallibilism. The first is infallibilism, i.e., the classical idea that certainty is achievable, and even necessary for knowledge (meant as *episteme*). The second is skepticism, the view that knowledge is plainly impossible or at least unattainable.

As far as different fallibilist positions are concerned, Niiniluoto (1984, ch. 3) has convincingly argued for a distinction between "strong" and "weak" (variants of) fallibilism. In short, "weak" fallibilists (like Bayesians, for instance) maintain that scientific hypotheses and ordinary beliefs are uncertain but probably true; according to "strong" fallibilists (like Popper), they are instead typically false, but truthlike or verisimilar (cf. Niiniluoto 1999, p. 13 and Sec. 4.1). The relevant notion of verisimilitude (or truthlikeness or approximation to the truth) was first introduced and systematically investigated by Popper (1963, ch. 10) himself, in the attempt to back up his realist conception of scientific progress (see Niiniluoto 2015; Cevolani and Tambolo 2013 for an assessment of recent developments along this line of inquiry). Popper maintained that all scientific statements, including our most successful theories and hypotheses, are typically not literally true but at best "close" or "similar" to the truth, and only accepted as such by the scientists. This implies that even in our most successful cognitive endeavor, we believe statements which are, strictly speaking, false, and we are rational in doing so.

As Popper (1963, p. 237) puts it, verisimilitude "represents the idea of approaching comprehensive truth [and] thus combines truth and content". In spite of other technical differences among them, all the various accounts of verisimilitude

[2] In this connection, Hansson (2013) highlights an interesting link between the Preface Paradox and the theory of belief revision, which also originated in the work of Makinson, together with Carlos Alchourrón and Peter Gärdenfors, in the eighties of the past century. Hansson notes that some belief changes are too complex to be performed, since they would exceed the capacities of real, cognitively-limited rational agents. Discussing the case of "belief contraction"—i.e., how to give up a previously entertained belief—he writes: "The problem of overly complex contractions was foreshadowed in David Makinson's preface paradox [...]. The author in Makinson's example [...] has reasons to contract by [*b*] but refrains from doing so since such a contraction would be cognitively unmanageable. [...] In cognitive terms, the agent may be described as being aware of a solution that goes beyond the reach of her abilities, for which reason she postpones the decision" (Hansson 2013, pp. 1024–1025). Hansson's suggestion is explored in more details in Cevolani (2016).

currently on the market revolve around the insight that high verisimilitude requires a "balance" of truth and information content (for a survey see Oddie 2014). An important consequence of such insight is that a false sentence may be more verisimilar than a true one. In fact, the latter may be too weak or "cautious" to be verisimilar, whereas a very informative or "bold" statement may be highly verisimilar, although false. For instance, Newton's theory, although strictly speaking false, seems clearly more verisimilar than the true, but much less informative, statement that planets orbit around the sun.

This explains why it is sometimes rational to accept a given statement, which the available evidence indicates as the most verisimilar at disposal, even if it is very likely that it will turn out to be false. Indeed, as philosophers of science have repeatedly emphasized, it is not unusual for scientists to believe a theory which is known to be false, for instance because it is highly idealized or because it faces some, still unresolved, anomaly. In particular, a hypothesis may be estimated as highly verisimilar on some evidence even if this same evidence proves it wrong, i.e., falsifies it (see, e.g., Niiniluoto 1999, p. 98). The technical work on verisimilitude by, among others, authors like Niiniluoto (1987, 1999, 2015) and Kuipers (2000, 2015) gave rise to a thoroughly fallibilist, and yet robustly realist, view of rational acceptance aiming at truth approximation, centered around the notion of fallible rational belief. Such notion, I submit, provides a straightforward way out of the Preface Paradox.

4 Fallible Belief and Expected Verisimilitude

The insight defended here is that, by publishing the book, the author accepts b as the most verisimilar sentence at his disposal given the available evidence, while acknowledging in the preface that b may be, or even is, false. To put it differently, what the author means by saying in the preface that the book will contain some error is not that he believes $\neg b$. Rather, the prefatory statement only emphasizes that b is his best attempt to approximate the truth about the target domain, and, as such, may (turn out to) be false. Thus, the situation of the author is not different from that of a researcher who accepts a highly successful, but presumably false or even falsified, scientific theory. Since the notion of verisimilitude was introduced by Popper as a cornerstone of his defense of a realist and fallibilist view of scientific inquiry, it comes as no surprise that it can be used to understand how a proposition can be fallibly and rationally accepted.

