Synthese (2019) 196:3433–3458
https://doi.org/10.1007/s11229-017-1604-8

Impossible worlds and partial belief

Edward Elliott[1]

Received: 19 April 2017 / Accepted: 23 October 2017 / Published online: 9 November 2017

Abstract One response to the problem of logical omniscience in standard possible worlds models of belief is to extend the space of worlds so as to include impossible worlds. It is natural to think that essentially the same strategy can be applied to probabilistic models of partial belief, for which parallel problems also arise. In this paper, I note a difficulty with the inclusion of impossible worlds into probabilistic models. Under weak assumptions about the space of worlds, most of the propositions which can be constructed from possible and impossible worlds are in an important sense *inexpressible*; leaving the probabilistic model committed to saying that agents in general have at least as many attitudes towards inexpressible propositions as they do towards expressible propositions. If it is reasonable to think that our attitudes are generally expressible, then a model with such commitments looks problematic.

Keywords Impossible worlds · Partial belief · Credence · Logical omniscience · Probabilistic coherence

Suppose we wish to model the total doxastic state of a typical (non-ideal) subject, whom we'll call α.[1] We'll need two main ingredients: one, a way to represent potential objects of thought, the kinds of things fit to serve as the contents of some cognitive mental state; and two, a way of representing which of these are the contents of α's attitudes.

[1] By 'total doxastic state' I mean the sum total of facts about the subject's doxastic attitudes broadly construed, i.e., α's full beliefs, partial beliefs, comparative degrees of confidence, and so on—generally, those aspects of her mental life which characterise how she takes the world to be.

✉ Edward Elliott
E.J.R.Elliott@leeds.ac.uk

[1] School of Philosophy, Religion and History of Science, University of Leeds, Room 3.11, Botany House, Leeds LS2 9JT, UK

If our model is to be faithful to the facts, then it's important that we don't end up representing α as being much more rational than she in fact is. What needs to be done to satisfy this desideratum depends on just how *ir*rational we think non-ideal agents can be, and opinions vary widely on this matter. But here is something that almost everyone agrees on: *we are not logically infallible*. The total doxastic state of any ordinary agent will usually be logically incoherent in some respect or other. Total belief sets probably aren't going to be closed under logical implication, even on those accounts that *seem* to make us look very rational indeed (e.g., Lewis 1982; Stalnaker 1984). And on the face of it, beliefs don't appear to be closed under even logical equivalence. The same applies to other kinds of doxastic attitudes: prima facie, one can be fully confident that *either it is raining or it's not* without thereby also being fully confident that *it's not the case that it's raining and not raining*. The intuitive data of logical incoherence and hyperintensionality needs to be accounted for—usually, by modelling the objects of belief using entities that cut finer than logical equivalence.

In this paper, I argue that one common strategy for modelling logically fallible agents and hyperintensional contents (*viz.*, through the use of impossible worlds) does not sit nicely with another very common approach to modelling total doxastic states (*viz.*, through the use of a numerically-valued function defined on a Boolean algebra of propositions; e.g., a probability function). Roughly, the source of the problem is that most of the propositions which can be constructed out of a sufficiently rich space of possible and impossible worlds are in a certain strong sense *inexpressible*, and any Boolean algebra defined on such a space will contain at least as many inexpressible propositions as expressible propositions. Since it's reasonable to think that most (if not all) of our doxastic attitudes are expressible, a model which commits us to widespread inexpressibility looks problematic. We can impose restrictions on the space of worlds which would prevent the inclusion of inexpressible propositions in the algebra, but only at the cost of reintroducing (a strong degree of) infallibility.

In Sect. 1, I outline an assumption about the expressibility of thought which be helpful in setting up my main argument. Then, in Sect. 2, I provide some background on the problems of logical omniscience as they apply to a standard way of modelling full belief, and discuss how the introduction of impossible worlds is supposed to help solve these problems. In Sect. 3, I introduce probabilistic analogues to the classical problems of logical omniscience, for which an analogous solution involving impossible worlds seems to apply. Finally, in Sect. 4 I present the central argument of the paper, and in Sects. 5 and 6, discuss responses.

Before moving on, it's worth noting some things that I'm *not* arguing. First, I do not think that the mere *existence* of inexpressible propositions should be considered problematic for the impossible worlds model—nor for that matter do I think that they would be especially problematic for the possible worlds model. I would not consider it a devastating problem if our formal models implied that inexpressible propositions exist, and could potentially serve as the objects of thought for some believers. I do, however, think that there is serious issue when our models commit us to saying that inexpressibility is the norm, and it is this problem that I intend to highlight here. (See Sect. 6 for more discussion on this point.) And second, my argument should not be read as being against the intelligibility of impossible worlds in general, nor do I want to claim that there are no benefits to including them within our ontology.

1 The expressibility hypothesis

In setting up my argument, I will presuppose the existence of an artificial language, $\mathcal{L}$, about which I will make some assumptions. $\mathcal{L}$ can be thought of as a class of declarative sentences, each a (possibly infinite) string of symbols taken from a (possibly infinite) alphabet, with a corresponding interpretation. We suppose that every sentence in $\mathcal{L}$ is non-ambiguous, precise, and for the sake of simplicity, context-independent. I'll stick to characterising $\mathcal{L}$ at the sentential level, since it is here that the issues we will be interested in arise. Nothing in what follows should be taken to suggest that there can be no quantifiers, modal operators, and so on, in $\mathcal{L}$.

Next, we will want $\mathcal{L}$ to be as expressive as possible with respect to α's (partial) beliefs, within the bounds allowed by the present assumptions.[2] The most straightforward version of my argument then proceeds on the basis of an assumption, which I will call the expressibility hypothesis: that $\mathcal{L}$ is *maximally expressive*, in the sense that for each distinct belief (or partial belief) that α has, there is a distinct sentence S in $\mathcal{L}$ which expresses the content of that exact belief and no other. $\mathcal{L}$ may be capable of saying much more than this as well, but to begin with we will assume that it is capable of saying at least this much.

Furthermore, besides having beliefs *simpliciter*, I assume that α can also have *negative* and *conjunctive* beliefs. For example, α might believe that *roses are red*, that *violets are blue*, and that *roses are red and violets are blue*, where the latter content intuitively has normative connections to the former two of the kind we might try to cash out in terms of conjunction introduction and elimination rules. If the content of the first belief is captured by a sentence S_1 of $\mathcal{L}$, and the content of the second by S_2, then we will use '$S_1 \wedge S_2$' to pick out the sentence (or a sentence) of $\mathcal{L}$ which express the third content. Likewise, if α comes to later believe that *roses are not red*, then there's another sentence, '$\neg S_1$', which expresses her changed belief.

In saying this, I'm not making any strong commitments in relation to the syntax of $\mathcal{L}$, which may consist entirely of 'atomic' sentences for all I've said here. But I see no good reason to think, *if* it is possible to have a language capable of expressing all of our beliefs at all, that there couldn't also be such a language which contains a unary connective and a binary connective corresponding to negation and conjunction respectively. Nor am I saying that α can *only* have atomic, negative, and conjunctive beliefs. She may also have *conditional* beliefs, e.g., a belief that *if roses are red then violets are blue*, where this is not just another way of saying that α believes that *it's not the case that: roses are red and violets are not blue*. In that case, we may also want to have primitive conditional sentences in $\mathcal{L}$. Likewise, α may believe that *roses are red or violets are blue*, where this is not the same thing as believing that *it's not the case that: roses are not red and violets are not blue*. We need not commit either way on these questions. It's perfectly reasonable to think that $\mathcal{L}$ has some non-trivial

[2] A note on this: I am ignoring any beliefs which might be, as Perry (1979) calls them, *essentially indexical*—e.g., the belief that *I am here*. The assumption that we can express irreducibly indexical beliefs in a language whose interpretation is by stipulation context-independent may rightly be doubted. But I am setting this complication aside because the arguments that follow can be naturally adapted to a centred worlds framework (see Lewis 1979), which would permit the inclusion of context-dependent sentences back into $\mathcal{L}$.

syntax at the sentential level. But we may well find that having just two connectives is fewer than we need to adequately distinguish between the full range of contents that a typical subject might believe, so we will remain neutral on just what that syntax is. (The upshot of these points will become apparent in the final paragraphs of Sect. 4.)

Whatever $\mathcal{L}$ is, it's obviously not English, nor any other natural language. But there is no need to interpret my talk of 'sentences' and 'languages' too closely on the model of natural languages. The 'language' in question may not be the sort of thing that any human being could speak, nor need it correspond very closely to the structure of thought. The 'sentences' may be purely mathematical objects, or arbitrary sets of abstracta. For example, one might want to simply let every object of belief just *be* a sentence of $\mathcal{L}$, and stipulate that every sentence expresses itself.[3] Alternatively, perhaps an appropriately constructed *Lagadonian* language would be expressive enough for our purposes.[4] In a series of recent works, Mark Jago has defended just this idea (see esp. his 2012; 2015a; b; cf. also Berto 2010). Indeed, the expressive richness of Jago's language is a central component of his use of ersatz possible and impossible worlds to model hyperintensional contents, in roughly the manner described in the next section. As he puts it, for sets of ersatz possible and/or impossible worlds to be an adequate model of hyperintensional content and to overcome the infamous 'problem of descriptive power', the world-building "language must be expressible enough to represent all of the possible and impossible situations we want to represent, and to represent distinct (possible or impossible) situations as distinct situations" (Jago 2015b, p. 718).

The reader may already be chomping at the bits to deny the expressibility hypothesis. I ask that they hold off their objections for now. I will return to discuss the matter in detail in Sect. 6, where I will argue three things—in order of importance,

(i) Although inconclusive, there are general reasons to accept the hypothesis.
(ii) There are prominent accounts of impossible worlds such that the hypothesis (or a close analogue thereof) is taken for granted, and would be difficult to deny.
(iii) Even if we ultimately ought to deny the hypothesis, the main thrust of the argument will be largely unchanged.

But that will have to wait until Sect. 6. The central thread of the argument goes through much smoother if we take the expressibility hypothesis for granted, and it's better to discuss the consequences of denying the hypothesis once its importance to my argument is clear.

[3] Compare the discussion on Daniel Nolan's account of impossible worlds in Sect. 6. Nolan constructs his space of (possible and impossible) worlds out of a 'language' consisting of the objects of thought directly, using a version of the Really Unrestricted Comprehension principle that I discuss below.

[4] A Lagadonian language is one wherein particulars are taken to be names of themselves, and properties and relations are taken to be predicates for themselves. For example, the content *Frank is taller than Mary* may (but need not) be treated as construction out of Frank, Mary, and the *is taller than* relation.

