

See discussions, stats, and author profiles for this publication at: <http://www.researchgate.net/publication/284398205>

The Problem of Logical Omniscience, the Preface Paradox, and Doxastic Commitments

ARTICLE *in* SYNTHESIS · DECEMBER 2015

Impact Factor: 0.74 · DOI: 10.1007/s11229-015-0979-7

READS

13

1 AUTHOR:



Niels Skovgaard Olsen

Universität Konstanz

9 PUBLICATIONS 7 CITATIONS

SEE PROFILE

The Problem of Logical Omniscience, the Preface Paradox, and Doxastic Commitmentsⁱ

Niels Skovgaard-Olsen

Abstract: The main goal of this paper is to investigate what explanatory resources Robert Brandom's distinction between acknowledged and consequential commitments affords in relation to the problem of logical omniscience. With this distinction the importance of the doxastic perspective under consideration for the relationship between logic and norms of reasoning is emphasized, and it becomes possible to handle a number of problematic cases discussed in the literature without thereby incurring a commitment to revisionism about logic. One such case in particular is the preface paradox, which will receive an extensive treatment. As we shall see, the problem of logical omniscience not only arises within theories based on deductive logic; but also within the recent paradigm shift in psychology of reasoning. So dealing with this problem is important not only for philosophical purposes but also from a psychological perspective.

1. Introduction

Recent years have witnessed a most fruitful exchange between formal epistemology and psychology of reasoning (Pfeifer & Douven, 2014). This development is in large part due to a shift of paradigms in psychology of reasoning. Whereas earlier work tended to be based on classical logic, more recent developments have begun importing Bayesian models from formal epistemology (Evans, 2002, 2012; Evans & Over, 2004; Oaksford & Chater, 2007, 2010). However, one problem with this interaction, which has not received the attention it deserves, is that it is common for models in formal epistemology to be based on the following norms (Spohn, 2012: ch. 4-5; Huber, 2013), which we shall call the *minimal requirements of rational beliefs*:

- (I) Rational beliefs are deductively closed
- (II) Rational beliefs are completely consistent
- (III) Every logically equivalent sentence is always believed to the same degree by the rational agent

If such models are to have any applications to psychology of reasoning, it is useful to step back to take a synoptic view and consider whether this normative foundation is too idealized to be

applicable to real agents. The way the present paper deals with this issue is by presenting one strategy for making this normative foundation less idealized. It does this by considering the explanatory resources that Brandom's (1994) distinction between acknowledged and consequential commitments affords in relation to the problem of logical omniscience.ⁱⁱ To accomplish this, existing literature is used to identify a number of problems that any adequate account of the relation between norms of reasoning and logic should be capable of meeting (section 3). In a second step, it will then be shown how a particular approach based on the abovementioned conceptual distinction is capable of delivering (what appears to be) satisfactory answers to all of them (section 4).

Briefly stated, *the problem of logical omniscience* is the problem that (I)–(III) appear to impose too demanding constraints on real agents by in effect presupposing that they are logically omniscient (cf. Stalnaker, 1999: ch. 13-14; Levi 1991: ch. 2, 1997: ch. 1). Indeed, this impression is strongly supported by considering the poor logical performance documented in the psychological literature ([Evans 2002, 2012](#), [Oaksford & Chater, 2007, 2010](#)). While psychology of reasoning has gradually moved away from theories based on deductive logic, such as mental logic and mental models theory, formal epistemology has continued to treat deductive closure and consistency as minimal conditions for rational belief sets. Moreover, through the common choice of propositions as the objects of beliefs, formal epistemology has automatically treated logically equivalent sentences as being believed to the same degree—irrespective of well-known, psychological findings such as the framing effect (Kahneman, 2012).ⁱⁱⁱ

In addition to discrepancies such as these with well-established empirical findings, the minimal requirements of rational beliefs have also come under considerable pressure from a range of problematic cases cited in the philosophical literature, which are introduced in section 3. So both the psychological and philosophical literatures suggest that the status of these minimal constraints on rational belief sets needs to be carefully scrutinized. However, it should be noted that the normative principles in question are as much a part of logic-based approaches, like belief revision theory, as they are of the probabilistic models that psychology of reasoning has begun to import from Bayesian epistemology. Hence, a shift from the former to the latter will not alleviate the present concerns.

[Christensen \(2007: 15ff.\)](#) thus argues that the probability calculus should not be viewed as a new logic for graded belief, but rather as “a way of applying standard logic to beliefs, when beliefs are seen as graded”. He makes his case by showing on the basis of the axioms of the probability calculus how the logical properties of propositions impose restrictions on probabilistic coherence.^{iv} An example is that probabilistic coherence requires of the agent that he believes $p \vee q$ at least as strongly as p , which follows directly from the fact that $p \vee q$ is entailed by p . Hence, just as logical

closure for binary beliefs would prohibit that the ideally rational agent believes p while not believing $p \vee q$, so probabilistic coherence for graded beliefs prohibits that he believes p to degree x while believing $p \vee q$ to a degree less than x . Moreover, just as logical consistency of binary beliefs would require that this agent does not believe both p and $\neg (p \vee q)$, probabilistic coherence of graded beliefs requires that his degree of belief in p and $\neg (p \vee q)$ does not sum up to more than one (Ibid.: 15-16).

So no matter whether binary, formal representations of beliefs are preferred (as in the old paradigm in psychology of reasoning), or probabilistic representations of degrees of beliefs are preferred (as in the new paradigm in psychology of reasoning), it holds that: “the prominent proposals for imposing formal constraints on ideal rationality are rooted in logic” (Ibid.: 18). It is only recently that there has been an awareness of this fact in the psychological literature.^v Evans (2012: 6) has aptly put his finger on the implication that this has for the celebrated paradigm shift in psychology of reasoning, when he says:

By around 2000 many researchers using the paradigm were questioning the idea that logic could provide a description of human reasoning, and many were also casting doubt on logic as an appropriate normative system (Evans, 2002; Oaksford & Chater, 1998). While these authors complained about “logicism” in the psychology of reasoning, it is again standard bivalent logic that they had in mind. Any well-formed mathematical system is a closed deductive system that can be regarded as a logic in which theorems (proven conclusions) are deduced from axioms (assumptions). Probability theory, which is much used in the new paradigm, actually reduces to binary logic when probabilities are set to 1 or 0. For example, if we set $P(A \text{ and } B) = 1$, we can conclude that $P(A) = 1$, thus preserving certainty (truth). So it is more accurate to say that authors were objecting to binary logic, which does not allow beliefs represented as subjective probabilities that range freely from 0 to 1, rather than logic per se.

Accordingly, the shift in psychology of reasoning is to be viewed as one concerning the need for representing degrees of beliefs that are concerned with our confidence in propositions rather than with necessary truth preservation of full beliefs. However, since the minimal constraints on rational beliefs have not been abandoned, we are still confronted with the problem of logical omniscience.

In this context, Brandom (1994) has made an interesting, conceptual distinction between acknowledged and consequential commitments, which can potentially throw new light on the normative issues at stake. Section 2 therefore introduces the pertinent features of his account.

2. Acknowledged and Consequential Commitments

2.1 Introducing the Brandomian Framework

Instead of theorizing about belief, Brandom (1994) chooses to theorize about public, doxastic *commitments*, which conversation partners attribute to one another on the basis of the assertions and retractions they make. In this type of interaction, the interlocutors alternate between taking up the role of *the speaker*, who makes the assertions, and *the scorekeeper*, who keeps track on the speaker's assertions by keeping score on the speaker's commitments and entitlements.

A doxastic commitment to p can be thought of as an obligation to defend p when appropriately challenged. For some of an agent's doxastic commitments it holds that the agent already counts as having redeemed his obligation to defend the corresponding claims (either because there are no standing challenges to his warrant that cannot be met, or because the claims *per default* have a defeasible status of not being in need of justification). For the commitments for which this holds, the agent is said to be (defeasibly) *entitled* to his assertions. Moreover, when an agent is attributed entitlement to a claim, it becomes possible for others to adopt a commitment to the claim in question while deferring back to the original speaker for the burden of justification.