For the sake of clarity, let me present here the main argument and result, and postpone the formal details to the "Appendix". I will assume that the author, as a rational inquirer, aims at approaching the comprehensive truth about the subject matter of the book. According to my reconstruction, by asserting the statements $b_1, b_2, \ldots, b_m$ in the book the author makes clear that he believes their conjunction b as the statement which is estimated as the closest to the truth, given the available evidence. This can be made a little more precise as follows. Let $vs(h)$ denote a measure of the actual verisimilitude or closeness to the truth of each statement h of the underlying language (see the "Appendix" for details). Moreover, suppose that a subjective probability distribution p expresses the author's rational degree of belief $p(h|e)$ in the truth of h, given some evidence e. On the basis of such a distribution, the author can then form an estimate of the actual verisimilitude of h, expressed as the degree $Evs(h|e)$ of expected verisimilitude of h on e. In turn, this notion of expected verisimilitude can be employed in the definition of the following rule of rational acceptance (Niiniluoto 1987, p. 416):[3]

Accept on evidence e the statement h which maximizes the value $Evs(h|e)$.

If, as I'm assuming here, the author follows the above rule, then by publishing the book he makes clear that b maximizes expected verisimilitude given the evidence available to him. In other words, according to the author $Evs(b|e)$ is greater than $Evs(h|e)$, for any possible alternative statement h. In this connection, it is worth noting again that $Evs(h|e)$ can be high even if $p(h|e)$ is low, or even zero, i.e., if the evidence falsifies h (Niiniluoto 1987, p. 274). Thus, no contradiction arises with the fact that, in the preface, the author acknowledges his own fallibility and emphasizes that b could be unlikely or even false. In such a way, the paradoxical impression raised by the prefatory statement is explained away.

An even stronger form of the argument just presented becomes viable if one focuses on some special cases. In particular, suppose that all claims made in the book are "basic" in the sense that $b_1, b_2, \ldots, b_m$ are atomic, logically independent statements about the domain under inquiry, or negations of such statements. In the present context, this assumption doesn't appear as particularly problematic; indeed, it is often made, at least implicitly, in many presentations of the Preface Paradox. For instance, it is customary to say that the author's book is one on history (cf. e.g. Christensen 2004; Foley 2009), apparently implying that it contains a great number of independent assertions on many different specific events; similarly, Leitgeb (2014a, pp. 11–12) considers the case of a "scientific lab publishing a database of a great many experimental results". In both cases, it seems safe to assume that the relevant claims in the book are basic statements in the defined sense.

Under the condition just specified, one can then show that $Evs(b|e)$ is maximal when b is the conjunction of all and only the basic statements b_i which are sufficiently probable according to the author (see the "Appendix" for a proof):

Theorem 1 *There is a threshold value σ such that, if b is the conjunction of all and only the b_i for which $p(b_i) > \sigma$, then $Evs(b)$ is maximal.*

In other words, if the author aims at maximizing expected verisimilitude, and he is sufficiently certain of the truth of each of the claims made in the book, then b is indeed the theory that he should accept, given his rational assessment of the available evidence as conveyed by the probability assigned to each b_i. Accordingly, Theorem 1 provides a straightforward way out of the paradox, showing how the author can rationally believe b even if b is improbable on e, or even if it is false.

[3] The reader acquainted with the research tradition inaugurated by Levi (1967) will note that this rule is a way of articulating a verisimilitude-based variant of (Bayesian) cognitive decision theory. See Niiniluoto (1987, ch. 12) for a full treatment of (expected) verisimilitude as a cognitive or epistemic utility, and the next section for relevant discussion.

5 Concluding Remarks: Belief, Acceptance, and Probability

How verisimilar is my verisimilitude-based solution of the Preface Paradox? I can anticipate at least two different kinds of objection to the present proposal. The former is of a more technical nature, and can be immediately presented, and answered, as follows. As was made clear in the foregoing section, Theorem 1 holds only under the assumption that the claims made in the book are basic statements in the sense defined there. Admittedly, no result as simple as Theorem 1 can be easily proved in the more general case where those claims are arbitrary statements. Still, both the notion of estimated verisimilitude and the acceptance rule based on maximizing expected verisimilitude can be adequately defined for all kinds of statements (see especially Niiniluoto 1987). Thus, even lacking results as strong as Theorem 1, the central insight that the preface author accepts b as the most verisimilar statement at disposal given the available evidence remains unaltered. In this sense, the suggested way out of the Preface Paradox is fully general.