2 The problems of logical omniscience

Suppose that Ω is a non-empty space of possible worlds. I remain neutral as to what worlds are; what's important is just that a world ω is the kind of thing such that it makes sense to say of a declarative sentence S that S is true at ω. In calling Ω a set of *possible* worlds, I'm saying that for every $\omega \in \Omega$ and $S, S_1, S_2, \ldots \in \mathcal{L}$:

Non-Contradiction:
At most one of S or $\neg$S is true at ω
Maximal Specificity:
At least one of S or $\neg$S is true at ω
Closure under Implication:
If $S_1, S_2, \ldots$ are true at ω and jointly imply S, then S is true at ω

What happens at worlds with respect to sentences that are *not* in $\mathcal{L}$ won't be important for our purposes, so in the sequel it should be assumed that the sentences S_1, S_2, etc., that I quantify over are always members of $\mathcal{L}$. I'll also assume that the relevant notion of implication (here and throughout) is at least as strong as that of classical sentential logic. If need be, we can also throw some conceptual or metaphysical necessities in as well, so as to rule out worlds with, e.g., married bachelors, four-sided triangles, non-dihydrogen oxide water, and the like.

Call any subset of Ω a *proposition*. This is a stipulative usage: I do not assume that propositions are objects of thought. The powerset of Ω, $\wp(\Omega)$, contains every proposition that can be formed from the worlds in Ω. Every sentence S in $\mathcal{L}$ can be mapped to some (possibly empty) member of $\wp(\Omega)$ which contains all and only the worlds where what S says is true. Following a standard notation, we'll designate this 'truth set' of S using $\|S\|$. Given the three assumptions I've made about Ω, logically equivalent sentences will always have the same truth sets. Moreover, $\neg$ and $\wedge$ will correspond to the basic set operations of complementation and intersection in the following way:

(i) $\|\neg S\| = \|S\|^C$
(ii) $\|S_1 \wedge S_2\| = \|S_1\| \cap \|S_2\|$

With this in place, we might make a start on modelling a total doxastic state. Begin with a model that originates with (Hintikka 1962), which focuses solely on full belief. Again, α is our subject. For each world ω, we can associate α with exactly one proposition in $\wp(\Omega)$, which we'll label $R_\alpha(\omega)$. $R_\alpha(\omega)$ is taken to represent the way the world must be given all of α's beliefs at ω. The worlds in $R_\alpha(\omega)$ are α's *doxastically accessible worlds* (at ω). In order to capture what α believes at ω, we might first say that any proposition in $\wp(\Omega)$ of which $R_\alpha(\omega)$ is a subset represents a content that α believes. The upshot is a compact model of α's *total belief state*. Fix an appropriate space of worlds Ω and $R_\alpha(\omega)$, and the rest of the work is done automatically by the subset relation.

But that's a little too quick. Even supposing that there are enough propositions in $\wp(\Omega)$ to represent all objects of belief, it may still be the case that $\wp(\Omega)$ also contains many propositions that correspond to nothing that can properly be believed. Modelling objects of belief as sets of worlds does not commit one to saying that every

set of worlds models an object of belief, and it shouldn't be taken for granted that every way the world might be corresponds to something that α can believe.[5] So let's make a very minor adjustment to the basic Hintikkan model. Suppose that $\mathcal{B} \subseteq \wp(\Omega)$ contains just those propositions that do model genuine objects of belief, and say:

α believes P iff $R_\alpha(\omega) \subseteq P$ and $P \in \mathcal{B}$

If every proposition is thinkable, then the inclusion of $\mathcal{B}$ adds nothing to the original model; if not, $\mathcal{B}$ serves to filter out any 'unthinkable' propositions.

Now it's well known that this model suffers from a cluster of issues that usually come under the heading of the *problems of logical omniscience*. Let me highlight three examples:

(i) If S_1 implies S_2 and $\|S_1\|, \|S_2\| \in \mathcal{B}$, then α believes S_1 only if she also believes S_2
(ii) If S is a tautology and $\|S\| \in \mathcal{B}$, then α believes S
(iii) α's beliefs are inconsistent only if $R_\alpha(\omega) = \emptyset$ (so α believes everything in $\mathcal{B}$)

The first is a result of *Closure under Implication*, which ensures that if S_1 implies S_2, then $\|S_1\| \subseteq \|S_2\|$. Corollary: if S_1 and S_2 are logically equivalent, then $\|S_1\| = \|S_2\|$. *Maximal Specificity* and *Closure under Implication* together imply that if S is a tautology, then $\|S\| = \Omega$, and since Ω is a superset of any proposition in $\mathcal{B}$, this gives rise to our second problem. With the addition of *Non-Contradiction* we also get that if S is a contradiction, then $\|S\| = \emptyset$, which ultimately leads to the third problem. Indeed, *Non-Contradiction* alone says that $\|S\|$ and $\|\neg S\|$ are disjoint, so α can believe both S and ¬S only if $R_\alpha(\omega) = \emptyset$.

There's a number of ways we might try to respond to these problems. Perhaps the error is in thinking that we can adequately model belief sets using unstructured sets of possible worlds and simple subset relations. Or, perhaps the error is in thinking that we can use a *single* set of worlds $R_\alpha(\omega)$ to encode an agent's total doxastic state at ω, which may be better represented using multiple 'fragments'. Or perhaps there isn't really a problem here after all, we really *are* logically omniscient and it is only the complexities of belief attribution in natural language and our imperfect access to our own beliefs which makes it seem otherwise. I think that each of these captures part of the truth, but my intention for this paper is not to suggest a positive solution to the problems of logical omniscience. Instead, I wish to focus on one common response, which begins with the thought that perhaps there are *not enough* propositions in $\wp(\Omega)$: we need to make our space of worlds bigger, to accommodate more fine-grained divisions amongst the objects of thought.

Suppose we make an extension to Ω, such that it now contains not only all of the original possible worlds, but also worlds where various kinds of impossible affairs

[5] For instance, in response to the Russell–Kaplan paradox (see Davies 1981, p. 262; Kaplan 1995), Lewis (1986, pp. 104–107) argues that there are many more ways the world might be than there are possible functional roles, and hence more than there are possible belief contents—at least $\beth_3$ for the former, and probably no more than $\beth_0$ for the latter.

obtain.[6] To make sure that Ω is rich enough, we will want worlds which are obviously inconsistent (where both S and ¬S are true), as well as worlds which are inconsistent in more subtle ways (e.g., worlds where S_1 and S_2 are true, but $S_1 \wedge S_2$ is not true). Indeed, if we think that agents are capable of extreme logical incoherence, then we will want to ensure that our worlds are not closed under any non-trivial consequence relation. It would not be very helpful to remove closure under classical consequence but retain closure under, e.g., intuitionistic consequence—otherwise, we're just swapping one sort of logical omniscience for another.

To really free up the model, then, proponents of impossible worlds will typically posit a highly permissive comprehension principle, along the following lines:[7]

Unrestricted Comprehension:
For any maximal set of sentences $\mathcal{S} \subseteq \mathcal{L}$, there will be worlds in Ω where every $S \in \mathcal{S}$ is true and no $S \in \mathcal{L} \backslash \mathcal{S}$ is true

Now Ω contains every logically possible world, plus every maximally specific impossible world. (Some impossible worlds theorists choose to drop even *Maximal Specificity* to allow for *incomplete* worlds as well. Whether or not we include incomplete worlds in Ω won't make a difference to the arguments of this section.)

By building a model around this expanded space of worlds, it's easy to block all three of the unwelcome 'omniscience' problems noted earlier. Indeed, we can say more than this. Let $\{S_1, S_2, \ldots\}$ be any consistent or inconsistent set of sentences, and let $R_\alpha(\omega)$ be the intersection of $\|S_1\|, \|S_2\|, \ldots$. Now $R_\alpha(\omega)$ will be non-empty, and for any S that's not in $\{S_1, S_2, \ldots\}$, there will be at least one maximally specific world in $R_\alpha(\omega)$ where S is not true. So, regardless of what we take α's set of beliefs $\{S_1, S_2, \ldots\}$ to be, we will be able to find some $R_\alpha(\omega)$ such that $R_\alpha(\omega) \subseteq \|S\|$ if and only if α believes S. That looks like a nice property for our model to have, and all we had to do was load Ω up with enough impossible worlds.

But note a consequence of *Unrestricted Comprehension*: there is no sentence S—at least, no sentence in $\mathcal{L}$—such that S is true at all and only the worlds in $R_\alpha(\omega)$ (assuming that α believes more than one thing). Say that a proposition P is *expressible* (relative to $\mathcal{L}$) just in case there is a sentence $S \in \mathcal{L}$ such that $P = \|S\|$. The set of expressible propositions, $\{\|S\| : S \in \mathcal{L}\}$, is an antichain of $< \wp(\Omega), \subseteq >$: for *any* two distinct sentences S_1, S_2, there will be worlds in Ω where S_1 is true and S_2 isn't true; so, $\|S_1\|$ will *never* be a subset of $\|S_2\|$. Suppose that α believes S_1 and at least one other thing S_2. Whatever $R_\alpha(\omega)$ ends up being, it will have to be a proper subset of

[6] The use of impossible worlds to help solve the problem of logical omniscience and related problems in epistemic logic was explicitly introduced in Rantala (1982), although the idea can also be found in Hintikka (1975) and Creswell (1973). Numerous authors have since made use of the idea, and a recent defence can be found in a series of works by Jago (2009, 2013, 2014a, 2015b, 2015a) and Berto (2010). See also Nolan (1997, 2013), though Nolan's general focus is on using impossible worlds to give a Lewisian semantics for counterpossible conditionals. I am focusing on proponents of the so-called "American stance" on impossible worlds; my arguments are not intended to touch upon the "Australasian" use of impossible worlds *qua* basis for an interpretation of some non-classical logic.

[7] By 'maximal set of sentences' $\mathcal{S} \subseteq \mathcal{L}$, I mean a set such that for any $S \in \mathcal{L}$, at least one of S or ¬S is in $\mathcal{S}$.

both $\|S_1\|$ and $\|S_2\|$. So, there's no S_3 such that $\|S_3\| = R_\alpha(\omega)$. $R_\alpha(\omega)$ is *inexpressible* in $\mathcal{L}$.[8]

At this point, let me bring in the expressibility hypothesis: for every one of α's beliefs, $\mathcal{L}$ includes a sentence S which expresses exactly that belief. If this is reasonable, then it's only natural to suppose that a proposition should be found in $\mathcal{B}$ only if it is expressible in $\mathcal{L}$; that is, $\mathcal{B} \subseteq \{\|S\| : S \in \mathcal{L}\}$. After all, what could it mean to represent α as believing a proposition P, where P is not characterised by any sentence in a language which, *ex hypothesi*, is capable of expressing every one of α's beliefs? And, since $R_\alpha(\omega)$ is inexpressible, $R_\alpha(\omega) \notin \mathcal{B}$.