To introduce the distinction between acknowledged and consequential commitments, Brandom says:

The commitments one is disposed to avow are *acknowledged* commitments. But in virtue of their inferentially articulated conceptual contents, assertional commitments have consequences. Undertaking a commitment to a claim with one content involves undertaking commitments to claims whose contents are (in the context of one's other commitments) its committive-inferential consequences. Undertaking a commitment to the claim that Pittsburgh is to the West of Philadelphia is one way of undertaking commitment to the claim that Philadelphia is to the East of Pittsburgh. These *consequential* commitments may not be acknowledged; we do not always acknowledge commitment to all the consequences of the commitments we do acknowledge. They are commitments nevertheless. (1994: 194)

For some of the doxastic commitments undertaken by the speaker, the scorekeeper will in other words note that they are acknowledged by the speaker. For others the scorekeeper can note that they are consequences of the *acknowledged commitments*, which the speaker might not acknowledge. One way of thinking about the underlying issue is this: by making an assertion one adopts a conditional task responsibility to defend the claim in light of appropriate challenges. If a doxastic commitment has other doxastic commitments as its consequences, then their falsity can be made part of the challenge posed to attempts of justifying the original claim; even if the speaker

is ignorant of the consequences of what he is saying. To take an example, suppose a speaker asserts both that ‘Berlin is to the North of Behrendorf’ and ‘Copenhagen is to the South of Behrendorf’, then the scorekeeper may challenge these claims by pointing out both that they introduce a consequential commitment to the claim that ‘Berlin is to the North of Copenhagen’ (due to transitivity), which we know to be false.

But to connect the present considerations to the issue of deductive closure above, it must be observed that Brandom talks about consequential commitments in relation to *material* (committive) *inferences* like the inference from one location being west of a second location to the second being east of the first.^{vi} Nowhere does he raise the issue in relation to the logical consequences of one’s beliefs that I am aware of. However, this shortcoming can easily be remedied, because Brandom analyzes the inferential articulation of conceptual content as consisting in the following relations (Brandom, 1994, MacFarlane, 2010):

Commitment preservation: The inference from premises Γ to q is *commitment-preserving* if a commitment to Γ counts as a commitment to q .

Entitlement preservation: The inference from premises Γ to q is *entitlement-preserving* if an entitlement to Γ counts (defeasibly) as an entitlement to q .

Incompatibility: p is incompatible with q if a commitment to p precludes an entitlement to q .

Since Brandom says that commitment-preserving inferences generalize the category of *deductive inferences*, and entitlement-preserving inferences generalize the category of *inductive inferences*, it seems reasonable, as a first approximation, to explicate the underlying reason relations in terms of Spohn’s (2012: ch. 6) account of reasons as follows:

Commitment preservation:

$$P(q|\Gamma) > P(q|\Gamma^c), P(q|\Gamma) = 1 \quad \text{[a probabilistic version]}^{\text{vii}}$$

Entitlement preservation:

$$P(q|\Gamma) > P(q|\Gamma^c), P(q|\Gamma) > b, \text{ for } b \geq 0.5^{\text{viii}}$$

where b denotes a contextually set threshold of when the speaker counts as having fulfilled his obligation to defend his assertions.

Moreover, it is possible to formulate both a weak and a strong notion of incompatibility, where the latter is the limiting case of the former and the case of logical inconsistency is an instance of strong incompatibility:

Weak Incompatibility:

$$P(q|p) < P(q|\neg p), P(q|p) < b, \text{ for } b \geq 0.5$$

Strong Incompatibility:

$$P(q|p) < P(q|\neg p), P(q|p) = 0$$

Hence, what was said above about consequential commitments should, *ipso facto*, apply to the logical consequences of the speaker's doxastic commitments, and what Brandom says about incompatibility should, *ipso facto*, apply to the case of logical inconsistency. We can thus begin to apply our conceptual distinctions to the problem of logical omniscience below in sections 3 and 4. (However, beyond this observation, the explications given above, which depict Brandom's inferential semantics as a probabilistic (or rank theoretic) reason relation semantics,^{ix} will play no further role in the course of the present argument.)

The point of introducing the distinction between acknowledged and consequential commitments is to avoid an ambiguity in belief talk:

In one sense, one believes just what one takes oneself to believe, what one is prepared to avow or assert. In another sense, one believes, willy-nilly, the consequences of one's beliefs (...). The sense of belief in which one is taken actually to believe what one ideally *ought* to believe (at least given what else one believes), call it *ideal* or *rational* belief, can conflict with the sense of belief for which avowal is authoritative. (...) The conflict arises precisely because one can avow incompatible beliefs, and fail to avow even obvious consequences of one's avowals. (Brandom, 1994: 195)

When we leave beliefs behind and focus on public, doxastic commitments, the analogue to cases of incompatible beliefs gets analyzed as cases, where *incompatible obligations* to defend claims have been undertaken. That is, such cases are viewed as the doxastic counterpart to cases, where agents have undertaken incompatible practical commitments by, for example, promising to be in two different places at once (Brandom, 1994: 196). In both cases we are dealing with instances of our general shortcoming as agents that we sometimes undertake multiple obligations that cannot all be redeemed at the same time.

Where things begin to get interesting is in relation to consequential commitments. As Kibble (2006b: 37) points out, just as it would be an inappropriate response to an agent, who has undertaken incompatible practical commitments, to attribute any arbitrary intention, it is a central feature of Brandom's pragmatic model of giving and asking for reasons that it would be inappropriate to follow the principle of *ex falso quodlibet* and attribute any arbitrary doxastic commitment to an agent, who has undertaken incompatible doxastic commitments. Instead the

appropriate response is to withhold attributions of entitlement to the particular claims that are incompatible (Brandom, 1994: ch. 3). Through this act, any further inheritance is blocked to these claims through testimony that would otherwise have allowed other agents to adopt a commitment to the claims in question while deferring back to the speaker for the burden of justification. However, this restriction to the applicability of *ex falso quodlibet* need not commit us to revisionism about logic, as we shall see in section 4.

2.2 Reinterpreting the Norms of Rational Belief

It is worth noticing that—as Milne (2009: 276) points out—the minimal principles of rationality have a natural justification on the basis of the norms of assertion. Extending a bit, the argument would go roughly as follows:

- (P₁) Making an assertion is to be understood as licensing others to use it as an uncontroversial starting point for further inquiry^x while deferring back to the speaker for the burden of justification (cf. Brandom 1994: 174, 2001:165).
- (P₂) The interlocutors would not be able to use an inconsistent set of propositions as an uncontroversial starting point for further inquiry.
- (P₃) The interlocutors would not be able to use a set of propositions that have unacceptable logical consequences as an uncontroversial starting point for further inquiry.
- (P₄) The interlocutors would not be able to use the speaker's assertions as an uncontroversial starting point for further inquiry, if they have logically equivalent formulations that are themselves unacceptable.
- (C) Hence, the speaker's obligation to defend the assertions he makes, when appropriately challenged, extends to the avoidance of their inconsistency, defending their logical consequences, and to defending their logically equivalent formulations.

Essentially the idea is that it is part of the epistemic use to which the speaker's interlocutors can reasonably put his assertions to exploit their logical properties for further computation. Consequently, it would constitute a failure if the speaker feeds them assertions that fail to meet its minimum requirements. As a result, the speaker's interlocutors are entitled to enlist the logical consequences of the speaker's acknowledged commitments as consequential commitments with an equal claim to form the basis of challenges as his acknowledged commitments.

Following this line, we can begin to view the minimal rationality constraints on beliefs introduced in section 1 as constraints governing the score of commitments and entitlements that the scorekeeper keeps on the speaker in the course of an argumentative dialogue. That is to say, in deciding whether the speaker has a constellation of commitments for which it both holds that there are no serious, unmet justificatory challenges, and that others would be permitted to inherit claims while deferring back to the speaker for the burden of justification, the scorekeeper can be viewed as engaged in the task of constructing a belief set, based on the speaker's public utterances, that is to be consistent and closed under logical consequence.

Viewing matters from this perspective allows us to regard the importance of these rationality principles as not consisting in whether speakers actually succeed in only avowing to consistent beliefs and all their logical consequences (which would be a claim of which the psychological literature suggests that we should remain highly skeptical). But rather as consisting in there being norms that we impose on others, when deciding whether it is safe to accept what they say, which we hold them accountable to in justificatory challenges. That is to say, what matters in this context is not so much the speakers' actual performance in their own individual reasoning, but whether they would accept challenges of their claims based on logical consequences of their claims that are themselves unacceptable. If the speakers accept such challenges, they can be taken to display the recognition of being bound by these norms; even if they are unable to comply with them by their own efforts.