The second kind of worry is more substantial, and has to do with the notion of belief itself. Most readers will feel uncomfortable with the idea that it is possible to believe propositions which one is pretty sure are false. However, this very idea is the core of the strongly fallibilist position that I have outlined in Sect. 3 and adopted in this paper. Of course, one can plainly reject this idea, and argue that what I am speaking about is not "belief" at all, but some form or other of (rational) "acceptance". Accordingly, one can opt for a weaker form of fallibilism and maintain that, while it is possible to rationally accept unlikely or even certainly false propositions, it is only possible to believe certainly true or at least highly probable ones. A full discussion of all the implications of this move is beyond the scope of this paper; still, it will be useful to quickly survey the main reasons why I find such an alternative unconvincing. I will start by briefly discussing the distinction between belief and acceptance in the present context; then, I will recall the problems of the "high probability" view of belief, and consider a solution of the Preface Paradox proposed by Leitgeb (2014a) as based on that view.

5.1 Belief and Acceptance

The notions of belief, on the one hand, and acceptance, on the other, have been distinguished along different dimensions.[4] For instance: acceptance, but not belief, would be voluntary in the sense that an inquirer can decide whether or not to accept some proposition, which is not the case for belief; acceptance is an all-or-nothing matter (an inquirer either accepts or doesn't accept a proposition) while belief can be graded; and so on. Perhaps the most crucial difference, however, is that belief would aim at truth while acceptance would aim at some other, pragmatic value, like utility or success.

As an example, an engineer may not believe Newtonian mechanics as a true picture of how the world is, but may still accept it as far as it can be usefully employed in calculations to design and build a bridge. According to this view, a

[4] The distinction between belief and acceptance has been elaborated by a number of philosophers, perhaps most prominently by Cohen (1992); Tuomela (2000) contains a still useful survey of some of the main positions in the debate.

literally false theory (like Newtonian mechanics) can be accepted, for pragmatic reasons, as a useful and successful tool; but it has no proper "cognitive" or "epistemic" value, at least insofar as cognitive values are truth-related theoretical utilities. And for this latter reason, one cannot believe a theory one knows or believes to be false, its pragmatic virtues notwithstanding.

However important this difference between belief and acceptance may be in other contexts, it seems irrelevant to me in the present one. In fact, acceptance as defined here cannot be distinguished from belief by saying that the latter, but not the former, would aim at truth. The reason is that the acceptance rule based on maximizing expected verisimilitude makes acceptance a truth-related notion exactly as belief is. In this sense, belief and acceptance become indistinguishable here, at least insofar as truth as the aim of inquiry is concerned. As Popper (1963, p. 236) made clear long ago, the two notions of verisimilitude and (high) probability need to be carefully distinguished exactly because they are both truth-related theoretical values:

> The differentiation between these two ideas [verisimilitude and probability] is the more important as they have become confused; because both are closely related to the idea of truth, and both introduce the idea of an approach to the truth by degrees. […] Logical probability […] represents the idea of approaching logical certainty […] Verisimilitude, on the other hand, represents the idea of approaching comprehensive truth.

In other words, while a probable statement is likely true in a given domain, a verisimilar one is like or close to the whole truth about the domain. Thus, if a distinction between belief and acceptance is relevant here, it cannot be based on the idea that belief is guided by truth-related, epistemic values, while acceptance is not.[5] In conclusion, I see no reason why a rational inquirer could not believe a false but close-to-the-truth proposition, as he believes a probable or likely true one. By accepting a statement which maximizes expected verisimilitude, an inquirer will, for this very same reason, aim at truth as the main cognitive goal of inquiry.