Is this a problem? I'm inclined to think that the inexpressibility of $R_\alpha(\omega)$ is not *by itself* problematic. It would perhaps have been problematic if we were forced to assume that $R_\alpha(\omega)$ must itself represent something that α believes, and hence that it should always be included within $\mathcal{B}$. However, nothing internal to the model I've described requires this to be the case. That $R_\alpha(\omega)$ should itself be a proposition that α believes was never a commitment of the original model, even when we were working with just possible worlds. What's needed for the representational system to work is that (a) if $\mathcal{P}_\alpha \subseteq \wp(\Omega)$ is the set of all and only those propositions towards which some agent α has beliefs at ω, then $\mathcal{P}_\alpha$ has some lower bound with respect to $\subseteq$ which we can designate as $R_\alpha(\omega)$; and (b) if $\mathcal{P}_\alpha \neq \mathcal{P}_\beta$, then $R_\alpha(\omega) \neq R_\beta(\omega)$. That is, every distinct total belief state can be uniquely represented by (at least one) set of doxastically accessible worlds. We can satisfy this by letting $R_\alpha(\omega)$ be the intersection of each proposition that α believes, without supposing that $R_\alpha(\omega)$ is itself something that α believes.

None of this is to say that the impossible worlds model of belief just developed is without problems—just that it doesn't commit us to saying that α believes something she cannot possibly believe. It is worth noting that if we can only believe expressible propositions, and no expressible proposition is a subset of any other expressible proposition, then there is a genuine question as to the *point* of using this kind of set-theoretic model to represent our beliefs in the first place. The machinery of set theory only comes into play at a single step, linking the (non-believed) proposition $R_\alpha(\omega)$ to the set of expressible propositions that α believes, the latter of which has no interesting set-theoretic structure. The only thing which unites the worlds in the proposition $R_\alpha(\omega)$ is that they are those worlds where each member of a set of sentences $S_1, S_2, S_3, \ldots$is true—and characterising *that* proposition amounts to just listing all and only those sentences which express something α believes. What we've done with $R_\alpha(\omega)$ and $\subseteq$, we could have done more perspicuously with a simple list of sentences. We gain nothing in economy by the addition of $R_\alpha(\omega)$, and modelling beliefs as supersets of $R_\alpha(\omega)$ doesn't seem to illuminate anything of interest.[9]

[8] Note that $R_\alpha(\omega)$ can be inexpressible even if we have a name a_i for each of the worlds within $R_\alpha(\omega)$ and $\mathcal{L}$ contains a way of saying "The actual world is a_1 or a_2 or …" (or something to that effect). Assuming that such a sentence exists in $\mathcal{L}$, if an unrestricted comprehension principle holds then the sentence will be true at *some* of the worlds in $R_\alpha(\omega)$, but it will also be false at some of those worlds (and true at some worlds outside of $R_\alpha(\omega)$).

[9] My complaint in this paragraph parallels one made by Bjerring and Schwarz in their (2017, §3). For discussion aimed at defending against these kinds of worries, see (Jago 2015a, pp. 595–597).

3 The problems of probabilistic coherence

So much for full belief. But if you're like me and you think that beliefs generally come in degrees (so that full belief is ultimately just a species of partial belief), then you will likely want your model of α's doxastic states *in general* to represent all of her partial beliefs, not just those that qualify as full beliefs. Luckily enough, there are natural ways to generalise the basic model outlined in the previous section. As Lewis puts it,

> [W]e must also provide for partial belief. Being a [doxastically accessible world] is not an all or nothing matter, rather it must admit of degree. The simplest picture, idealised to be sure, replaces the sharp-edged class of [doxastically accessible worlds] by a subjective probability distribution. …We can say that a [doxastically accessible world] *simpliciter* is a possible [world which] gets a non-zero (though perhaps infinitesimal) share of probability, but the non-zero shares are not all equal. (1986, p. 30)

In the rest of this paper, I want to focus on partial belief. In the present section, I will note how problems analogous to those of the traditional (full belief) problems of logical omniscience arise under a probabilistic model, and how different assumptions about the structure of Ω affect it.

For the sake of concreteness, I outline *one* way to generalise the full belief model to partial beliefs, along the lines suggested by Lewis. I want to stress that what follows is an illustrative example only: many of the specific details are not crucial to my main argument (e.g., the use of a probability mass function $\mathcal{D}$ to induce the credence function $\mathcal{C}r$). Readers already familiar with the idea of extending probability theory to an impossible worlds framework may choose to skim this section.

Let Ω be any non-empty space of possible and/or impossible worlds.[10] This time, instead of assigning a single proposition $R_\alpha(\omega)$ as α's doxastically accessible worlds, we will instead represent α's total doxastic state using a probability distribution $\mathcal{D}: \Omega \mapsto [0, 1]$. One could interpret $\mathcal{D}$ as representing α's degree of belief that the actual world is ω, for each ω in Ω, to the extent at least that (singleton sets of) worlds are to be included amongst the purported objects of partial belief. But this interpretation is unnecessary: $\mathcal{D}$, like $R_\alpha(\omega)$ earlier, should in the first instance be understood as a formal tool for modelling doxastic states in the manner to be outlined presently.

In the simplest case, $\mathcal{D}$ assigns 0 to all but countably many ω in Ω, and a real value between 0 and 1 to the remaining worlds such that those values sum to unity. We can then use $\mathcal{D}$ to induce a function $\mathcal{C}r$ on any subset $\mathcal{B}$ of $\wp(\Omega)$ by stipulating that for each $P \in \mathcal{B}$,

$$\mathcal{C}r(P) = \sum_{\omega \in P} \mathcal{D}(\omega)$$

Independent of any assumptions about what kinds of worlds are in Ω and what propositions get into $\mathcal{B}$, we know that $\mathcal{C}r$ will satisfy:

[10] To be clear: we are *not* making any assumptions yet about which worlds get into Ω; we will see how different assumptions about Ω impact upon the probabilistic model as we go along.

Nonnegativity:
If $\emptyset$ is in $\mathcal{B}$, then $\mathcal{C}r(\emptyset) = 0$
Normalisation:
If Ω is in $\mathcal{B}$, then $\mathcal{C}r(\Omega) = 1$
Monotonicity:
For all pairs P_1, P_2 in $\mathcal{B}$, if $P_1 \subseteq P_2$, then $\mathcal{C}r(P_1) \leq \mathcal{C}\mathrm{r}(P_2)$
Σ-Additivity:
If $\mathcal{P}$ is any countable set of disjoint propositions in $\mathcal{B}$ whose union ($\bigcup\ \mathcal{P}$) is also in $\mathcal{B}$, then $\mathcal{C}r(\bigcup\ \mathcal{P}) = \sum_{P \in \mathcal{P}}\ \mathcal{C}r(P)$

If $\mathcal{B}$ contains $\emptyset$, then $\mathcal{C}r$ is a measure on $\mathcal{B}$. But it's not *yet* a probability function, as usually understood. For that, we need to make the additional assumption that $\mathcal{B}$ is some Boolean sub-algebra of $\wp(\Omega)$. That is, given what we've just said, $\mathcal{C}r$ is a *probability function* just in case:

Booleanism:
For all P, P_1, $P_2 \in \wp(\Omega)$,

(i) If $P \in \mathcal{B}$, then $P^{\mathrm{C}} \in \mathcal{B}$
(ii) If P_1, $P_2 \in \mathcal{B}$, then $P_1 \cap P_2 \in \mathcal{B}$

From (i) and (ii), it follows that if P_1, $P_2 \in \mathcal{B}$, then $P_1 \cup P_2 \in \mathcal{B}$. *Booleanism* is standard for the large majority of models of partial belief and a background requirement for many of the results in probability theory. I return to discuss it again in Sects. 4 and 5. As we'll see, it leads to problems if we assume that Ω has a certain minimal structure, and that $\mathcal{B} \subseteq \{\|\mathrm{S}\| :\ \mathrm{S}\ \in\ \mathcal{L}\}$.

But for now, suppose only that $\mathcal{B}$ includes all and only those propositions towards which α has partial beliefs, whatever they may be. In that case, a very natural way to read $\mathcal{C}r$ is as a representation of α's total degree of belief state:

P is believed by α to degree x if and only if $\mathcal{C}r(P) = x$

This generalises the earlier model of full belief quite nicely. On the simplest generalisation, say that full belief equates to degree of belief 1. Then, we will be able to characterise $R_\alpha(\omega)$ as just that set of worlds which are assigned some positive value by $\mathcal{D}$; thus, $\mathcal{C}r(\|\mathrm{S}\|) = 1$ for every $\|S\| \in \mathcal{B}$ such that $R_\alpha(\omega) \subseteq \|\mathrm{S}\|$. But now we can also represent each of the many non-extremal grades of belief that α can have towards any proposition in $\mathcal{B}$, removing the sharp edges between belief and non-belief.

However, if Ω is a space of possible worlds, then it's easy to see that the new model will have its very own problems with logical omniscience. Corresponding to *Nonnegativity*, *Normalisation*, *Monotonicity* and *Σ-Additivity* respectively, we can quickly derive the following constraints of probabilistic coherence:

(i) If S is a contradiction and $\|\mathrm{S}\| \in \mathcal{B}$, then $\mathcal{C}r(\|\mathrm{S}\|) = 0$
(ii) If S is a tautology and $\|\mathrm{S}\| \in \mathcal{B}$, then $\mathcal{C}r(\|\mathrm{S}\|) = 1$
(iii) If S_1 implies S_2 and $\|\mathrm{S}_1\|$, $\|\mathrm{S}_2\| \in \mathcal{B}$, then $\mathcal{C}r(\|\mathrm{S}_1\|) \leq \mathcal{C}r(\|\mathrm{S}_2\|)$
(iv) If S_1 and S_2 are inconsistent and $\|\mathrm{S}_1\|$, $\|\mathrm{S}_2\|$, $\|\neg(\neg\mathrm{S}_1 \wedge \neg\mathrm{S}_2)\| \in \mathcal{B}$, then $\mathcal{C}r(\|\neg(\neg\mathrm{S}_1 \wedge \neg\mathrm{S}_2)\|) = \mathcal{C}r(\|\mathrm{S}_1\|) + \mathcal{C}r(\|\mathrm{S}_2\|)$

Additionally, if full belief is degree of belief 1, then the new model implies that it's not even possible for α to have inconsistent beliefs: $\mathcal{D}$ must assign a positive value to

at least one possible world ω, and the set of propositions P such that $\mathcal{C}r(P) = 1$ will be consistent. On an alternative account, full belief might be characterised in terms of exceeding some threshold degree t, for $t < 1$. In that case, there may be no $R_\alpha(\omega)$ such that α believes P if and only if $R_\alpha(\omega) \subseteq P$, and it may be possible for α's beliefs to be inconsistent. However, if $t > 0.5$, then it will be impossible for α to believe both S and ¬S simultaneously; and as long as $t > 0$, it will be impossible for α to believe any contradictions.