A particularly nice example is Russell's famous letter to Frege, where Russell directs Frege's attention to the inconsistency generated by set theory known as Russell's paradox, of which Spohn (2012: 48) writes:

Was Frege irrational because he firmly believed in an inconsistent axiomatization of set theory? Clearly not. He would only have been irrational if he had stuck to the axiomatization after Russell's discovery of the inconsistency. But of course he immediately saw the impact of Russell's antinomy.

This example clearly illustrates the following points that we will be emphasizing throughout this paper: (i) we cannot use Frege's formal system as an uncontroversial starting point for further inquiry as it stands due to Russell's paradox, (2) it is appropriate to pose challenges to Frege based on this inconsistency, (3) Frege would not be permitted to ignore Russell's challenge, yet (4) the fact that there was an inconsistency generated by Frege's belief set does not show that he was being irrational, as suggested by the quote above.

Notice that the point of this reorientation in the interpretation of (I)–(III) is not to shift the burden of conformity to (I)–(III) from the speaker to the scorekeeper. To be sure, such a shift would constitute a substantial improvement. The reasons are that: (a) we are probably better at

recognizing errors in the reasoning of others than in our own individual reasoning (cf. Mercier & Sperber, 2011), and (b) the general logical powers of the scorekeeper are boosted, because talk about ‘the scorekeeper’ is really a placeholder for talk about an open-ended community of agents, who engage in a critical discussion about how to assess the speaker’s claims. The reason why this position is nevertheless not advanced in this paper is that even an open-ended community of scorekeepers is bound to violate (I)–(III) some of the time. Instead the position endorsed below is that we should shift the focus from the actual conformity to (I)–(III) to investigating the recognition of being bound by (I)–(III) displayed by the justificatory challenges posed by scorekeepers and the challenges accepted by speakers. The Brandomian lesson to be learned here is that it is not through our conformity that we are bound by (I)–(III) in our reasoning, but rather through the normative attitudes we express in our sanctioning practices (cf. Brandom, 1994: ch. 1). Taking this line moreover allows us to follow Levi (1991: ch. 2, 1997: ch. 1) in viewing (I)–(III) not as principles that we succeed in implementing in our actual behavior, but rather as regulatory ideals that we recognize our obligation to approximate through a continuous effort to refine our abilities by technological and educational means.

3. Four Possible Gaps between Logic and Norms of Reasoning

In an unpublished manuscript that is too good not to be cited, MacFarlane (manuscript) considers 36 possible bridge principles between norms of reasoning and logical consequence that take the following form:

If A, B, \models C, then (normative claim about believing A, B, and C)

The different versions are produced by varying the following four parameters: (1) the type of deontic operator (i.e. whether facts of logical validity give rise to *obligations*, *permissions*, or *defeasible reasons* for beliefs), (2) the polarity (i.e. whether the obligations, permissions, or defeasible reasons concern *believing* or *not disbelieving*), (3) the scope of the deontic operator, and (4) whether the facts about logical validity have to be known by the agent. But the preceding discussion has already brought out further parameters that MacFarlane’s otherwise comprehensive discussion fails to consider: (5) beliefs vs. public commitments, (6) acknowledged commitments vs. consequential commitments, and (7) the doxastic perspective of the speaker vs. that of the scorekeeper.

So to illustrate the attractiveness of transposing the normative issues in the way outlined above by thinking of the rationality principles not as principles of private beliefs, but as principles of public commitments, which are imposed from a scorekeeping perspective, it is instructive to review some of the puzzle cases that MacFarlane discusses. More specifically, we are going to look at the arguments posed by Harman (1986) to show the lack of a connection between logical consequence and norms of reasoning, which have been succinctly summarized by Hartry Field (2009: 252-3) as follows:

Problem 1:

Reasoning (change of view) doesn't follow the pattern of logical consequence. When one has beliefs A_1, \dots, A_n , and realizes that they together entail B , sometimes the best thing to do isn't to believe B but to drop one of the beliefs A_1, \dots, A_n .

Problem 2:

We shouldn't clutter up our minds with irrelevancies, but we'd have to if whenever we believed A and recognized that B was a consequence of it we believed B .^{xi}

Problem 3:

It is sometimes rational to have beliefs even while knowing they are jointly inconsistent, if one doesn't know how the inconsistency should be avoided.

Problem 4:

No one can recognize all the consequences of his or her beliefs. Because of this, it is absurd to demand that one's beliefs be closed under consequence. For similar reasons, one can't always recognize inconsistencies in one's beliefs, so even putting aside point 3 it is absurd to demand that one's beliefs be consistent.

An example of problem 3 is the preface paradox, where the author of a book finds that he has supporting evidence for every single claim made in his book, yet knowledge of his own general fallibility cautions him to disbelieve the conjunction of all his claims. If beliefs are closed under conjunction, he thereby finds himself with an inconsistent belief set. Yet it is not clear what he should do about it as all of his beliefs seem quite reasonable.

A further example is given by Hartry Field in his second John Locke lecture:

any rational person would have believed it impossible to construct a continuous function mapping the unit interval onto the unit square until Peano came up with a famous proof about how to do it, so the belief that no such function could exist was eminently rational but inconsistent, and there are many more examples of a similar nature.^{xiii}

Below in section 4 bridge principles will be formulated that differ from those MacFarlane considers by introducing parameters (5)-(7), which are capable of handling problems 1-4, as well as three further constraints that MacFarlane (manuscript) considers. This is a significant contribution insofar as MacFarlane presents these desiderata as standing in a tension and thus requiring some sort of trade-off. But first we start out with some initial observations and a treatment of the preface paradox.

3.1 Preliminary Observations

The first thing to notice is that we can simply grant Harman (1986), Foley (1993), and others that there are cases like the preface paradox, where it, from the speaker's point of view, may make sense to give in and learn to live with an inconsistency, if it is either too hard or costly to deal with the problem. Moreover, logic does not provide a guide for the speaker for how to manage his acknowledged commitments, if it comes to his attention that they have a logical consequence that is better avoided, because there are always more ways of resolving the issue, as problem 1 indicates.

However, this does not mean that the principles of rationality cease to impose norms of reasoning, and that the scorekeeper should cease to treat the speaker as *obligated* to avoid inconsistencies and accept the logical consequences of his acknowledged commitments (as long as they have not been withdrawn), as we shall see in detail below. Furthermore, the speaker can be seen as recognizing that these norms are still in force, if he accepts the appropriateness of challenges based on his failure to repair his "public belief set".

As we have seen, the outcome of the scorekeeper's failure to construct a deontic score for the speaker that meets the minimal constraints on belief sets is not that the speaker fails to have any rational beliefs. For first of all, we are treating these principles as requirements of public commitments and not as requirements of (rational) beliefs. Secondly, the speaker's failure to comply with them does not even mean that he does not have any public, doxastic commitments. It just means that he has undertaken an obligation to defend a constellation of claims that he cannot redeem (because they would require him to accept as consequential commitments logical consequences, which in turn cannot be defended). Thirdly, the consequence of this failure is that the speaker for the moment cannot be attributed entitlement (and be treated as a source of entitlement for others) *with respect to the afflicted assertions*.^{xiii} But this may be a consequence that the speaker may have to live with at times, where there is no obvious repair to the constellation of obligations that he has undertaken. The rationale for this penalty is to avoid the propagation of error, and indeed both Foley (1993: 119) and Harman (1986: 15-7) agree that it would be a mistake to base further inquiry on inconsistent propositions even if they are sometimes unavoidable.

Because the consequential commitments are only used as an aid in deciding whether entitlement can be attributed, the possibility is not precluded that the speaker may sometimes be rationally permitted to manage his acknowledged commitments in ways that temporarily exclude him from attributions of entitlements. In such cases, the agent's assertions can be treated temporarily as not being a source of information that can be used unproblematically as a base for further inquiry. If it happens regularly, the agent can be blacklisted (see also Kibble (2006b)). In this way it is possible to drive a wedge between our assessments of the agent's rationality and of the information that we want to use for further inquiry. For rational agents it need not be possible to be a source of valuable information under all circumstances—no matter how paradoxical the requirements they are confronted with.^{xiv}

A case in point may be the preface paradox, which we will return to shortly. In this context, it is also worth noting the situation that Harman (1986: 16) argues that most of us are in when it comes to the liar paradox:^{xv}

the rational response for most of us may simply be to recognize our beliefs about truth are logically inconsistent, agree this is undesirable, and try not to exploit this inconsistency in our inferences.