5.2 Belief and High Probability

Why should one invoke the distinction between belief and acceptance, which, as I argued above, isn't really relevant in the present context? The main reason, I think, is that one would like to keep the intuitive link between belief and high probability

[5] This is not to deny that one can discriminate these two notions in a meaningful way. What I'm saying is simply that, if belief aims at truth while acceptance doesn't, then belief, and not acceptance, is the relevant notion for the strongly fallibilist view put forward here (I thank an anonymous referee for pressing me to clarify this point). Admittedly, a full defense of such view would require a separate paper. In what follows, I shall limit myself to highlighting some differences between the verisimilitude-based and the probability-based views of rational belief (or acceptance). Cognitive decision theorists like Levi (1967), Niiniluoto (1987, ch. 12), and Maher (1993, ch. 6) have systematically explored these notions and provided useful discussions of their interplay with truth and other pragmatic and cognitive values. Recent work in so called accuracy-first epistemology (Pettigrew 2016) has revived this line of inquiry under the heading of "epistemic utility theory"; connections with the research program on truthlikeness are studied in Oddie (2015).

intact. This link is the core intuition behind the weak variant of fallibilism, according to which scientific and ordinary beliefs are typically uncertain but probably true. It is then useful to see how a "purely probabilistic" analysis of the Preface Paradox would run (cf. Easwaran 2015, Sec. 1.2).

Suppose that our author has a high degree of belief in each of the claims $b_1, b_2, \ldots, b_m$ made in the book. If such degrees of belief are expressed as degrees of probability, then the probability calculus alone prescribes that the conjunction b of those claims cannot be more probable than any of them. But this only means that the author will have a low degree of belief in b or, equivalently, a high degree of belief in its negation, i.e., in the prefatory statement $\neg b$. Thus, no paradox arises, if we refrain from talking about belief or acceptance at all, and only take into account the degrees of belief, or credences, expressed by the relevant probabilities.

The price for this quick (dis)solution of the paradox is making belief-talk simply meaningless. Some find this price too high, and argue that a notion of "all-or-nothing" or "qualitative" belief is required in order to make sense of the concept of belief or acceptance commonly used in both scientific and ordinary reasoning. The so-called Lockean thesis is the most natural way of providing such a notion (see, e.g., Foley 2009; Leitgeb 2014b). According to this thesis, the propositions accepted by a rational agent are just the ones which are assigned a high degree of belief: in other words, one should rationally believe some proposition h if and only if one's assessment of $p(h)$ is greater than (or equal to) some stipulated threshold (between 0.5 and 1).

Unfortunately, it is well-known that the Preface Paradox can be construed exactly as an argument against the idea that high probability and belief are connected in this way. More precisely, the Preface Paradox suggests that a high probability value is not necessary for belief, thus blocking one half of the Lockean thesis (the one saying that believing h entails that $p(h)$ is high). This can be seen as follows. Suppose that the author's beliefs are logically closed; in particular, that they are closed under conjunction, so that if the author accepts $b_1, b_2, \ldots, b_m$ then he also accepts b. Then, there is some proposition (namely, b) which is believed although its probability is small. Moreover, as m increases, $p(b)$ can become arbitrarily small, even if the probability of all b_i is very high; consequently, $p(b)$ can be made smaller than any value that one may want to propose as the relevant threshold. In particular, a value of this threshold greater than 0.5 is not necessary for rational belief.[6]

A number of scholars have sought a way out from this puzzling situation by rejecting some highly general background assumptions behind the Preface Paradox. For instance, Foley (2009) rejects the idea that the beliefs of a rational inquirer should be closed under conjunction, and Christensen (2004) that they have to be logically consistent. The solution proposed in this paper amounts instead to

[6] For instance, suppose that the $b_1, b_2, \ldots, b_m$ are probabilistically independent and that $p(b_i) = 0.9$ for all b_i. Then the probability of their conjunction is $p(b) = 0.9^m$, which quickly tends to zero as m increases. Similarly, the well-known Lottery Paradox shows that no value of the threshold smaller than 1 is sufficient for rational belief. As Foley (2009, p. 39) notes, the Lottery Paradox and the Preface Paradox "create a pincer movement on the Lockean thesis" and on the supposed connection between high probability and belief. See also Maher (1993, Sec. 6.2.4) who makes this same point without explicit reference to the Preface Paradox.

rejecting one of the premises of the paradox, i.e., that the author really accepts the negation of b, the conjunction of all statements in the book. According to my analysis, the author believes b, even if he has reasons to suppose that b is false, and rejects $\neg b$, even if he has reason to suppose it true. Thus, the prefatory statement should not be read as meaning that the author believes the negation of at least one claim in the paper (he does not, since he accepts their conjunction) but that he accepts b just as the most verisimilar, if likely false, statement at disposal given the evidence.