If logical omniscience is bad, then strict probabilistic coherence seems much worse. And the problems aren't limited to just probabilistic representations. For instance, Dubois and Prade's (1988) possibility theory allows us to systematically construct a degree of belief function on the basis of what they call a *possibility distribution*; i.e., a function $\mathcal{D}'$ from Ω into [0, 1] such that $\mathcal{D}'(\omega) = 1$ for at least one world ω. Taking $\mathcal{D}'$ as the basis for our model instead of $\mathcal{D}$, we can define $\mathcal{C}r$ on any subset $\boldsymbol{\mathcal{B}}$ of $\wp(\Omega)$ as follows:

$$\mathcal{C}r(\emptyset) = 0, \text{ and if } P \neq \emptyset, \text{ then } \mathcal{C}r(P) = sup\{\mathcal{D}'(\omega) : \omega \in P\}$$

Defining $\mathcal{C}r$ in this way implies that it is sub-additive:

If $P_1, P_2, P_1 \cup P_2 \in \boldsymbol{\mathcal{B}}$, then $\mathcal{C}r(P_1 \cup P_2) = max\{\mathcal{C}r(P_1), \mathcal{C}r(P_2)\} \leq \mathcal{C}r(P_1) + \mathcal{C}r(P_2)$

So, to a limited extent, using possibility distributions would let us avoid strict probabilistic coherence—though, sub-additivity is still a very strong constraint! More importantly, $\mathcal{C}r$ so-defined will still satisfy *Nonnegativity*, *Normalisation*, and *Monotonicity*, and so $\mathcal{C}r$ will still be constrained by (i)–(iii). In that sense, the possibilistic model still has to deal with a version of the problems of probabilistic coherence.

The same applies more generally: the vast majority of formal systems for the representation of partial beliefs will have $\mathcal{C}r$ satisfy at least one of *Nonnegativity*, *Normalisation*, and *Monotonicity* (or something very similar). For example, Choquet capacities (Choquet 1954; applied in, e.g., Tversky and Kahneman 1992), Dempster–Shafer belief and plausibility functions (Dempster 1968; Shafer 1976), ranking functions (Spohn 2012), and the set-valued functions of Levi (1974) and Kyburg (1992). Where Ω consists of only possible worlds, all of these models will have to deal with very strong coherence constraints.

But never fear—impossible worlds to the rescue! If we were to instead define the probability distribution $\mathcal{D}$ on a space of worlds Ω that satisfies *Unrestricted Comprehension*, then $\mathcal{C}r$ need not satisfy any of the constraints (i)–(iv). Indeed $\mathcal{C}r$ can be *almost* as wild and wacky as we want it to be. For instance, suppose that $\mathcal{D}$ assigns a positive value only to worlds where S and S ∧¬S are both true, and never to worlds where ¬S or ¬(S ∧¬S) are true. Now, assuming that all of the relevant propositions are in $\mathcal{C}r$'s domain, $\mathcal{C}r(\|\mathrm{S}\|) = \mathcal{C}r(\|S \wedge \neg\mathrm{S}\|) = 1$, and $\mathcal{C}r(\|\neg\mathrm{S}\|) = \mathcal{C}r(\|\neg(\mathrm{S} \wedge \neg\mathrm{S})\|) = 0$.

Proviso: if *Maximal Specificity* holds and $\|S\|, \|\neg\mathrm{S}\| \in \boldsymbol{\mathcal{B}}$, then $\mathcal{C}r(\|S\|) + \mathcal{C}r(\|\neg\mathrm{S}\|) \geq 1$.[11] So *Unrestricted Comprehension* does not give us total freedom

[11] *Maximal Specificity* says that $\|\mathrm{S}\| \cup \|\neg\mathrm{S}\| = \Omega$. *Normalisation* plus Σ *-Additivity* then imply that $\mathcal{C}r(\|\mathrm{S}\| - \|\neg\mathrm{S}\|) + \mathcal{C}r(\|\neg\mathrm{S}\| - \|\mathrm{S}\|) + \mathcal{C}r(\|\mathrm{S}\| \cap \|\neg\mathrm{S}\|) = 1$. Since $\mathcal{C}r(\|\mathrm{S}\| \cap \|\neg\mathrm{S}\|) \geq 0, \mathcal{C}r(\|\mathrm{S}\|) \geq \mathcal{C}r(\|\mathrm{S}\| - \|\neg\mathrm{S}\|)$ and $\mathcal{C}r(\|\neg\mathrm{S}\|) \geq \mathcal{C}r(\|\neg\mathrm{S}\| - \|\mathrm{S}\|)$, it follows that $\mathcal{C}r(\|\mathrm{S}\|) + \mathcal{C}r(\|\neg\mathrm{S}\|) \geq 1$.

to let $\mathcal{C}r$ assign values to expressible propositions however we like. But we can fix this if we're willing to expand Ω even further, to allow for non-maximal worlds:

> **Really Unrestricted Comprehension:**
> For any set of sentences $\boldsymbol{\mathcal{S}} \subseteq \boldsymbol{\mathcal{L}}$, there will be worlds in Ω where every $S \in \boldsymbol{\mathcal{S}}$ is true and no $S \in \boldsymbol{\mathcal{L}} \backslash \boldsymbol{\mathcal{S}}$ is true

Now if you want $\mathcal{C}r$ to assign 0 to both $\|S\|$ and $\|\neg S\|$, you just need to make sure that $\mathcal{D}$ assigns positive values only to worlds where neither S nor $\neg S$ is true. More generally, for *any* way you might want $\mathcal{C}r$ to distribute values across a countable set of expressible propositions, we'll be able to find a $\mathcal{D}$ which generates exactly that distribution. A quick example to demonstrate the point. Let S_1, S_2 and S_3 be any three distinct sentences whatsoever. Suppose we want a $\mathcal{C}r$ such that:

$$\mathcal{C}r(\|S_1\|) = x, \mathcal{C}r(\|S_2\|) = y, \mathcal{C}r(\|S_3\|) = z, \text{ and } \mathcal{C}r(P) = 0 \text{ otherwise,}$$

where $x > y > z \geq 0$. To accomplish this, we let $\mathcal{D}$ be as follows. Where ω_1 is the world where only S_1, S_2 and S_3 are true, $\mathcal{D}(\omega_1) = z$. Where ω_2 is the world where only S_1 and S_2 are true, $\mathcal{D}(\omega_2) = y - z$. Where ω_3 is the world where only S_1 is true, $(\mathcal{D}\omega_3) = x - y$. The 'empty world' (where no sentences whatsoever are true) is then assigned $1 - x$, and every other world is assigned 0. It follows that $\mathcal{C}r(\|S_1\|) = x, \mathcal{C}r(\|S_2\|) = y, \mathcal{C}r(\|S_3\|) = z$, and $\mathcal{C}r(P) = 0$ otherwise. Given my assumptions about $\mathcal{D}$, the same basic trick can be adopted for any $\mathcal{C}r$ that assigns a positive value to countably many expressible propositions.

The idea to use a probability function over a space of possible and impossible worlds in order to model probabilistically incoherent agents is common in conversation, but also shows up at several points in the literature. Cozic (2006) has recently advocated the strategy, and Halpern and Pucella (2011, §4) make similar points. Lipman (1997) and (1999) attempts to deal with logical non-omniscience by deriving a probabilistic expected utility representation from an agent's preferences, where the probability function in question is defined over a state-space involving both possibilities and impossibilities. Easwaran (2014, esp. pp. 1–2, 29) also suggests using impossible worlds in our probabilistic models of agents' doxastic states, albeit in a slightly different context.

At the risk of belabouring a point that will already be clear to many, let me summarise the discussion of this section. We can see the 'problems of probabilistic coherence' as a consequence of a sequence of modelling choices. First, we need to choose what kinds of worlds get into Ω. Second, we need to define the function $\mathcal{C}r$, and characterise the structure of its domain, $\boldsymbol{\mathcal{B}}$. And finally, we need to say something about how we are going to interpret $\mathcal{C}r$. In this respect, things are closely analogous to the problems of logical omniscience, and the same basic strategies for response are applicable. The response we've discussed centres upon the first modelling choice: by introducing enough impossible worlds into Ω, we can avoid all of the probabilistic coherence constraints (i) through (iv) above, and indeed, we can make $\mathcal{C}r$ appear as irrational as we like.

4 The problem of inexpressibility

In this section, I will argue that if Ω satisfies a very weak (and very plausible) richness assumption, then either *Booleanism* is false, or our model won't plausibly represent highly logically fallible agents—which, of course, was the central motivation for introducing impossible worlds in the first place. The most straightforward way to make the argument begins with the premise that whatever $\mathcal{B}$ is, it should contain only propositions which are expressible in $\mathcal{L}$.

For any S_1, take the set of all worlds in Ω where S is true, and consider its complement $\|S_1\|^C$. If *Unrestricted Comprehension* holds, then there is no S_2 such that $\|S_2\| = \|S_1\|^C$. As we've already noted, for any pair of sentences S_1 and S_2, there will be worlds where S_1 and S_2 are both true. And if *Really Unrestricted Comprehension* also holds, then there will be also be worlds where neither S_1 nor S_2 is true. In either case, $\|S_1\|$ and $\|S_2\|$ cannot be complements of one another. Hence, if $\|S_1\|$ is expressible, then $\|S_1\|^C$ is inexpressible. And since we've assumed that $\mathcal{B}$ is closed under complementation, it follows that there must be at least as many inexpressible propositions in $\mathcal{C}r$'s domain as there are expressible propositions. And that's not a nice result: $\mathcal{L}$ is supposed to include a sentence capable of expressing every object of thought towards which we might have partial beliefs, and yet the model we've now developed is assigning nonsensical values to propositions expressed by no sentences of $\mathcal{L}$.

We could get around the foregoing argument if (and only if) we adopt the following restriction on Ω:

> **Restriction R1:**
> For every S_1 such that $\|S_1\| \in \mathcal{B}$, there is an S_2 such that for any $\omega \in \Omega$, exactly one of S_1 or S_2 is true

I'll have more to say about R1 in a moment, but first, note that merely imposing R1 on Ω won't solve all our problems. We've also supposed that $\mathcal{B}$ is closed under (at least finite) intersections, and with only R1 in place the set of expressible propositions (in $\mathcal{B}$) will still be an antichain of $\langle \wp(\Omega), \subseteq \rangle$. (The only difference from before is that $\{\|S\| : S \in \mathcal{L}\}$ will now be closed under complementation.) So take any two sentences S_1 and S_2 such that $\|S_1\| \neq \|S_2\|$: there is no S_3 such that $\|S_3\| = \|S_1\| \cap \|S_2\|$. After all, nothing about R1 implies that there must be any sentences in $\mathcal{L}$ which are true at a world if and only if two other sentences are true at that world. Likewise, there is no S_3 such that $\|S_3\| = \|S_1\| \cup \|S_2\|$. Consequence: even with R1 in place, there will *still* be at least as many inexpressible propositions in $\mathcal{C}r$'s domain as there are expressible propositions.

The following is necessary and sufficient for ensuring that the intersection of any two expressible propositions (in $\mathcal{B}$) is itself expressible:

> **Restriction R2:**
> For every pair S_1, S_2 such that $\|S_1\|, \|S_2\| \in \mathcal{B}$, there is an S_3 such for any $\omega \in \Omega$, S_1 and S_2 are both true at ω if and only if S_3 is true at ω

Given R1, R2 also implies that the union of any two expressible propositions (in $\mathcal{B}$) is expressible. That is, for any pair of expressible propositions $\|S_1\|, \|S_2\|$ in $\mathcal{B}$, there

is some sentence S_3 such that S_3 is true at ω if and only if at least one of S_1 or S_2 is true at ω.