Furthermore, Foley (1993: 115-7) discusses a number of interesting cases, where he, *inter alia*, makes the point that sometimes the optimal strategy is not the one that has a small chance of arriving at an ideal outcome, where no mistakes are made, but rather one that minimizes the expected number of mistakes (even if one can thereby be certain that mistakes are made some of the time). Indeed, a case could be made that this is exactly the type of situation we find ourselves in when we have to rely on what is known to be fallible sources of information, which is surely the normal course of events. Of course, this leads us directly to the preface paradox.

3.2 Dealing with the Preface Paradox

There are various desiderata that an adequate solution to the preface paradox should be capable of meeting. On the one hand, we want to continue to take measures to avoid errors from propagating by treating inconsistency as a defect for a set of commitments, which makes the afflicted assertions incapable of functioning as an uncontroversial starting point for further inquiry. On the other, Foley (1993: 117) seems right to insist that it is a desideratum for any decent theory that agents should not be deemed irrational for recognizing their own fallibility. Indeed, it seems that, if anything, it is part of being an epistemically responsible agent to do just that. Furthermore, we want to avoid the absurd result that the set of commitments undertaken in a book by epistemically

responsible agents ends up not being a suitable starting point for further inquiry by our own standards due to its inconsistency.

In meeting these constraints, we will use reflections about what the function is of the various parts of a book as our clue. As it turns out, the resulting approach ends up fitting nicely with Spohn's observation that the problem generated by the preface paradox arises due to a mixture of epistemic perspectives.^{xvi}

To make things a bit more concrete, I will use Olsen (2014) as an example. In it chapters I–V serve the function of advancing substantial claims about a number of subject matters ranging from methodological issues, the semantics of conditionals, performance on psychological experiments, and the nature of rationality. In contrast, the preface served the opportunity to make a statement about *the epistemological status* of the claims advanced in chapters I–V (in addition to its more rudimentary functions of advertising what is to come and acknowledging the influence of others). There is thus a sense in which all the substantial claims made in that book are contained within chapters I–V and that nothing of consequence about its subject matter is stated in the preface. Accordingly, if the reader wants to look up what its author thinks about some topic to challenge it, then he or she should turn to chapters I–V and can safely ignore the preface. Hence, chapters I–V contain all the claims that I undertake an obligation to defend in writing that book *qua* author.

In contrast, in commenting in the preface on the epistemological status of the claims advanced in chapters I–V, I am already beginning to comment on what in the book can be used as a starting point for further inquiry. However, that is the task of the scorekeeper. So in a preface of this type, the author is already beginning to act as his own scorekeeper, as it were, and it is here the source of the problems is to be located.

To disentangle the roles of these different epistemic perspectives, it is useful to take a look at what Brandom (1994: ch. 8) has to say in general about the interaction between the doxastic perspectives of the speaker and the scorekeeper. Elsewhere I have laid out these matters more carefully. But for present purposes let the following brief sketch suffice. According to Brandom, it is a structural feature of the scorekeeping perspective that a principled distinction is drawn between *what is actually correct* and *what is merely taken to be correct*. He holds that this normative distinction is expressed through the use of *de dicto* and *de re* ascriptions, when attributing doxastic commitments to the speaker. That is, in describing the claims that the speaker has undertaken an obligation to defend on the basis of his assertions, the scorekeeper can either express the assertions in the speaker's own vocabulary, in a form that he would acknowledge having undertaken, or he can specify which entities the speaker is *talking about*, and what claims he is making *of* these entities,

using his own vocabulary. Of the two, the latter is the form used for making truth assessments, as the following example illustrates:

Bruja: “Pachamama will yield a poor harvest unless she is treated properly”.

Scorekeeper: “The Bruja is claiming *of* the earth *that* it will yield a poor harvest unless it is treated properly”.

Once stated in its *de re* form, the Bruja can be treated as having made an acceptable assertion that any farmer will give his assent to (although they may have their disagreements about exactly what counts as a proper way of treating the earth once we go beyond the most basic maxims).^{xvii} In its former *de dicto* version the scorekeeper might have had some reservations.

In making the distinction between *what appears to be correct*, according to the doxastic perspective under assessment (i.e. the claim about Pachamama), and *what is correct*, once this claim has received a *de re* specification, the scorekeeper needs a supply of propositions stating how the world actually is. To him, it will appear that his own collateral commitments make up this set (because why else accept these propositions unless they appeared to express how the world actually is to him). So in effect he is comparing the commitments of the doxastic perspective under assessment with his own doxastic commitments—in spite of the fact that it appears to him *as if* he is comparing what the Bruja takes to be correct to how things actually are.

Now the point of introducing this bit of Brandom’s account is that it puts into a new light what the author is doing in the preface when starting to act as his own scorekeeper. When acting as a scorekeeper in relation to foreign doxastic perspectives, the scorekeeper is bound to make some attributions of error simply due to the differences in their collateral commitments. So here the scorekeeper has no problem with complying with the maxim that no agent is to be treated as infallible as he will attribute mistakes to the commitments under assessment part of the time. However, when he is acting as a scorekeeper on a book written by himself, his comparisons of what the author takes to be correct with what is actually correct all end up falling out favorably as he is in effect comparing a set of propositions with itself.^{xviii} So in this case the maxim that no agent is to be treated as infallible is violated and he cannot express a recognition of the fallibility of the author without producing the inconsistency expressed by the preface paradox. Actually, the problem is twofold. On the one hand, there is the problem of denying the proposition that every claim in the book is correct *qua* scorekeeper while simultaneously being committed to defending that very claim *qua* author. On the other, there is the problem that supposing that there is an error in the book—in spite of the fact that each claim is assessed as correct—ends up indicating that the

set of propositions is error-prone that is presupposed to express how the world actually is in his own truth assessments.

This is how things look from the author's side of the story. When we turn to his readers, the present suggestion is that they should construct two scores of commitments that they attribute to the author. The first is the author's deontic score *qua* author and it contains the propositions that the author has undertaken an obligation to defend during his treatment of the subject matter under investigation. In our example, this would be the propositions expressed in chapters I–V. The second is the author's deontic score *qua* acting as a scorekeeper on his own work and it contains the propositions that the author has undertaken an obligation to defend through his remarks in the preface. Of the two, the latter is guaranteed to be an inconsistent set, so the afflicted propositions cannot be attributed entitlement, and the former is only inconsistent, if the author happens to have produced an inconsistency in his treatment of the issues dealt with in chapters I–V.

For the author each claim in chapters I–V appears to be justified and correct and he states so in the preface. But the reader is well-advised not to be predisposed to accept all of the author's claims about the epistemological status of the claims made in the book due to the inconsistency. Rather, the reader should weigh the author's fallibility higher than the fact that each claim in chapters I–V appears to be justified to the author. For what the author's fallibility amounts to is exactly this: part of the time he makes claims that appear to be correct to him despite the fact that they are actually mistaken. In contrast, the author is unable to weigh the information about the epistemological status of his assertions in the same manner, if it would mean that he should stop acting on what he perceives to be a good justification for making a particular claim. What he can do is to improve his skills at evaluating and obtaining evidence. But no matter how good he gets, there will always be a point, where he just has to rely on what he perceives to be a good justification in spite of his continued fallibility.

So the way the present account seeks to avoid the absurd consequence that we can no longer use the claims advanced in books as a starting points for further inquiry is by demarcating the inconsistency produced in the preface to the deontic score of the author *qua* acting as a scorekeeper on his own claims. That the epistemically responsible author refuses in the preface to undertake an obligation to defend the claim, that every proposition in the book can be used as an uncontroversial starting point for further inquiry, does not make the negation of the conjunction of all the propositions in chapters I–V part of the actual claims advanced in the book. Surely, the point of writing the book was not just to present the reader with yet another large conjunction of claims that he should not accept.