Interestingly, the opposite route of rejecting the other premise of the paradox—i.e., that the author accepts b—has been recently proposed by Leitgeb (2014a). He suggests that, by publishing the book, the author does not accept b, but only the weaker claim that "the vast majority" of $b_1, b_2, \ldots, b_m$ are true. Of course, this is logically compatible with asserting, at the same time, that at least one of them is false. More precisely, let k be greater than 0 and smaller than m, but "sufficiently close" to m. According to Leitgeb (2014a, p. 12), what the author accepts by publishing the book is not b itself, but its "statistical weakening" $S_k(b)$, defined as the disjunction of all the conjunctions of k different statements among $b_1, b_2, \ldots, b_m$.[7] From my point of view, $S_k(b)$ is too weak a statement to be accepted by the author as a good approximation to the truth, since it doesn't provide enough information about the target domain and hence cannot be (expected to be) highly verisimilar.[8]

Leitgeb's solution is in line with the high-probability view of belief, since $p(S_k(b))$ can be high even if $p(b)$ is low (Leitgeb 2014a, p. 14). In this sense, his proposal can be seen as a way of articulating a weakly fallibilist way out of the Preface Paradox, as opposed to the strongly fallibilist one defended in this paper. Of course, I don't claim to have convinced the reader that the latter position, and the corresponding verisimilitude-based notion of fallible rational belief, is correct and without problems. What I hope to have shown is that this position is defensible and, if correct, would provide a straightforward way out of the Preface Paradox.

Acknowledgements I would like to thank Vincenzo Crupi, Roberto Festa, Theo Kuipers, Hannes Leitgeb, Diego Marconi, Graham Oddie, Gerhard Schurz, and Luca Tambolo for useful feedback on a previous draft of this paper, and one anonymous reviewer for exceptionally helpful comments. Financial support from the FIRB project "Structures and Dynamics of Knowledge and Cognition" (Italian Ministry of Scientific Research, Turin unit, D11J12000470001) and from the project "Assessing information models: exploring theories and applications of optimal information search" (University of Turin and Compagnia di San Paolo, D16D15000190005) is gratefully acknowledged.

[7] For instance, if $m = 4$ and $k = 3$, then the statistical weakening of $b = b_1 \wedge b_2 \wedge b_3 \wedge b_4$ is

$$S_3(b) = (b_1 \wedge b_2 \wedge b_3) \vee (b_1 \wedge b_2 \wedge b_4) \vee (b_1 \wedge b_3 \wedge b_4) \vee (b_2 \wedge b_3 \wedge b_4).$$

The actual value of k is highly context-dependent and does not need to be explicitly stated, not even by the author of the book (Leitgeb 2014a, pp. 12, 14). In any case, it seems that k should be at least greater than $\frac{m}{2}$.

[8] Nor is $S_k(b)$ a good approximation to b itself. One may want to say that $S_k(b)$ "approximates" b in the sense that it says that k of the m claims in b are true, and k is close to m. However, there are many other sentences which are weaker than b but closer to it than $S_k(b)$. In particular, each of the disjuncts of $S_k(b)$ is a better approximation to b than $S_k(b)$ itself, since each such disjunct provides much more correct information about b than $S_k(b)$ (for a more detailed analysis, see Cevolani 2016).

Appendix

In this section, I provide formal definitions of all the notions used throughout the paper, and a proof of Theorem 1 from Sect. 4.