Exactly how restrictive R1 and R2 end up being depends heavily on which expressible propositions end up included in $\mathcal{B}$. We can safely assume that whatever $\mathcal{B}$ is, it will be richly populated with plenty of expressible propositions, so R1 and R2 are never trivially satisfied. On the other hand, if there are sentences whose characteristic propositions are not in $\mathcal{B}$, then R1 and R2 are consistent with certain a degree of freedom in relation to those sentences. But this is not especially interesting: since $\mathcal{B}$ contains all of the propositions in $\mathcal{C}r$'s domain, whatever is true of the expressible propositions *not* in $\mathcal{B}$ will be irrelevant to the model of α's degrees of belief that we are left with. Hence, we can simplify the discussion and pretend henceforth that $\mathcal{B} = \{\|S\| : S \in \mathcal{L}\}$.

The key point in what follows will be that how R1 and R2 can be implemented is constrained by what kinds of worlds we want to *keep* in Ω. For example, if we were to require that Ω contains at least all of the logically possible worlds, then the S_2 referred to in R1 must be logically equivalent to $\neg S_1$ (if not identical to $\neg S_1$): every logically possible world where S_1 doesn't hold is one where $\neg S_1$ holds, and if S_2 and $\neg S_1$ are true at the very same logically possible worlds then they must be logically equivalent.

I will not assume that Ω contains every logically possible world, though I think that something in the vicinity must be true if we want to use $\mathcal{C}r$ as a model of ideal agents as well as non-ideal agents. Instead, I will assume something much weaker. Say that S_1 is *blatantly inconsistent* with S_2 just in case either $S_1 = \neg S_2$ or $S_2 = \neg S_1$. Then my assumption can be expressed as follows:

Minimal Richness:
For any consistent triple S_1, S_2, S_3, there is at least one world $\omega \in \Omega$ such that:

(i) S_1, S_2, and S_3 are all true at ω, and
(ii) If S_4 is blatantly inconsistent with any of S_1, S_2, or S_3, then S_4 is not true at ω

Minimal Richness should be uncontroversial, *especially* since it can be motivated by precisely the same sorts of considerations which motivate including a rich space of impossible worlds into our models in the first place.[12] Consider: if S_1, S_2, and S_3 are

[12] A referee suggests in response to this point that there may be limits on our capacity to believe or have varying degrees of belief towards multiple contents simultaneously even when they're jointly consistent; e.g., if the contents expressed by S_1, S_2, and S_3 are each particularly complex, then representational storage limits could prevent all three from being simultaneously believed to some positive degree or other. In that case, it may not be possible for α to have confidence regarding each of S_1, S_2, and S_3 at the same time, undercutting any immediate formal need for having a world ω in Ω such that each of S_1, S_2, and S_3 are true.

There may well be representational storage limits, as a contingent matter of fact, for certain kinds of non-ideal agents. But suppose we restate *Minimal Richness* such that it quantifies only over triples S_1, S_2, S_3 such that it is possible for α to have doxastic attitudes towards the contents expressed by S_1, S_2, and S_3 simultaneously. Now, the restricted richness condition in conjunction with R1 will imply that if S_2 is true at all and only the worlds where S_1 is not true, then *either* S_2 is logically equivalent to $\neg S_1$, *or* it's not possible for α to have attitudes regarding S_2 while having attitudes regarding S_1. Likewise, given R2 the restricted version of the condition implies that if S_3 is true at all and only the worlds where S_1 and S_2 are each true for a pair S_1 and S_2 which can be simultaneously entertained, then *either* S_3 is logically equivalent to $S_1 \wedge S_2$, *or* α cannot have attitudes regarding S_3 while also having attitudes towards S_1 and S_2.

In each case, the latter disjunct would already be problematic given *Booleanism*. If $\mathcal{C}r(\|S_1\|)$ is defined then so is $\mathcal{C}r(\|S_1\|^C)$. That is, if $\|S_1\|$ is in $\mathbf{B}$, then $\|S_1\|^C$ is in $\mathbf{B}$, so we should want to be able to say that

jointly consistent, then it is surely possible for α to have a confidence of, say, greater than 2/3 in their simultaneous truth, which will only be possible if there is a world in Ω where each of the three sentences is true.[13] Similarly, it's surely possible to have the same high degree of confidence regarding their simultaneous truth while having zero confidence towards any S_4 that's blatantly inconsistent with S_1, S_2, or S_3. And this is would only be possible if there's a world where S_1, S_2, and S_3 are all true and S_4 isn't—for otherwise, $\|S_1\| \cap \|S_2\| \cap \|S_3\| \subseteq \|S_4\|$, and since $\mathcal{C}r(\|S_1\| \cap \|S_2\| \cap \|S_3\|) > 0$, we know that $\mathcal{C}r(\|S_4\|) > 0$.

So let's consider R1, which states that every S_1 can be paired with another sentence S_2 which is true at a world ω if and only if S_1 is not true at ω. If *Minimal Richness* holds, then whatever S_2 ends up being, it must be logically equivalent to $\neg S_1$. For suppose that S_2 is not logically equivalent to $\neg S_1$. Then either S_2 does not imply $\neg S_1$, or $\neg S_1$ does not imply S_2 (or both). If S_2 does not imply $\neg S_1$, then $\{S_2, S_1\}$ is consistent, and there will be at least one world where S_2 and S_1 are both true, which contradicts R1. On the other hand, if $\neg S_1$ does not imply S_2, then $\{\neg S_1, \neg S_2\}$ is consistent and there will be worlds where $\neg S_1$ and $\neg S_2$ are both true. Since S_1 and S_2 are blatantly inconsistent with $\neg S_1$ and $\neg S_2$ respectively, this would have to be a world where neither S_1 nor S_2 is true, which also contradicts R1. Hence, any sentence S_2 that satisfies R1 must be logically equivalent to $\neg S_1$, if *Minimal Richness* is true.

This leaves us with a limited range of options for implementing R1. The most straightforward way would be to let the required sentence S_2 just *be* $\neg S_1$. In effect, this is just to assume that the worlds in Ω satisfy *Non-Contradiction* and *Maximal Specificity*. And it's easy enough to think of some plausible motivations for assuming *Non-Contradiction*: one *could* argue that no model of a minimally rational agent's doxastic state should represent her as having any degree of belief that both S_1 and $\neg S_1$ could be true simultaneously (cf. Lewis 2004; Bjerring 2013; Jago 2014b). To the extent that we make errors of logical reasoning, they tend to be more subtle—e.g., a failure to deduce a downstream consequence of what we believe, rather believing in blatant inconsistencies.

Motivating *Maximal Specificity* is a little more difficult, as it amounts to removing all incomplete worlds from Ω. Some are independently happy to do this (e.g., Bjerring 2014; Bjerring and Schwarz 2017, p. 28; cf. Stalnaker 1996). For others, incomplete worlds are a crucial aspect of the model (Jago 2014a, b). Furthermore, it'll be a consequence of assuming *Non-Contradiction* and *Maximal Specificity* together that we lose the capacity to have $\mathcal{C}r$ assign wholly independent values to the pairs $\|S\|$ and

Footnote 12 continued
α can have attitudes towards both propositions simultaneously. So, whatever sentence S_2 holds at all and only the worlds where S_1 doesn't hold had better be logically equivalent to $\neg S_1$; and likewise regarding $\|S_3\| = \|S_1\| \cap \|S_2\|$. Furthermore, the main upshot of the discussion that follows is that if the relevant sentences S_2 and S_3 are not $\neg S_1$ and $S_1 \wedge S_2$ respectively, then the worlds we are left with in Ω are closed under apparently quite arbitrary inference rules which we have no good reasons to believe are adhered to in general by ordinary agents. Were we to suppose that whenever S_1 and S_2 are true, there's a third sentence S_3 which is also true such that (i) S_3 is not logically equivalent to $S_1 \wedge S_2$ *and* (ii) cannot be entertained by α alongside S_1 and S_2, then we won't have closed Ω under any less baffling inference patterns.

[13] The proof of this is straightforward given that $\mathcal{C}r$ is a probability function. Suppose that $\mathcal{C}r(P_1) = \mathcal{C}r(P_2) = \mathcal{C}r(P_3) > 2/3$. Then $\mathcal{C}r(P_1 \cap P_2) > 1/3$, and since $\mathcal{C}r(P_3 \cap (P_1 \cap P_2)) = \mathcal{C}r(P_3) + \mathcal{C}r(P_1 \cap P_2) - \mathcal{C}r(P_3 \cap (P_1 \cap P_2)) = 1$, $\mathcal{C}r(P_3 \cap (P_1 \cap P_2))$ must be greater than 0.

$\|\neg S\|$. Indeed, the worlds we are left with are closed under the rules of *double negation introduction* and *elimination*, with $\mathcal{C}r$ satisfying $\mathcal{C}r(\|S\|) = \mathcal{C}r(\|\neg\neg S\|)$ for all $\|S\|$ in $\mathcal{B}$. This is already quite a strong restriction.

Nevertheless, there are good reasons to think that if the implementation of R1 is to be even remotely well-motivated, then S_2 shouldn't be anything *other* than $\neg S_1$. Suppose that S_2 is any sentence that's logically equivalent to $\neg S_1$ other than $\neg S_1$ itself—say, $\neg\neg\neg S_1$. We might then keep some non-maximally specific and/or contradictory worlds in Ω, but now our worlds will be closed under the rules of *sextuple negation introduction* (SNI) and *elimination* (SNE):

(SNI) From S, infer $\neg\neg\neg\neg\neg\neg S$
(SNE) From $\neg\neg\neg\neg\neg\neg S$, infer S

Any reasons we might have had to avoid closing worlds under the (relatively simple) rules of double negation would apply with all the more force here: to the extent that ordinary agents might generally accept something like (SNI) and (SNE), it's *because* they accept that S_1 is true if and only if $\neg S_1$ is not true. Given *Minimal Richness*, the very best case we can make for implementing R1 involves letting S_2 be $\neg S_1$. Anything else would look implausible and arbitrary.

But it is in combination with R2 that R1 most worrisome. R2 states that every pair of sentences S_1, S_2 can be paired with a some S_3 such that S_3 is true at a world if and only if both S_1 and S_2 are true at that world. Given *Minimal Richness*, we know that S_3 must be logically equivalent to $S_1 \wedge S_2$. The argument here is similar to the one earlier with R1. Suppose that S_3 is not logically equivalent to $S_1 \wedge S_2$. Then S_3 doesn't imply $S_1 \wedge S_2$, or $S_1 \wedge S_2$ doesn't imply S_2. If S_3 doesn't imply $S_1 \wedge S_2$, then at least one of the following is consistent:

$$\{S_3, \neg S_1, \neg S_2\},\ \{S_3, \neg S_1, S_2\},\ \{S_3, S_1, \neg S_2\}$$

In each case, there will be at least one world in Ω where S_3 is true and at least one of S_1 or S_2 is not true, which would contradict R2. If $S_1 \wedge S_2$ does not imply S_3, then S_1 and S_2 do not jointly imply S_3, so $\{S_1, S_2, \neg S_3\}$ is consistent and there is at least one world in Ω where S_1 and S_2 are both true and S_3 is not. This would also contradict R2. So, S_3 must be at least logically equivalent to $S_1 \wedge S_2$.