No, the author's score *qua* author begins and ends with chapters I–V. And in relation to this set of commitments business is as usual. That is to say, if the author is reliable, then the fact that a

claim appears to him to be justified should be allowed to carry some weight. But ultimately the readers ought to make their own assessments of whether entitlement can be attributed to each individual claim and be prepared to make some attributions of mistakes on the grounds of the author's general fallibility.

Since the attribution of inconsistency is only used as a way of stopping error from propagating, the present account moreover allows the scorekeeper to treat the author as continuing to be rational. The inconsistency in the author's score *qua* acting as a scorekeeper on his own work is only produced, because the agent is acting on incompatible obligations, each of which seems eminently rational in its own right. On the one hand, he continues to be the author of the book and is therefore committed to defend the claims advanced in chapters I–V (and will use these in his assessments of how things actually are). On the other, the author is trying to give his readers some instruction in the preface about how to assess his own claims based on how he would have assessed them, if they had been advanced by someone else. In this, the author tries to express a recognition of his own fallibility, which is surely the only responsible thing to do. Unfortunately, in attempting to combine both concerns he ends up producing an inconsistency in the second deontic score. But the fault lies with his incompatible obligations and not in his lack of rationality.^{xix}

As we have seen, this account is thus able to meet all of the desiderata for dealing with the preface paradox identified above.

4. The Bridge Principles and Problems 1-4

To return to MacFarlane's (manuscript) bridge principles, I extend this list by the following candidates, which are loosely inspired by Brandom's account. As said, these bridge principles differ from those MacFarlane considers in dealing with public commitments instead of belief, introducing the focus on acknowledged and consequential commitments, and in emphasizing the doxastic perspective of the speaker and the scorekeeper:

- (I) If $A, B, \models C$, then the speaker ought to see to it that if he/she acknowledges a commitment to A and a commitment to B , and A and B are challenged via C ,^{xx} he/she acknowledges a commitment to C .

Commentary: the speakers' means for acknowledging a commitment to C consists in accepting challenges to A and B based on challenges to C .

- (II) If $A, B, \models C$, then if the speaker acknowledges a commitment to A and B , the scorekeeper is permitted/entitled to attribute a consequential commitment to C .

Moreover, since all relations of commitment preservation are entitlement-preserving, it holds that:

- (III) If $A, B, \models C$, then if the speaker acknowledges a commitment to A and B , and the scorekeeper both attributes an entitlement to A and B and a consequential commitment to C , the scorekeeper ought to attribute an entitlement to C .
- (IV) If $A, B, \models C$, then if the speaker is entitled to adopt a commitment to A and B , the speaker is entitled to adopt a commitment to C .

It is to be noticed that the deontic operator is given a wide scope over the whole conditional in (I). As a result, (I) describes the conditional task responsibility of the speaker to acknowledge a commitment to C , *if* he/she acknowledges a commitment to A and B . However, this is an obligation that can be fulfilled by either acknowledging a commitment to C or by withdrawing the commitment from A or B . So the first of Harman's problems is avoided (cf. MacFarlane, manuscript). We can also set aside problem 3, as it has already received an extensive treatment above.

(It should, moreover, be noted that principle (III) and (IV) were mainly stated for the purpose of completeness. They will play no further role in our treatment of problems 2 and 4 below.)

4.1 Dealing with Problems 2 and 4

One of the ramifications of making it the task of the scorekeeper to construct a (public) belief set for the speaker, on the basis of his assertions, is that problem 2 and 4 need to be addressed both from the perspective of the speaker and from that of the scorekeeper.

If we start out with the speaker's perspective, the first observation to be made is that the speaker has only adopted the conditional task responsibility to defend his commitments whenever appropriately challenged. Hence, the speaker need not worry about the excessive demand of having to defend all the consequences of his claims in the absence of scorekeepers, who are capable of identifying the corresponding consequential commitments and posing suitable challenges. However, as the knowledge of the implications grows, the speaker continues to run the risk of having to retract his earlier claims, if he cannot provide an adequate response to the novel challenges.

So to see how the speaker can fulfill the requirements of bridge principle (I) in light of problem 4, it suffices to notice that the context in which the speaker would have to acknowledge a commitment to the logical consequence of his acknowledged commitments is when challenges are posed to the consequential commitments as a way of challenging his acknowledged commitments. So what the speaker would need to do to comply with this bridge principle is merely to accept such challenges and be prepared to withdraw his commitment to A or B in the case the challenges to C prove to be too severe.

Moreover, problem 2 is easily avoided. To the extent that challenges are hardly going to be based on trivial (and irrelevant) logical consequences of the speaker's acknowledged commitments, the speaker does not stand in danger of having to devote precious, cognitive resources to dealing with irrelevancies.

When we turn to the scorekeeping perspective, one way of dealing with this same problem of clutter avoidance would be to hold that “the algorithm” for adding logical consequences to the speaker's score as consequential commitments terminates whenever its operation does not immediately contribute to the task of finding out whether entitlement can safely be attributed. That is, there will be no need for the scorekeeper to go through infinite sequences of conjuncts and disjuncts, if it is already clear from the outset that they are irrelevant for determining whether entitlement can be attributed.

This way of addressing problem 2 moreover opens up for a way to avoid being committed to revisionism about logic due to the restriction of *ex falso quodlibet* noted in section 2.1. Accordingly, one way of getting around this problem would be to hold that “the algorithm” for adding logical consequences to the score terminates for a particular set of assertions as soon as an inconsistency has been detected. For then the task of assessing whether entitlement can be attributed has already been solved, and the scorekeeper can proceed to challenge the speaker and criticize others that adopt commitments to the claims in question through deference to the speaker.

If we apply bridge principle (II) to problem 4 for the scorekeeping perspective, we notice that the task of assessing whether entitlement can be attributed does not impose excessive demands on the scorekeeper, because although the scorekeeper is *permitted* to add all the logical consequences as consequential commitments to the speaker's score—and to challenge him on this basis—he is *not required* to do so.^{xxi} Similarly, although the scorekeeper is *permitted* to run complete consistency checks on the speaker's score using all the logical consequences as consequential commitments, he is *not required* to do so. Nor is he required to check every logical equivalent formulations of the speaker's acknowledged commitments.

As we have seen, the scorekeeper is entitled to take these measures to enforce his duty to prevent error from propagating, when the speaker puts forward his assertion as something that

others can use as an uncontroversial starting point for further inquiry. But the scorekeeper can, of course, refrain from fully exercising this right by not investigating *all* the logical consequences of the speaker's assertions, if he is willing to run the risk of letting an error slip in. Indeed, at some point he must terminate prematurely due to the undecidability of logical consequence. But even if consequence were decidable, he would still have to terminate prematurely due to: (1) the complexity involved in discovering that $A, B, \models C$ may exceed what would have been physically possible for him given the best proof systems available, (2) the fact that there are infinitely many consequences of $A \& B$, which cannot be investigated in a finite amount of time, and (3) his limited logical competence.^{xxii}

Potentially the algorithm for executing this task takes the form of a fast and frugal heuristics (cf. Gigerenzer, 2010), which only adds the most salient consequential commitments that would be needed for the context of conversation. For surely there is a trade-off to be made between the cost of continuing to probe the speaker's (public) belief set by adding logical consequences and the potential cost of sometimes adopting error-prone claims through testimony. However, this does not mean that we have to give in to problem 4, because, as Levi (1991: ch. 2, 1997: ch. 1) has emphasized, the important question is not, whether our actual performance succeeds in implementing the requirements of the principles of rationality. But rather whether we continue to recognize that we are in need of improvement whenever they don't. That is to say, to the extent that we continue to refine our abilities to detect consequential commitments through, for instance, education and technological assistance (e.g. use of computers, paper and pencil, and handbooks of tables), we express our recognition that there is a regulatory ideal that we stand under an obligation to approximate.

4.2 Three Further Constraints

In addition to the cases we have already considered, MacFarlane (manuscript: pp. 11-2) uses the following constraints to adjudicate between possible bridge principles. Since his concern is with the relationship between logical consequence and rational beliefs, we will need to consider whether something equivalent holds for the case of public commitments.

The first is *the strictness test*, which holds that for the general case, the agent has not done everything that he ought to, if he only believes p but not its logical consequence, q .

Although our first bridge principle did not capture the exact wording of this constraint, a case could be made that it managed to capture the gist of it by requiring that the speaker accepts challenges based on the logical consequences of his acknowledged commitments. At this point it is unclear whether anything further is needed or whether this conditional task responsibility already succeeds in making the relation between p and its logical consequences sufficiently strict.