To keep things simple, suppose that the domain under inquiry is described by a propositional language with n atomic sentences $a_1, a_2, \ldots, a_n$.[9] These atomic statements a_i and their negations $\neg a_i$ are called the "basic" statements of the language, and describe what we may call the "basic features" of the underlying domain. If an inquirer is only interested in the basic features of the world, his beliefs can be represented by a "basic theory"—i.e., the strongest (non-contradictory) conjunction of basic statements that he is willing to accept. The strongest possible conjunctions of this kind are the so-called state descriptions—or (propositional) constituents—of the language, which are the most informative descriptions of a possible state of affairs (a "possible world") within the conceptual resources of the language. Each constituent c_i can be written as a conjunction of n atomic sentences, negated ($-$) or not ($+$), as follows:

$$\pm a_1 \wedge \pm a_2 \wedge \ldots \wedge \pm a_n.$$

One can check that there are 2^n constituents, and that only one of them, call it $c_\bigstar$, is true. Thus, $c_\bigstar$ represents "the whole truth" about the domain, since it is the complete description of the actual world as expressed in the language. Accordingly, the verisimilitude of any statement, hypothesis or theory h can be expressed by an adequate measure of the closeness or similarity of h to $c_\bigstar$. If h is a basic theory in the sense defined above—i.e., a conjunction of k basic statements, with $k \leq n$—then there is a simple way to define such a measure.

Let c_i be an arbitrary constituent and let $T(h, c_i)$ be the set of "matches" between h and c_i, i.e., of the conjuncts that h and c_i have in common or, so to speak, of the conjuncts of h which are "true in" c_i. Similarly, $F(h, c_i)$ will denote the set of "mismatches" between h and c_i, i.e., of the conjuncts of h whose negation is true in c_i. Then, the similarity or closeness of h to c_i can be defined as the weighted difference of the normalized number of matches and mismatches between h and c_i:

$$s_\phi(h, c_i) = \frac{|T(h, c_i)|}{n} - \phi \frac{|F(h, c_i)|}{n} \tag{1}$$

where $\phi > 0$. Intuitively, different values of ϕ reflect the relative weight assigned to truths and falsehoods: the greater ϕ, the less similar h is to c_i due to the mismatches in $F(h, c_i)$. The verisimilitude of h can then be defined as the similarity of h to the true constituent $c_\bigstar$:

$$vs_\phi(h) = s_\phi(h, c_\bigstar) = \frac{|T(h, c_\bigstar)|}{n} - \phi \frac{|F(h, c_\bigstar)|}{n} \tag{2}$$

Thus, $vs_\phi(h)$ is maximal (and equals 1) just in case h is the truth $c_\bigstar$.

[9] I'm adopting here the so called basic feature approach to verisimilitude developed by Cevolani et al. (2011); however, as mentioned at the beginning of Sect. 5, the crucial intuition behind my proposal can be extended to basically all other existing accounts.

Since the truth is usually unknown, one cannot use Eq. 2 to calculate the actual degree of verisimilitude of h. However, if a probability distribution p is defined on the set of constituents of the language—such that $p(c_i)$ expresses the rational degree of belief of the inquirer in the truth of c_i—, then the expected verisimilitude of h can be defined as the expected value of $vs_\phi(h)$:

$$Evs_\phi(h) = \sum_{c_i} s_\phi(h, c_i) \times p(c_i) \tag{3}$$

In words, $Evs_\phi(h)$ expresses the inquirer's best estimate of the actual verisimilitude of h given the available evidence.

The main formal result of this paper is Theorem 1 from Sect. 4: if the claims $b_1, b_2, \ldots, b_m$ in the book are basic statements and b their conjunction, there is a threshold value σ such that, if $p(b_i) > \sigma$ for all b_i, then $Evs_\phi(b)$ is maximal. In the following, x and y will denote arbitrary basic statements of the language, and h an arbitrary basic theory in the sense defined above.

Let us first note that the similarity measure defined in Eq. 1 is additive in the sense that:

$$s_\phi(h, c_i) = \sum_{x:h \vDash x} s_\phi(x, c_i); \tag{4}$$

i.e., the similarity of h to c_i is just the sum of the similarities of the conjuncts x of h to c_i (in fact, note that $s_\phi(x, c_i) = \frac{1}{n}$ if x is true in c_i, and $s_\phi(x, c_i) = -\frac{\phi}{n}$ otherwise). It follows from this that the expected verisimilitude measure defined in Eq. 3 is also additive in the same sense:

Lemma 1 $Evs_\phi(h) = \sum_{x:h \vDash x} Evs_\phi(x)$

Proof

$$\begin{aligned}
Evs_\phi(h) &= \sum_{c_i} p(c_i) s_\phi(h, c_i) \\
&= \sum_{c_i} p(c_i) \sum_{x:h \vDash x} s_\phi(x, c_i) \quad \text{by eq. (4)} \\
&= \sum_{c_i} \sum_{x:h \vDash x} p(c_i) s_\phi(x, c_i) \\
&= \sum_{x:h \vDash x} \sum_{c_i} p(c_i) s_\phi(x, c_i) \\
&= \sum_{x:h \vDash x} Evs_\phi(x) \quad \text{by eq. (3)}
\end{aligned}$$

The following theorem specifies under what conditions the conjunction of h with an arbitrary basic statement y (not already a conjunct of h) has greater expected verisimilitude than h itself.