An argument analogous to that given for R1 then immediately suggests how we ought to implement the restriction, if at all: require that all worlds in Ω satisfy $\wedge$-*Consistency*:

$\wedge$-Consistency:
For all $S_1, S_2 \in \mathcal{L}$, S_1 and S_2 are both true at ω if and only if $S_1 \wedge S_2$ is true at ω

Certainly, it would be absurd to suppose that R2 is not satisfied by $S_1 \wedge S_2$, but rather some other sentence equivalent to $S_1 \wedge S_2$. For suppose that R2 was satisfied by, say, $\neg(\neg S_1 \wedge S_2) \wedge \neg(\neg S_2 \wedge S_1) \wedge \neg(\neg S_1 \wedge \neg S_2)$. Then our models would have us representing α as someone who, without fail, always infers back and forth between S_1, S_2 and $\neg(\neg S_1 \wedge S_2) \wedge \neg(\neg S_2 \wedge S_1) \wedge \neg(\neg S_1 \wedge \neg S_2)$, while potentially skipping over the much more natural and direct inferences between S_1, S_2 and $S_1 \wedge S_2$. But anyone who doesn't reliably follow the rules of *conjunction introduction* and *elimination*

is not going to be unfailingly adhere to any inference rules which link S_1, S_2 and $\neg(\neg S_1 \wedge S_2) \wedge \neg(\neg S_2 \wedge S_1) \wedge \neg(\neg S_1 \wedge \neg S_2)$ to one another. (To be sure, one *could* in principle describe a consequence relation such that the later inferences are admitted but the former are not. But why would we think that closing the worlds in Ω under *that* relation makes for a good model any doxastic agent, let alone the typical believer?)

In conjunction with *Non-Contradiction* and *Maximal Specificity*, $\wedge$-*Consistency* guarantees that $\neg(\neg S_1 \wedge \neg S_2)$ is true at any world where at least one of S_1 or S_2 are true: for any $\|S_1\|$ and $\|S_2\|$,

$$\|S_1\|^C = \|\neg S_1\|, \text{ and } \|S_1\| \cap \|S_2\| = \|S_1 \wedge S_2\|,$$

Hence,

$$(\|S_1\|^C \cap \|S_2\|^C)^C = \|\neg(\neg S_1 \wedge \neg S_2)\| = \|S_1\| \cup \|S_2\|$$

In fact, they imply that (a) every Boolean combination of expressible propositions will be expressible by some sentence involving $\neg$ and/or $\wedge$, and more generally that (b) every world in Ω will be closed under the $\{\neg, \wedge\}$ fragment of classical propositional logic. We're fast running out of impossibilities—and with them, our capacity to represent logically non-ideal subjects.

Now I want to be clear that I've not yet said that Ω contains no impossible worlds whatsoever. If there are irreducibly disjunctive sentences in $\mathcal{L}$, then a sentence like $S_1 \vee S_2$ may still behave erratically by, e.g., not being true at all and only the worlds where at least one of S_1 or S_2 is true. Likewise, if $\mathcal{L}$ contains a primitive conditional connective $\rightarrow$ (i.e., where $S_1 \rightarrow S_2$ is *not* simply a shorthand for $\neg(S_1 \wedge \neg S_2)$), then we've not said anything to guarantee that the worlds in Ω must validate even very simple inference rules like *modus ponens*. Thus, there may still be plenty of logically impossible worlds in Ω. Nevertheless, with *Non-Contradiction* and *Maximal Specificity*, $\wedge$-*Consistency* alone we've managed to close Ω under a very strong consequence relation. Indeed, Ω is already only apt for modelling agents who are very good logical reasoners: for every classically valid inference pattern $S_1, S_2, \ldots \Rightarrow S$, the worlds in Ω will be closed under an corresponding inference which replaces each of $S_1, S_2, \ldots$and S with a classically equivalent sentence expressed using only $\neg$ and $\wedge$. For instance, while Ω might not be closed under *disjunction introduction*, we do know that at any world where either S_1 or S_2 is true, $\neg(\neg S_1 \wedge \neg S_2)$ will also be true. And at any world where S_1 and S_2 are true, $\neg(\neg S_1 \wedge \neg S_2) \wedge \neg(\neg S_1 \wedge S_2) \wedge \neg(S_1 \wedge \neg S_2)$ is true. What we have, in effect, is a model of an agent who is logically infallible with respect to a huge range of sometimes very complex inferences. That the agent might *also* be logically incompetent with respect to other very basic inferences hardly seems to help.

In summary: given *Minimal Richness*, if we want to preserve *Booleanism* alongside the expressibility hypothesis, then we have to close Ω under some (classically valid) inferences. We have a certain degree of choice as to what inferences these might be (e.g., *double negation elimination* versus *sextuple negation elimination*). But closing Ω under the most simple and natural rules—that is, those rules which ordinary agents are most likely to consistently follow—leads us directly into closing Ω under a complete fragment of classical logic, and, plausibly, under classical logic *simpliciter*.

5 Responses

At the end of Sect. 3, I noted that the problems of probabilistic coherence result from a sequence of choices, about the formal properties and interpretation of Ω, $\mathcal{B}$, and $\mathcal{C}r$. All standard models of partial belief presuppose that $\mathcal{B}$ satisfies *Booleanism*, and $\mathcal{C}r$ satisfying at least one of *Nonnegativity*, *Normalisation*, and *Monotonicity* or something very similar; combined with a space of worlds limited only to the possible, these quickly get us to some very strong coherence constraints on degrees of belief. We can avoid these constraints without making any significant changes to the standard models if Ω includes enough impossible worlds, but doing so will generate a problem with expressibility.

There are a lot of moving parts here, and consequently, plenty of ways to respond. As a (non-exhaustive) list of options, we might:

1. Keep the standard probabilistic model of partial belief, and bite the bullet on the matter of probabilistic coherence.
2. Develop a non-standard model of partial belief which keeps *Booleanism* but avoids the probabilistic coherence without resorting to impossible worlds.
3. Develop a non-standard model of partial belief which involves impossible worlds but doesn't presuppose *Booleanism*.
4. Offer an alternative interpretation of $\mathcal{C}r$ (i.e., such that $\mathcal{C}r$ being defined for inexpressible propositions does not conflict with the expressibility hypothesis).
5. Reject the expressibility hypothesis.

I'm inclined to think that some combination of the first strategy and (to a much lesser extent) the second strategy is our best bet. It would be better if we didn't have to throw out most of what we've managed to achieve regarding the formal representation of partial belief, so making very significant changes to the basic model outlined in Sect. 3 seems like a rash decision. But moreover, just how bad the problems of probabilistic coherence actually are depends on just how probabilistically irrational the typical human is, and there are reasons to think that the probabilistic (possible worlds) model isn't too far from the truth (appearances to the contrary notwithstanding). But that is a big debate, and arguing the point is best left for a different discussion. To conclude, then, I will in this section say a few words about the third and fourth types of response, and discuss the fifth type of response in Sect. 6.

With respect to the third strategy, it's worth noting that *Booleanism* is not something to be given up lightly. To be sure, the definition of $\mathcal{C}r$ in terms of a probability distribution $\mathcal{D}$ that I gave in Sect. 3 in no way required any special assumptions about the structure of $\mathcal{B}$; so it's clear that we *can* construct a recognisably 'probabilistic' model of partial belief without assuming *Booleanism*. But then we can raise a version of the point made at the end of Sect. 2: if we let Ω satisfy *Really Unrestricted Comprehension*, and simply define $\mathcal{B}$ as $\{\|S\| : S \in \mathcal{L}\}$, then while it's true that $\mathcal{D}$ will let us encode any arbitrary assignment of values into $\mathcal{C}r$, it's hard to see why we should *want* to use a probability distribution in the first place. $\mathcal{D}$ itself doesn't directly represent anything about α's doxastic state—no S will be true at just one world ω, so $\mathcal{D}(\omega)$ cannot be interpreted as a degree of belief towards the singleton proposition

{ω}. What we really have is just a complicated way of listing out α's degree of belief states, with the probabilistic aspects adding nothing to efficiency or illumination.

But that isn't the only worry in the vicinity. A more important concern, I think, arises from the fact that *Booleanism* frequently comes up as a basic assumption in various representation theorems, where the requirement that $\mathcal{B}$ has some minimally rich algebraic structure is a prerequisite for our being able to assign numerical values to the contents of $\mathcal{B}$ in a meaningful and systematic way. For example, the assumption plays a role throughout Jeffrey's (1990) representation theorem for expected utility theory—where, if we were to assume that the space of thinkable propositions $\mathcal{B}$ was such that none of its members is a subset of any other members, almost all of his axioms would be either meaningless or trivial. *Booleanism* is a standard assumption for theories of decision making and uncertainty, with almost all axiomatic decision theories being built around it. Or consider the common approach to characterising numerical degrees of belief defined in terms of qualitative belief orderings over propositions, based on the work of de Finetti (1931) and Scott (1964). Representation theorems which take us from qualitative belief orderings to probabilities are importantly dependent on $\mathcal{B}$ having a rich algebraic structure. Without something like the axiom of *qualitative additivity*—that if P_1 and P_2 both have null intersection with P_3, then one holds P_1 to be more likely than P_2 if and only if one holds $P_1 \cup P_3$ to be more likely than $P_2 \cup P_3$—the qualitative belief ordering would lack a sufficiently rich structure to support anything more than a simple (and representationally inadequate) ordinal scale.[14]

With respect to the fourth strategy, we could perhaps keep the probabilistic model as it is (more or less), but make changes to how we interpret $\mathcal{C}r$.[15] For instance, instead of saying that $\mathcal{C}r(P) = x$ if and only if α has degree of belief x towards some object of belief represented by P, we might instead say that $\mathcal{C}r$ represents α's degrees of belief only where the propositions in question are expressible. But what then of the values that $\mathcal{C}r$ assigns to inexpressible propositions? One thought would be to say that while $\mathcal{C}r$ represents α's degrees of belief when P is expressible, it represents some other propositional attitude ϕ when P is inexpressible. For instance, one might think that if P is expressible, then $\mathcal{C}r(P^{\mathrm{C}})$ represents α's *degree of rejection* towards P, which plausibly is $1 - \mathcal{C}r(P)$. However, this kind of 'rejectionist' proposal will only work if the complement of every inexpressible proposition is expressible, which is not in general the case. In particular, the domain of $\mathcal{C}r$ has to be closed under intersections and unions, and the complement of the (inexpressible) intersection or union of two expressible propositions will often be itself inexpressible.