The second, perhaps more important one, is whether the proposed bridge principle is capable of getting the priority right, so that we can still say that:

We seek logical knowledge so that we will know how we ought to revise our beliefs: not just how we *will* be obligated to revise them when we acquire this logical knowledge, but how we are obligated to revise them even now, in our state of ignorance. (ibid.)

This concern arises, because if we were only normatively constrained by known logical consequences, it seems that “[t]he more ignorant we are of what follows from what, the freer we are to believe whatever we please” (ibid.), which seems to get things backwards.

More specifically, the concern in our context might be that since the speaker only has to acknowledge the logical consequences of his acknowledged commitments as consequential commitments by accepting suitable challenges, the speaker gets off the hook more easily the more ignorant his scorekeepers are. In response, it can be pointed out that the speaker’s responsibility to accept such challenges does not come with an expiration date.^{xxiii} So he will continue to be liable to criticism, if his assertions are shown to be logically incoherent as our knowledge about the logical consequences grows. (Or rather, the expiration date is the point, where we can no longer consider the agent’s assertions as uncontroversial starting points for further inquiry, because our knowledge has grown too much in the intermediary time. But clearly this does not guard the original agent from revision through ignorance, because what it means is merely that the assertions will lose their epistemic significance once the ignorance is overcome, if there was anything problematic about them in the first place.)

With an open-ended responsibility to answer justificatory challenges based on consequential commitments incurred by the logical implications of his claims, the speaker has a motivation to seek logical knowledge to understand what sort of challenges he makes himself liable to. Similarly, it holds for the scorekeeper that—although he is only permitted and not required to add the logical consequences as consequential commitments to the speaker’s score according to bridge principle (II)—he risks contributing to the propagation of error whenever he refrains from exercising this right. And he was assigned a general duty to prevent error from propagating above. So he too is under pressure to overcome a state of ignorance. Hence, there will be a pressure for both the speaker and the scorekeeper towards overcoming our state of ignorance on the present proposal.

The final constraint consists in being able to maintain that an agent, who refuses to take a stand on a logical consequence (e.g. the conjunction of his beliefs), is acting in a way that he ought not to. As we have seen, bridge principle (I) postpones the need for the speaker to take a stand on the logical consequences of his acknowledged commitments until a suitable challenge emerges, and it is this feature of the present account that ensures that excessive demands are not imposed on the

speaker. But on the other hand, it is not clear why the agent should be forced to take a stance on *all* the logical consequences of his acknowledged commitments in the absence of a well-grounded suspicion about unmet, severe challenges. It might be *prudent* for the speaker to consider a good deal of logical consequences of his assertions before making them to avoid having to withdraw them immediately in the face of embarrassing challenges. But it is not obvious why it would constitute a failure of his epistemic responsibility, as long as he is prepared to withdraw them if severe challenges emerge. And, of course, at that point bridge principle (I) no longer licenses him to refrain from taking a stance on the logical consequences of his acknowledged commitments.

According to bridge principle (II), the scorekeeper is not required to take a stance on all the logical consequences of the speaker's acknowledged commitments. And it is this feature of the present account that ensures that excessive demands are not imposed on the scorekeeper. But here too it is unclear why it should be problematic that the scorekeeper refuses to take a stance on whether a logical consequence could be added to the speaker's score as a consequential commitment, unless there was some well-grounded suspicion that the scorekeeper might thereby contribute to the propagation of error. So here too our bridge principles don't seem to collide with MacFarlane's (manuscript) criteria of adequacy.

5. Conclusions and Future Work

It then appears that the present account is capable of handling the problematic cases that Harman (1986) discusses, as well as the further constraints that MacFarlane (manuscript) considers. The significance of this contribution consists in that MacFarlane presents these various desiderata as standing in a tension and thus requiring some sort of trade-off, which has been avoided on the present account.

By theorizing about public commitments instead of beliefs, we are able to treat cases of inconsistency as harmless cases of incompatible obligations that cannot all be redeemed at once. By invoking the distinction between doxastic perspectives, and making it the task of the scorekeeper to construct a deontic score for the speaker that meets the minimal requirements of rational beliefs to decide whether entitlement can be attributed, we are able to drive a wedge between assessments of the speaker's rationality and assessments of which information we want to use for further inquiry. This move allows the speaker to be rationally permitted to maintain inconsistent doxastic commitments, when confronted with conflicting requirements, while allowing his scorekeepers to take measures to prevent errors from propagating. Moreover, we have

seen that it comes with the further nicety that we can continue to remain uncommitted about revisionism about logic, while avoid letting *ex falso quodlibet* ruin the deontic score of the speaker by adding commitment to random propositions, whenever the speaker finds himself in situations of this kind.

An area for further investigation is a general comparison between the respective advantages and disadvantages of formulating the bridge principles in terms of public commitment or rational beliefs. It is surely of central importance when dealing with this issue that while it is not completely voluntarily what we believe (in the sense that if we really believe something, we cannot just decide to stop believing in it whenever we want to (ibid: 15)), our acknowledged commitments is something that we can exercise full control over. For this reason it might be more natural to think about potentially conflicting obligations in terms of public commitments than in terms of beliefs, which would thereby restrict a central tool for dealing with inconsistencies to bridge principles formulated in terms of public commitments. However, it is at the same time clear that a more complete, comparative discussion would, *inter alia*, have to compare the present bridge principles formulated in terms of public commitments with those formulated in terms of beliefs advanced in MacFarlane (manuscript) and H. Field (2009).

The upshot of this paper has been that one can make the normative foundation of models in formal epistemology more palatable by viewing it as applying to public commitments attributed in argumentative contexts instead of to beliefs in individual reasoning. As such, this approach to the problem of logical omniscience opens up for a new avenue of research in psychology. The take home message has been that if we are interested in the extent to which consistency, deductive closure, and the equivalent treatment of logically equivalent sentences provide a suitable normative foundation, we should not look at whether the participants actually succeed in complying to these norms in their own individual reasoning, but rather at the extent to which they recognize being bound by them in argumentative contexts through the justificatory challenges they pose and accept.

More generally, this reorientation connects with the work of Hahn & Oaksford (2007) and Mercier & Sperber (2011), who have recently made an influential case that the primary function for which reasoning evolved is the production and evaluation of arguments. In support of this claim, Mercier & Sperber (2011) cite a range of circumstantial evidence. Probably the most convincing of which is the finding that once the Wason selection task^{xxiv} was posed in groups, where the participants could deliberate about the solution in an argumentative context, the performance went up from the usual ca. 10%^{xxv} to about 70% (and even to 80% when they had first been presented with the problems on an individual basis). Moreover, this drastic improvement in performance was not merely the result of there being one individual in each group, who had come up with the correct solution and shared it with the others, as the verbal transcripts clearly

show how some groups were able to jointly assemble all the pieces of the puzzle (Moshman & Geil, 1998).

Of course, such findings do not conclusively settle the issue about the evolutionary function of reasoning. But they do make it interesting to follow the approach sketched in this paper to test whether the norms are being recognized in an argumentative setting as opposed to being followed in individual reasoning.

References

- Brandom, R. (1994), *Making it Explicit*. Cambridge, MA.: Harvard University Press.
- (2001), *Articulating Reasons*. Cambridge, MA.: Harvard University Press.
- Christensen, D. (2007), *Putting Logic in Its Place: Formal Constraints on Rational Belief*. Oxford: Oxford University Press.
- Dorn, G. J. W. (2005), 'Eine komparative Theorie der Stärke von Argumenten', in: *Kriterion*, 19: 34-43.
- Dutilh Novaes, C. (forthcoming), 'A dialogical, multi-agent account of the normativity of logic', in: *Dialectica*.
- Evans, J. St. B. T. (2002), 'Logic and human reasoning: An assessment of the deduction paradigm', in: *Psychological Bulletin*, 128: 978-996.
- (2012), 'Questions and challenges for the new psychology of reasoning', in: *Thinking & Reasoning*, 18 (1): 5-31.
- Evans, J. St. B. T. & Over, D. (2004), *If*. Oxford: Oxford University Press.
- Field, H. (2009), 'What is the normative role of logic?', in: *Proceedings of the Aristotelian Society Supplementary Volume*, 83(1): 251-68.
- Foley, R. (1992), 'The Epistemology of Belief and the Epistemology of Degrees of Belief', in: *American Philosophical Quarterly*, 29 (2): 111-24.
- Gigerenzer, G. (2010), *Rationality for Mortals: How People Cope with Uncertainty*. Oxford: Oxford University Press.
- Hahn, U. & Oaksford, M. (2007), 'The rationality of informal argumentation: A Bayesian approach to reasoning fallacies', in: *Psychological Review*, 114: 704-732.
- Harman, G. (1986), *Change in View: Principles of Reasoning*. Cambridge, MA.: The MIT Press.
- Huber, F. (2013), 'Formal Representations of Belief', in: Zalta, E. N., *The Stanford Encyclopedia of Philosophy*. (Summer 2013 Edition) URL = <http://plato.stanford.edu/archives/sum2013/entries/formal-belief/>
- Kahneman, D. (2012), *Thinking, Fast and Slow*. London: Penguin Books.