Lemma 2 $Evs_\phi(h \wedge y) > Evs_\phi(h)$ iff $p(y) > \frac{\phi}{\phi+1}$.

Proof

$$
\begin{array}{lll}
 & Evs_\phi(h \wedge y) \gtreqless Evs_\phi(h) & \\
\text{iff} & Evs_\phi(h) + Evs_\phi(y) \gtreqless Evs_\phi(h) & \text{by lemma 1} \\
\text{iff} & Evs_\phi(y) \gtreqless 0 & \\
\text{iff} & \sum_{c_i} p(c_i) s_\phi(y, c_i) \gtreqless 0 & \text{by eq. (3)} \\
\text{iff} & \left(\sum_{c_i \vDash y} p(c_i) \times \frac{1}{n}\right) + \left(\sum_{c_i \vDash \neg y} p(c_i) \times -\frac{\phi}{n}\right) \gtreqless 0 & \\
\text{iff} & \frac{1}{n} p(y) - \frac{\phi}{n} p(\neg y) \gtreqless 0 & \\
\text{iff} & p(y) - \phi(1 - p(y)) \gtreqless 0 & \\
\text{iff} & p(y) \gtreqless \frac{\phi}{\phi + 1} &
\end{array}
$$

Finaly, from Lemma 2, the Proof of Theorem 1 from Sect. 4 easily follows.

Proof Let be $\sigma = \max(\frac{\phi}{\phi+1}, 0.5)$. For the sake of conciseness, let us say that a basic statement x is "likely" iff $p(x) > \sigma$ and "unlikely" otherwise (i.e., iff $p(x) \leq \sigma$). Suppose that b is the (consistent) conjunction of all and only the likely basic statements $b_1, b_2, \ldots, b_m$ of the language. We have to show that, provided there is one such conjunction b, $Evs_\phi(b)$ is maximal, i.e., that it is greater than $Evs_\phi(h)$ for any h different from b. Suppose first that some of the conjuncts of h are unlikely; it follows from (an iterated application of) Lemma 2 that any basic theory obtained from h by removing an unlikely conjunct has greater expected verisimilitude than h itself. So $Evs_\phi(h)$ cannot be maximal. Suppose now that h contains only likely conjuncts. If these are all the likely basic statements of the language then h is the same as b. Otherwise, if b_i is a likely statement not already in h, then it follows from Lemma 2 that $h \wedge b_i$ has greater expected verisimilitude than h. In sum, $Evs_\phi(h)$ is maximal just in case h is identical with b.

A technical comment on the definition of the threshold σ above is in order, to explain why σ is not simply defined as $\frac{\phi}{\phi+1}$, but it is required to be the greater of 0.5 and $\frac{\phi}{\phi+1}$. As pointed out by an anonymous reviewer, if ϕ (the "weight of falsehood" in Eq. 2) is chosen as smaller than 1, then $\frac{\phi}{\phi+1}$ becomes smaller than 0.5. As a consequence, for some basic proposition x, both $p(x)$ and $p(\neg x)$ may be above the threshold $\frac{\phi}{\phi+1}$. In such case, Theorem 1 implies that the conjunction b maximizing expected verisimilitude is inconsistent, since it contains both x and $\neg x$. Assuming that a rational agent should not believe logically false propositions, one needs to exclude the case above; this can be done in at least two ways. The first is to require that $\phi \geq 1$, and hence that $\frac{\phi}{\phi+1} > 0.5$; this solves the problem by restricting the application of measure vs_ϕ to those contexts (arguably the most common ones) in which the "loss" due to accepting a basic falsehood is greater (in absolute value) than the "gain" obtained from accepting a basic truth. The second solution—suggested by the reviewer and adopted here—is to require that $p(x)$ is greater than both $\frac{\phi}{\phi+1}$ and 0.5 or, which is the same, that x is more probable than its negation and moreover it passes the relevant threshold. This guarantees that there is at most one conjunction of all and only the basic propositions which are likely in the defined sense, and that such conjunction, if it exists, maximizes expected verisimilitude.