Of course, there may exist some other broadly 'doxastic' attitude ϕ that I've not considered, which takes inexpressible propositions as its objects—but what reason

[14] To be sure, there are non-Boolean 'probability' theories—for example, quantum probabilities are constructed around involutive algebras which need not satisfy *Booleanism*. I suspect that similar problems as those raised in Sect. 4 will also arise in most circumstances where $\mathcal{B}$ is taken to satisfy a number of basic algebraic closure conditions, but I have not argued for this.

[15] Note that the interpretation of $\mathcal{C}r$ will still have to be recognisably *doxastic*, otherwise we're no longer dealing with a model of α's doxastic states. I have nothing to say about non-doxastic interpretations of $\mathcal{C}r$.

do we have for positing the existence of this ϕ, beyond the desire to preserve some modelling assumptions?

6 The expressibility hypothesis (again)

Finally, one may want to go after the assumption that there exists an $\mathcal{L}$ of the kind described in Sect. 1, in which everything that α believes or partially believes is expressible. If this is false, then the presence of inexpressible propositions in the domain of Cr is perhaps even to be expected, not shunned. Maybe we have just discovered that sometimes our partial beliefs towards expressible propositions comes hand-in-hand with partial beliefs towards inexpressible propositions; the latter are perfectly legitimate objects of thought, but not all such objects are expressible.

First things first, it should be noted that there are accounts of what worlds are which cannot plausibly avoid a version of my argument by denying the expressibility hypothesis. For example, Nolan (1997) favours an approach where (in his terminology) 'propositions'—the meanings of sentences and the objects of thought—are taken to be the fundamental entities from which worlds are constructed. On this picture, possible worlds are maximal consistent sets of propositions *à la* Adams (1974), while impossible worlds are those sets of propositions which are inconsistent and/or non-maximal. Adopting this view, we could let $\mathcal{L}$ simply be the class of all propositions *qua* objects of thought, trivialising the question as to whether $\mathcal{L}$ is 'expressively rich enough' to capture every belief that α might have. We can then easily see that once something like *Unrestricted Comprehension* holds, there will be sets of worlds with no proposition in common amongst their members. These sets of worlds will not only be linguistically inexpressible, but quite literally unthinkable.[16]

Furthermore, I have already noted Jago's work on the expressiveness of Lagadonian languages in Sect. 1, which undergirds his linguistic ersatz account of impossible worlds as arbitrary sets of sentences taken from a pre-specified 'world-making' language $\mathcal{L}$. And note the central importance of the expressibility hypothesis to the account, according to which a set-of-worlds proposition P represents some content C just in case, for every world ω in P, there is a sentence S in ω which expresses that C. In general, this brand of linguistic ersatzer argues for the representational adequacy of their propositions *qua* sets of 'worlds' by arguing first that the basic world-making language is up to the task of distinguishing between all possible contents of belief, from which it quickly follows that sets of sets of these sentences can distinguish between different belief contents—for the simple reason that there is a one-to-one correspondence between the set of sentences S of a language, $\mathcal{L}$, and the set of $P \subseteq \mathcal{L}$ such that $S \in P$. The expressiveness of the ersatz sets-of-worlds model is directly grounded in

[16] This point is not unknown to Nolan, who notes in his (1997, p. 563) that there will be sets of worlds on his account which correspond to no proposition *qua* object of thought. In personal correspondence, Nolan has also pointed out that any set of worlds containing only possible worlds will be inexpressible if Ω satisfies *Unrestricted Comprehension*. For any set of possible worlds $\{\omega_1, \omega_2, \ldots\}$ there will be an impossible world ω_i such that (a) everything true at all of the worlds in $\{\omega_1, \omega_2, \ldots\}$ is true at ω_i, and (b) some impossibility $\perp$ is also true at ω_i. Since $\perp$ isn't true at any possible world, there is nothing that's true at all *and only* the worlds in $\{\omega_1, \omega_2, \ldots\}$.

the expressiveness of the language it's built upon, with propositional representation achieved directly through the meanings of the sentences shared by the worlds within the propositions.

To be sure, one can imagine an ersatzer who begins with a langauge $\mathcal{L}$ which is expressively inadequate, and claims that those beliefs which cannot be represented by any sentence of $\mathcal{L}$ are nevertheless represented by those subsets which have no sentences in common. But how is this representation achieved? Certainly, not in the standard way. Indeed, what reason could we have for thinking that sets of sets of world-making sentences which have nothing in common will do a reasonable job of representing the purportedly 'inexpressible' beliefs? What content would the inexpressible proposition $\{\emptyset\}$ represent, and does it represent anything different than the distinct inexpressible proposition $\emptyset$? And what does $\{\{S_1, S_2\}, \{S_3\}\}$ represent? That either $S_1 \wedge S_2$, or S_3? We already have a content for that:

$$\{\omega \subseteq \Omega : (S_1 \wedge S_2) \vee S_3 \in \omega\}$$

It's hard to imagine any sort of systematic story about how ersatz propositions with nothing in common amongst their members could nevertheless serve to represent a genuine content. And absent such a story, we're stuck with the standard approach, which presupposes the expressive adequacy of the world-making language $\mathcal{L}$.

But I don't want my argument to rest upon specific approaches to characterising worlds. So, to conclude the discussion, I will proceed as follows. First, I'll make a few general points in favour of the expressibility hypothesis. I don't take any of these to be conclusive; much like the present state of the literature on the expressibility of thought, there is plenty of space for disagreement here. It is enough to show, however, that denying the expressibility hypothesis is no trivial matter. Secondly, and much more importantly, I'll end by saying why I don't think that denying the expressibility hypothesis is the right way to respond to the argument.

Let me start then by noting that although there are surprisingly few philosophical discussions regarding whether every possible object of thought is linguistically expressible, to the extent that the question *has* been discussed the usual presumptive answer has been affirmative; e.g., Searle (1969, pp. 19ff), Katz (1978; 1981), Schiffer (2003, p. 71), Priest (2006, p. 54), and Hofweber (2006). Michael Dummett goes so far as to state a priori that:

> Thoughts differ in all else that is said to be among the contents of the mind in being wholly communicable: it is of the essence of thought that I can convey to you the very thought I have […] It is of the essence of thought, not merely to be communicable, but to be communicable, without residue, by means of language. (1978, p. 142)

Most of these discussions focus on natural languages, which makes it a little hard to apply them to the non-natural language $\mathcal{L}$. Of particular note is that natural languages will contain a variety of context-dependent expressions which serve to expand their expressiveness, whereas I've stipulated that the sentences of $\mathcal{L}$ have their meanings independent of context. Since I've made very few substantive assumptions about $\mathcal{L}$, it's hard to see why there would be any particular problems for applying lessons drawn

from natural languages to an language $\mathcal{L}$ besides those which arise from context-sensitivity. Certainly, the fact that the interpretation of $\mathcal{L}$'s sentences are unambiguous and precise shouldn't give us any reason to think that it's *less* likely we'll find the right sentences in $\mathcal{L}$.

We *could* re-run the argument without supposing that $\mathcal{L}$ contains only context-insensitive expressions. We would then need to speak not of expressibility and inexpressibility *simpliciter*, but rather expressibility relative to a context. But, if it's not already plausible that every object of belief is expressible in a context-*in*sensitive language, then it's not clear why every content of belief should be expressible in a context-sensitive language in a specific context. A better option, if we thought that every belief were expressible in some natural language $\mathcal{L}_n$, would be to take $\mathcal{L}_n$ as the basis for the construction of $\mathcal{L}$, which proceeds by systematically eliminating the context-sensitivity of $\mathcal{L}_n$ while preserving overall expressibility. The received view is that such an elimination is entirely possible—and indeed, *easy*. As Stalnaker puts it, it seems at first pass "easy to eliminate context-dependence [since for] any proposition expressed in context c by sentence S, we may simply stipulate that some other sentence S′ shall express, in all contexts, that same proposition" (Stalnaker 1984, pp. 151–152).[17] *If* this kind of elimination strategy is viable, then we have every reason to think that whatever we can say in, e.g., English, we can say in a spruced up and context-independent version of English.

But all this depends on a more general assumption that our beliefs ought to be linguistically expressible *somehow or other*, which the reader may very well doubt. Nevertheless, the existence of something much like $\mathcal{L}$ is strongly suggested by a wide variety of positions in philosophy. The assumption plays a role in important attempts to explain mental representation. If one accepts the arguments for the existence of a Language of Thought as the psychological basis for our capacity to have propositional attitudes, then the existence of a language like $\mathcal{L}$ seems hard to deny. According to this popular view, thinking in general is a computational process sensitive only to the (context-independent) syntax of strings of symbols in a compositional Language of Thought, and one has a belief with content P only in the event that they are appropriately related to a sentence in this language which means that P. The existence of a language rich enough to express each of our beliefs is also presupposed a number of models of mental content. For instance, and besides the Lagadonian approaches already mentioned, Chalmers models the contents of thoughts—including our partial beliefs—as sets of *scenarios*, with each scenario being an 'epistemically complete' description of way the world might be for all we know a priori in an idealised language consisting of vocabulary for describing the microphysical and phenomenal characteristics of the world (see his 2011, 2012). That is, each scenario is a (potentially infinitary) conjunction of sentences in an ideal language, with each scenario being inconsistent with every other scenario. To express any set of scenarios in this language, a (potentially infinitary) disjunction of scenarios will suffice.

With all that said, the recent literature has seen some purported counterexamples to my assumption about the expressibility of belief. Shaw (2013) develops a variation

[17] Of course, not everyone agrees with the received view. See Recanati (1994) and Carston (2002, p. 30ff) for a more conservative perspective on whether this kind of elimination strategy is clearly viable.

on the Berry paradox to argue for the existence of a kind of inexpressible thought content—an instance of a case which he says "happens on extremely rare occasions due to a particular kind of linguistic technicality" (p. 70). Hellie (2004) has also argued that there may be truths about phenomenal experience which we can appreciate but cannot express linguistically. And if one thinks that there is a one-to-one correspondence between ways the world might be and possible belief contents, then there are also classic expressive inadequacy arguments involving qualitatively indiscernible individuals and alien properties, to the effect that no language can describe every possibility (e.g., Lewis 1986, p. 157ff; Bricker 1987). I will not discuss any of these points in detail. Perhaps each gives rise to a genuine problem for the expressibility hypothesis. But acquiescing on this point hardly seems to help with the problem currently at hand. The inexpressibility of most of $\mathcal{C}r$'s domain cannot be explained by an occasional linguistic technicality. And moreover, the inexpressible propositions that we have been describing are not plausibly *about* some ineffable aspect of our phenomenal experience, alien properties, or qualitatively indiscernible individuals.