- Kibble, R. (2005), 'Beyond BDI? Brandomian commitments for multi-agent communication', in: *Proceedings of Normative Multi-Agent Systems Workshop*, AISB Symposium, University of Hertfordshire.
- (2006a), 'Reasoning about propositional commitments in dialogue', in: *Research on Language & Computation*. Dordrecht: Springer.
- (2006b), 'Speech acts, commitment and multi-agent communication', in: *Computational & Mathematical Organization Theory*. Dordrecht: Springer.
- Levi, I. (1991), *The Fixation of Belief and its Undoing*. Cambridge: Cambridge University Press.
- (1997), *The Covenant of Reason*. Cambridge: Cambridge University Press.
- MacFarlane, J. (2010), 'Pragmatism and Inferentialism', in: Weiss, B. & Wanderer, J. (ed.), *Reading Brandom. On Making it Explicit*. London: Routledge: 81-95.
- (draft), 'In What Sense (If Any) Is Logic Normative for Thought?', URL = <http://johnmacfarlane.net/work.html>
- Manktelow, K. (2012), *Thinking and Reasoning: An Introduction to the Psychology of Reason, Judgment and Decision Making*. Hove: Psychology Press.
- Milne, P. (2009), 'What is the normative Role of Logic', in: *Aristotelian Society Supplementary Volume*, 83 (1): 269-98.
- Moshman, D. & Geil, M. (1998), 'Collaborative Reasoning: Evidence for Collective Rationality', in: *Thinking and Reasoning*, vol. 4 (3): 231-48.
- Oaksford, M. & Chater, N. (2003), 'Conditional Probability and the Cognitive Science of Conditional Reasoning', in: *Mind and Language*, 18: 359-79.
- (2007), *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.
- (2009), 'Precis of "Bayesian Rationality: The probabilistic approach to human reasoning"', in: *Behavioral and Brain Sciences*, 32: 69-120.
- (ed.) (2010), *Cognition and Conditionals. Probability and Logic in Human Thinking*. Oxford: Oxford University Press.
- Olsen, N. S. (2014). *Making Ranking Theory Useful for Psychology of Reasoning*. PhD dissertation, University of Konstanz.
- URL = <http://kops.uni-konstanz.de/handle/123456789/29353>
- Pfeifer, N. & Douven, O. (2014), 'Formal Epistemology and the New Paradigm Psychology of Reasoning', in: *Review of Philosophy and Psychology*, 5(2): 199-221.
- Pfeifer, N. & Kleiter, G. D. (2007), 'Towards a mental probability logic'.
- URL = <http://www.pfeifer-research.de/pdf/belg4.pdf>

[Updated version of: Pfeifer, N. & Kleiter, G. D. (2005), 'Towards a mental probability logic', in: *Psychologica Belgica*, 45(1): 71-99.]

Piwek, P. (2011), 'Dialogue Structure and Logical Expressivism', in: *Synthese*, 183(S1): 33-58.

--- (2014), 'Towards a Computational Account of Inferentialist Meaning', in: *The online proceedings of the AISB50 Convention*.

URL = <http://doc.gold.ac.uk/aisb50/AISB50-S21/AISB50-S21-Piwek-paper.pdf>

Raidl, E. & Skovgaard-Olsen, N. (forthcoming), 'Bridging Ranking Theory and the Stability Theory of Belief'.

Spohn, W. (2012), *The Laws of Belief. Ranking Theory and its Philosophical Applications*. Oxford: Oxford University Press.

Stalnaker, R. C. (1999), *Context and Content*. Oxford: Oxford University Press.

Walton, D. N. & Krabbe, E. C. W. (1996), *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*. Albany: State University of New York Press.

ⁱ Acknowledgment: this paper profited greatly from discussions with Wolfgang Spohn, Michael De, Lars Dänzer, Eric Raidl, and the other members of a reading group on *The Laws of Belief* at the University of Konstanz. I would also like to thank the participants at Thomas Müller's colloquium, Keith Stenning, Paul Piwek, the reviewers of *Synthese*, and the audience at AISB50 for insightful discussions.

ⁱⁱ Predecessors: in a way Levi (1991: ch. 2, 1997: ch. 1) was the first to emphasize that one could make progress with respect to the problem of logical omniscience by thinking of it in terms of commitments rather than in terms of belief. Subsequently, Milne (2009) has gone down a similar path. What the present investigation adds is giving it a more Brandomian spin (which was already implicit in Milne (2009)) and by formulating bridge principles that are capable of dealing with the constraints presented in MacFarlane (manuscript). After writing this paper, I discovered that Dutilh Novaes (forthcoming) has a very similar agenda, and it seems that the two papers complement each other very well.

ⁱⁱⁱ Explication of the framing effect: it has been shown that different ways of presenting the same information will give rise to different emotions, which in turn affects our judgments and decision making. Accordingly, the statement 'the chance of survival one month after surgery are 90%' will be found more reassuring than the equivalent statement 'mortality within one month of surgery is 10%' (Kahneman, 2012: 88). As a result, participants will respond differently to these two statements in spite of their logical equivalence.

^{iv} Reference: see also Pfeifer & Kleiter (2007).

^v Qualification: however, as one of the reviewers points out, one can find quotes in Oaksford & Chater (2009: 107-8) foreshadowing the observations Evans (2012) makes in the quote below. And indeed the same holds for Pfeifer & Kleiter (2007: 24). The point is just that the implications that such observations have for issues such as the problem of logical omniscience have not made their way into the general awareness of the psychological literature.

^{vi} On Brandom's notion of material inferences: it should be noted that material inferences are used as a generic notion for content-based inferences in the writings of Brandom. To be sure, Brandom does not accept the analytic/synthetic distinction for familiar Quinean reasons. But his notion of material inferences covers what would have traditionally been thought of as following in both of these categories. In his writings one thus not only finds examples of material inferences that sound like analytical inferences, like the example in the text, but also examples like inferring that a banana is ripe from its being yellow (Brandom, 2010: 104), which sounds like synthetic judgments with an inductive basis.

^{vii} Ranking theoretic explications based on Spohn (2012):

If we let ' $\mathbf{E}_{s1}(A)$ ' denote that the agent, S1, is entitled to A and ' $\mathbf{C}_{s1}(A)$ ' denote that S1 is committed to A, and we specify entitlements and commitments relative to the ranking function at the context of assessment, τ_{score} , we have that: $\mathbf{E}_{s1}(A) := \tau_{\text{score}}(A) > 0$. I.e. $\tau(A) > 0$ is satisfied at the context of assessment. A stricter threshold $b > 0$ could also have been chosen, but in ranking theory $\tau(A) > 0$ already indicates a full acceptance of A, and any value above 0 would just indicate how entrenched A is compared to other propositions that are likewise fully accepted.