References

Cevolani, G. (2016). Another way out of the preface paradox? In L. Felline, F. Paoli, & E. Rossanese (Eds.), *New developments in logic and philosophy of science*. College Publications (forthcoming).

Cevolani, G., Crupi, V., & Festa, R. (2011). Verisimilitude and belief change for conjunctive theories. *Erkenntnis*, *75*(2), 183–202.

Cevolani, G., & Tambolo, L. (2013). Progress as approximation to the truth: A defence of the verisimilitudinarian approach. *Erkenntnis*, *78*(4), 921–935.

Christensen, D. (2004). *Putting logic in its place: Formal constraints on rational belief*. Oxford: Oxford University Press.

Cohen, L. J. (1992). *An essay on belief and acceptance*. New York: Clarendon Press.

Easwaran, K. (2015). Dr. Truthlove, or how I learned to stop worrying and love Bayesian probabilities. In *Noûs*. doi:10.1111/nous.12099.

Foley, Richard. (2009). Beliefs, degrees of belief, and the Lockean thesis. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of belief* (pp. 37–47). Dordrecht: Springer.

Hansson, S. O. (2013). Bootstrap contraction. *Studia Logica*, *101*(5), 1013–1029.

Hartshorne, C., Paul, W., & Burks, A. W. (eds.). (1931–1958). The collected papers of Charles Sanders Peirce. Vols. I–VI edited by Hartshorne, C., & Weiss, P. (1931–1935), vols. VII–VIII edited by Burks, A. W. (1958). Cambridge, MA: Harvard University Press.

Hetherington, S. C. (2005). "Fallibilism". In *The internet encyclopaedia of philosophy*. University of Tennessee. http://www.iep.utm.edu/fallibil/.

Kaplan, M. (2013). Coming to terms with our human fallibility: Christensen on the preface. *Philosophy and Phenomenological Research*, *87*(1), 1–35.

Kitcher, P. (2001). Real realism: The Galilean strategy. *Philosophical Review*, *110*(2), 151–197.

Kuipers, T. A. F. (2000). *From instrumentalism to constructive realism*. Dordrecht: Kluwer Academic Publishers.

Kuipers, T. A. F. (2015) Models, postulates, and generalized nomic truth approximation. In *Synthese*. doi:10.1007/s11229-015-0916-9

Leitgeb, H. (2014a). A way out of the preface paradox? *Analysis*, *74*(1), 11–15.

Leitgeb, H. (2014b). The stability theory of belief. *Philosophical Review*, *123*(2), 131–171.

Levi, I. (1967). *Gambling with truth*. Cambridge, MA: MIT Press.

Maher, P. (1993). *Betting on theories*. Cambridge: Cambridge University Press.

Makinson, D. C. (1965). The paradox of the preface. *Analysis*, *25*(6), 205–207.

Niiniluoto, I (2015) Optimistic realism about scientific progress. In *Synthese*. Forthcoming. doi:10.1007/s11229-015-0974-z.

Niiniluoto, I. (1984). *Is science progressive?*. Dordrecht: Reidel.

Niiniluoto, I. (1987). *Truthlikeness*. Dordrecht: Reidel.

Niiniluoto, I. (1999). *Critical scientific realism*. Oxford: Oxford University Press.

Oddie, G. (2014). Truthlikeness. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Summer 2014.

Oddie, G. (2015) What accuracy could not be, Manuscript.

Pettigrew, R. (2016). *Accuracy and the laws of credence*. Oxford: Oxford University Press.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge* (3rd ed.). London: Routledge and Kegan Paul.

Rescher, N. (1998). Fallibilism. In E. Craig (Ed.), *Routledge enciclopaedia of philosophy*. London and New York: Routledge.

Schurz, G., & Weingartner, P. (2010). Zwart and Franssen's impossibility theorem holds for possible-world-accounts but not for consequence-accounts to verisimilitude. *Synthese*, *172*, 415–436.
Tuomela, R. (2000). Belief versus acceptance. *Philosophical Explorations*, *3*(2), 122–137.