If $\boldsymbol{\mathcal{L}}$ lacks the expressive power to represent our thoughts about such things—so be it. Let $\boldsymbol{\mathcal{L}}$ represent a language capable of expressing only those more mundane beliefs which *are* expressible, like the belief that *roses are red*. (If need be, let $\boldsymbol{\mathcal{L}}$ be the set of declarative sentences of English, and fix a context.) What kind of content could the set of worlds where 'Roses are red' is *not* true represent, if not that *roses are not red*? Clearly, it has something to do with roses and redness—but what? We can't express it, sure, but it doesn't even seem like there's anything content-like in the vicinity for us to believe. At best, the inexpressible propositions we've been talking about look like an artefact of the model, not some newly discovered kind of content towards which most of our beliefs are directed.

This is, of course, a version of the argument above against the hypothetical linguistic ersatzer who denies the expressibility hypothesis. The point here is general, and constitutes the central reason why going after the expressibility hypothesis looks like the wrong strategy. An *adequate* response to the argument of Sect. 4 can't be to just point out that there may be *some* possible things that α *could* believe which are not expressible. The odd inexpressible object of thought here and there isn't an immediate cause for concern: the underlying problem survives mere counterexamples to the existence of $\boldsymbol{\mathcal{L}}$. Unless we make serious changes to the basic probabilistic model of our beliefs, then so long as *Booleanism* and (*Really*) *Unrestricted Comprehension* are true, if you have a degree of belief x towards $\|\mathrm{S}\|$ you will have a degree of belief $(1-x)$ towards the mysteriously inexpressible proposition $\|\mathrm{S}\|^{\mathrm{C}}$; and if you have degrees of belief x and y towards $\|\mathrm{S}_1\|$ and $\|\mathrm{S}_2\|$ then you'll have some degree of belief $z \leq x, y$ towards the inexpressible $\|\mathrm{S}_1\| \cap \|\mathrm{S}_2\|$ and $((x+y)-z)$ towards $\|\mathrm{S}_1\| \cup \|\mathrm{S}_2\|$. Inexpressibility on this model is not some esoteric phenomenon resting on a technicality, nor does it seem to be limited to a specific kind of topic (e.g., phenomenology, alien properties, and indiscernible individuals) about which we *might* have beliefs.

For similar reasons, I am not moved by simple cardinality arguments aimed at showing that we must accept the existence of inexpressible propositions, regardless of whether we adopt impossible worlds into our ontology or not. Some vigorously intuit that for any subset $\boldsymbol{\mathcal{S}}$ of any language $\boldsymbol{\mathcal{L}}$, α might (partially) believe that all and only the sentences of $\boldsymbol{\mathcal{S}}$ are true. If $\boldsymbol{\mathcal{L}}$ is set-sized, then the cardinality of the $\wp(\boldsymbol{\mathcal{L}})$ is

strictly greater than that of $\mathcal{L}$. It follows that $\mathcal{L}$ cannot contain a unique sentence S for each subset $\mathcal{S} \subseteq \mathcal{L}$ to the effect of 'All and only the elements of $\mathcal{S}$ are true'. Thus, either the content in question is not expressible at all, or it cannot be expressed in $\mathcal{L}$—either way, $\mathcal{L}$ is not up to the task of expressing everything that α might believe. But even if the intuition underlying this argument is correct—and it is by no means obvious that it is—the conclusion is merely that we must accept that we *might* have some inexpressible (partial) beliefs. What the argument doesn't do is give us any reason to think that the algebra of propositions $\mathcal{B}$ that constitutes what α actually has partial beliefs towards is filled to the brim with inexpressible propositions. Indeed, it's perfectly consistent with the argument's conclusion that $\mathcal{B}$ contains no inexpressible propositions at all!

We get to keep the model only if we're happy with the implication that thinkers systematically have at least as many partial beliefs towards inexpressible propositions as they do towards expressible propositions. And that is a hard pill to swallow. If we're to be expected to swallow it, we'll need good reasons to think that (a) these inexpressible propositions exist, (b) that they have such-and-such systematic relations to the expressible propositions, and (c) that they can and indeed always are believed. And those reasons can't be just that these are consequence of a model which includes possible and impossible worlds.

The probabilistic analogues of the problems of logical omniscience require some response. The solution we end up with *may* involve the introduction of impossible worlds, but this looks to be a viable solution only if we drop the very standard—and very important—assumption of *Booleanism*, or if we embrace the inexpressibility of most of our thoughts. Neither option seems particularly appealing, and we may well do better to look for a solution without the impossible.

Acknowledgements Special thanks to to Daniel Nolan and Robbie Williams for helpful discussions on this paper and closely related topics. Further thanks are due to Jessica Isserow for comments on numerous drafts, Thomas Brouwer, Paolo Santorio, the Leeds NatRep and CMM seminar groups, and two anonymous referees for *Synthese*. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 312938. In addition, this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No 703959.

References

Adams, R. M. (1974). Theories of actuality. *Nous*, *8*, 211–231.

Berto, F. (2010). Impossible worlds and propositions: Against the parity thesis. *The Philosophical Quarterly*, *40*, 471–86.

Bjerring, J. C. (2013). Impossible worlds and logical omniscience: An impossibility result. *Synthese*, *190*, 2505–24.

Bjerring, J. C. (2014). Problems in epistemic space. *Journal of Philosophical Logic*, *43*, 153–170.

Bjerring, J. C., & Schwarz, W. (2017). Granularity problems. *The Philosophical Quarterly*, *67*(266), 22–37.

Bricker, P. (1987). Reducing possible worlds to language. *Philosophical Studies*, *52*(3), 331–355.

Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell.
Chalmers, D. (2011). The nature of epistemic space. In A. Egan & B. Weatherson (Eds.), *Epistemic modality* (pp. 60–107). Oxford: Oxford University Press.
Chalmers, D. (2012). *Constructing the world*. Oxford: Oxford University Press.
Choquet, G. (1954). Theory of capacities. *Annales de l'institut Fourier*, *5*, 131–295.
Cozic, M. (2006). Contributions to economic analysis. In R. Topol & B. Walliser (Eds.), *Impossible states at work: Logical omniscience and rational choice* (pp. 47–68). Amsterdam: Elsevier.
Creswell, M. J. (1973). *Logics and languages*. London: Methuen.
Davies, M. (1981). *Meaning, quantification, necessity: Themes in philosophical logic*. London: Routledge & Kegan Paul.
de Finetti, B. (1931). Sul significato soggettivo della probabilita. *Fundamenta Mathematicae*, *17*(1), 298–329.
Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society Series B (Methodological)*, *30*, 205–247.
Dubois, D., & Prade, H. (1988). *Possibility theory. An approach to computerized processing of uncertainty*. New York: Plenum.
Dummett, M. (1978). *Truth and other enigmas*. Cambridge: Harvard University Press.
Easwaran, K. (2014). Regularity and hyperreal credences. *Philosophical Review*, *123*(1), 1–41.
Halpern, J. Y., & Pucella, R. (2011). Dealing with logical omniscience: Expressiveness and pragmatics. *Artificial Intelligence*, *175*(1), 220–235.
Hellie, B. (2004). There's something about Mary. In Y. Nagasawa (Ed.), *Inexpressible truths and the allure of the knowledge argument* (p. 333-64). Cambridge: MIT press.
Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. Ithaca: Cornell University Press.
Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic*, *4*, 475–84.
Hofweber, T. (2006). Inexpressible properties and propositions. In D. Zimmerman (Ed.), *Oxford studies in metaphysics*. Oxford: Oxford University Press.
Jago, M. (2009). Logical information and epistemic space. *Synthese*, *167*, 327–341.
Jago, M. (2012). Constructing worlds. *Synthese*, *189*, 59–74.
Jago, M. (2013). The logica yearbook 2012. In V. Puncochar & P. Svarny (Eds.), *Are impossible worlds trivial?* (pp. 35–50). London: College Publications.
Jago, M. (2014a). *The impossible: An essay on hyperintensionality*. Oxford: Oxford University Press.
Jago, M. (2014b). The problem of rational knowledge. *Erkenntnis*, *79*, 1151–1168.
Jago, M. (2015a). Hyperintensional propositions. *Synthese*, *192*(3), 585–601.
Jago, M. (2015b). Impossible worlds. *Nous*, *49*(4), 713–728.
Jeffrey, R. C. (1990). *The logic of decision*. Chicago: University of Chicago Press.
Kaplan, D. (1995). A problem in possible worlds semantics. In W. Sinnott-Armstrong, et al. (Eds.), *Modality, morality and belief: Essays in honor of Ruth Barcan Marcus* (pp. 41–52). Cambridge: Cambridge University Press.
Katz, J. (1978). Effability and translation. In F. Guenthner & M. Guenthner-Reutter (Eds.), *Meaning and translation* (pp. 157–189). New York: NYU Press.
Katz, J. (1981). *Language and other abstract objects*. Oxford: Basil Blackwell.
Kyburg, H. E. (1992). Getting fancy with probability. *Synthese*, *90*, 189–203.
Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy*, *71*(13), 391–418.
Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, *88*(4), 513–543.
Lewis, D. (1982). Logic for equivocators. *Nous*, *16*(3), 431–441.
Lewis, D. (1986). *On the plurality of worlds*. Cambridge: Cambridge University Press.
Lewis, D. (2004). The law of non-contradiction: New philosophical essays. In G. Priest (Ed.), *Letters to beall and priest* (pp. 176–177). Oxford: Clarendon Press.
Lipman, B. L. (1997). Epistemic logic and the theory of games and decisions. In M. Bacharach (Ed.), *Logics for nonomniscient agents: An axiomatic approach* (pp. 193–216). Berlin: Springer.
Lipman, B. L. (1999). Decision theory without logical omniscience: Toward an axiomatic framework for bounded rationality. *The Review of Economic Studies*, *66*(2), 339–361.
Nolan, D. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic*, *38*, 535–72.
Nolan, D. (2013). Impossible worlds. *Philosophy Compass*, *8*(4), 360–372.
Perry, J. (1979). The problem of the essential indexical. *Nous*, *13*(1), 3–21.
Priest, G. (2006). *In contradiction: A study of the transconsistent*. Oxford: Oxford University Press.

Rantala, V. (1982). Impossible worlds semantics and logical omniscience. *Acta Philosophica Fennica*, *35*, 106–15.

Recanati, F. (1994). Foundations of speech act theory: Philosophical and linguistic perspectives. In S. Tsohatzidis (Ed.), *Contextualism and anti-contextualism in the philosophy of language* (pp. 156–166). London: Routledge.

Schiffer, S. (2003). *The things we mean*. Oxford: Oxford University Press.

Scott, D. (1964). Measurement structures and linear inequalities. *Journal of Mathematical Psychology*, *1*(2), 233–247.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton: Princeton University Press.

Shaw, J. R. (2013). Truth, paradox, and ineffible propositions. *Philosophy and Phenomenological Research*, *86*(1), 64–104.

Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical application*. Oxford: Oxford University Press.

Stalnaker, R. (1996). Impossibilities. *Philosophical Topics*, *24*, 193–204.

Stalnaker, R. C. (1984). *Inquiry*. London: The MIT Press.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.