Entitlement preservation: $\tau_{\text{score}}(C|\Gamma) > \tau_{\text{score}}(C|\Gamma^c)$, $\tau_{\text{score}}(C|\Gamma) > 0$. That is to say, if the degree of perceived justification of C is raised above the threshold by Γ , then an entitlement to Γ introduces an entitlement to C. Furthermore, in ranking theory we have the updating rule of result-oriented conditionalization (Spohn, 2012: section 5.4) to ensure that $\mathbf{E}_{s1}(\Gamma)$ introduces $\mathbf{E}_{s1}(C)$ for cases of $\mathbf{E}_{s1}(\Gamma)$ that are based on less than full certainty (thus paralleling Jeffrey conditionalization from note viii):

$$\kappa_{A \rightarrow n}(w) = \begin{cases} \kappa(w|A) & \text{for } w \in A \\ \kappa(w|\bar{A}) + n & \text{for } w \in \bar{A} \end{cases}$$

Commitment preservation: $\tau_{\text{score}}(C|\Gamma) > \tau_{\text{score}}(C|\Gamma^c)$, $\tau_{\text{score}}(C|\Gamma) = \infty$. That is to say, if the degree of perceived justification of C is raised to absolute certainty by Γ , then $\mathbf{C}_{s1}(\Gamma)$ introduces $\mathbf{C}_{s1}(C)$. To be sure, some threshold $c < \infty$ could also have been chosen. But here we preserve Brandom's idea that commitment preservation generalizes deductive relations.

Weak Incompatibility: $\tau_{\text{score}}(q|p) < \tau_{\text{score}}(q|\neg p)$, $\tau_{\text{score}}(q|p) < 0$

Strong Incompatibility: $\tau_{\text{score}}(q|p) < \tau_{\text{score}}(q|\neg p)$, $\tau_{\text{score}}(q|p) = -\infty$

Notice that the relationship between the probabilistic and rank-theoretic explications presented here is one of formulating similar ideas in competing formal frameworks. However, had the novel translation between the two in Raidl & Skovgaard-Olsen (forthcoming) been employed, the probabilistic explications would have had to be reformulated.

^{viii} Refinement through J-conditionalization: to allow for cases of entitlement to Γ where $P(\Gamma) < 1$, the second condition could be replaced by Jeffrey conditionalization as follows: $\sum_{i=1}^n [P_{\text{initial}}(q|\gamma_i) \cdot P_{\text{new}}(\gamma_i)] > b$, for $P_{\text{initial}}(\gamma_i) > 0$ and $\sum_{i=1}^n P_{\text{new}}(\gamma_i) > b$.

^{ix} Inferentialism as a probabilistic reason relation semantics: by exploiting the idea from Spohn (2012: ch. 6) that p is a reason *for* q whenever $\tau(q|p) > \tau(q|\neg p)$, and that p is a reason *against* q whenever $\tau(q|p) < \tau(q|\neg p)$, the weak and the strong notions of incompatibility are treated as cases of when p is an inductive or a deductive reason *against* q , and entitlement preservation and commitment preservation are treated as cases, where the set Γ counts as an inductive or a deductive reason *for* q . This explication treats inferentialism as a *probabilistic (or rank theoretic) reason-relations semantics*, and it is in general agreement with Dorn's (2005) account of the strength of arguments. However, this explication can only be partial, because it needs to be supplemented by Brandom's pragmatic account of the conditions under which the scorekeeper should add and subtract commitments and entitlements from the speaker's score, which Kibble (2005, 2006a, 2006b) and Piwek (2011, 2014) have begun to formalize. See also Walton & Krabbe (1996).

^x Clarification on assertion: actually on Brandom's view, making an assertion is to be viewed as putting forward a claim as something that the hearer can *use as a premise in his/her own reasoning* and not: putting it forward as *an uncontroversial starting point for further inquiry*. The reason why the latter formulation is preferred here is to bracket the issue of reductions. The point is that while reductions use the speaker's assertions as premises in one's own reasoning, the premises in reductions cannot be thought of as uncontroversial starting points for further inquiry. Rather I take it that reductions can be seen as a dialectical tool that scorekeepers use to show that there is a problem with the speaker's constellation of commitments. (I thank Michael De for forcing me to clarify this point.)

^{xi} Dispositional Beliefs vs. Occurrent Beliefs: actually a case could be made that this problem could be set aside as a misunderstanding. The reason is that the problem of cluttering up our minds with logical consequences seems to concern beliefs understood as *occurrent beliefs*, whereas formal epistemology is usually taken to model *dispositional beliefs*.

^{xii} Reference: <http://podcasts.ox.ac.uk/people/hartry-field>.

^{xiii} Separating a weak and a strong version: notice that it would also be possible to hold the view that the deontic score built up in the course of a conversation would be completely ruined by an inconsistency. Instead, a weaker version was put forward here, according to which entitlement is only withheld with respect to the assertions producing the inconsistency (e.g. p and q , where q entails non- p) and not with respect to the whole deontic score. However, repeated instances of such failures can diminish one's trust in the agent, which is why the idea of blacklisting recurrent sinners is introduced below.

^{xiv} Potential objection: here Paul Piwek (p.c.) raises the reasonable objection that the most sensible thing to do in cases where agents are confronted with paradoxical requirements may simply be to refrain from undertaking any public commitments at all and stick to their private beliefs. While this is a good point in general, I think it loses its intuitive force when applied to the preface paradox. Surely, we should not encourage authors not to write books or refrain from expressing a recognition of their own fallibility when doing so.

^{xv} Explication of the liar paradox: one version of the liar paradox runs as follows. The second sentence in this endnote is not true. Suppose the second sentence is true, then it is true that the second sentence is not true, and so the second sentence must not be true. Suppose it is not true, then things are as the second sentence says they are, and so it must be true.

^{xvi} Reference: (personal communication). Moreover, after writing this paper, I discovered that a similar response is suggested already in Walton & Krabbe (1996: 60).

^{xvii} Acknowledgement: thanks to one of the reviewers for pointing out this further possibility of a *de re* and *de dicto* ambiguity in the interaction between the Bruja and the scorekeeper.

^{xviii} Qualification: as one of the reviewers pointed out this may in fact be an idealization, because surely cases can arise, where there is some difference between the views of the author expressed in the book and the views of the author when writing the preface simply due to a temporal delay. In response, I would point out that the problems associated with acting as a scorekeeper on one's own views do not arise for cases, where one is dealing with the past self. Accordingly, the points made in the main text deal exclusively with the paradigmatic case, where an author corrects any claim that he deems to be false prior to the publication so that the views articulated in the book actually express the claims the author wants to commit himself to at the time of publication.

^{xix} Parallel to Moore's paradox: in exhibiting this difficulty in asserting something about one's own doxastic perspective that one would be able to assert about a foreign doxastic perspective, the preface paradox bears some resemblance to Moore's paradox. Moore's paradox consists in that we cannot assert sentences such as '*p*, but I do not believe that *p*' or '*p*, but I believe that non-*p*' without it sounding paradoxical—in spite of the fact that it is perfectly possible for any agent that *p* is the case and that this agent either believes that non-*p* or fails to believe that *p* (cf. Brandom 1994: 605). In both cases we seem to be faced with things that we know hold with respect to any other doxastic perspective (and *a fortiori* with respect to our own), which, however, we cannot assert directly about our own (present) doxastic perspective. Perhaps the best that the author can do is to restrict himself to counterfactuals about how he *would have* acted as a scorekeeper if the book *had been* written by someone else.

^{xx} Acknowledgement: I thank Paul Piwek for pointing out the need to introduce a qualification here.

^{xxi} Qualification: as pointed out by one of the reviewers, I may actually be weakening Brandom's position at this point, since he has quotes indicating that every committive-inferential consequence of an acknowledged commitment should be added to the deontic score (cf. Brandom, 1994: 190). However, in that case, I would hold that the present version of the position constitutes an improvement, insofar as it allows us better to deal with problem 4 for the scorekeeping perspective.

^{xxii} Acknowledgement: I thank Michael De for helping me to clarify this point.

^{xxiii} On commitments without an expiration date: as the practice of defending the works of deceased philosophers shows, the deontic score of an agent can outlive his/her biological time in virtue of other agents stepping in and administering the commitments of a deceased agent either as he/she would have been disposed to or in the way that would have been most optimal.

^{xxiv} Explication of the Wason selection task: in this task, the participants are presented with four cards, which have D, K, 3, and 7 respectively faced up and given the conditional rule 'If there is a D on one side of any card, then there is a 3 on its other side'. The task then consists in determining which cards to turn over to decide, whether the rule is true or false. To check for its falsity, the participants would have to select the D and the 7 card. Yet, most tend to select D and 3 (Manktelow, 2012: ch. 3).

^{xxv} Reference: Evans & Over (2004: 74).