

BAYESIANISM FOR NON-IDEAL AGENTS

Mattias Skipper · Jens Christian Bjerring

Penultimate draft, forthcoming in *Erkenntnis*

Abstract: Orthodox Bayesianism is a highly idealized theory of how we ought to live our epistemic lives. One of the most widely discussed idealizations is that of *logical omniscience*: the assumption that an agent's degrees of belief must be probabilistically coherent to be rational. It is widely agreed that this assumption is problematic if we want to reason about bounded rationality, logical learning, or other aspects of non-ideal epistemic agency. Yet, we still lack a satisfying way to avoid logical omniscience within a Bayesian framework. Some proposals merely replace logical omniscience with a different logical idealization; others sacrifice all traits of logical competence on the altar of logical non-omniscience. We think a better strategy is available: by enriching the Bayesian framework with tools that allow us to capture what agents can and cannot *infer* given their limited cognitive resources, we can avoid logical omniscience while retaining the idea that rational degrees of belief are in an important way constrained by the laws of probability. In this paper, we offer a formal implementation of this strategy, show how the resulting framework solves the problem of logical omniscience, and compare it to orthodox Bayesianism as we know it.

Keywords: Bayesianism · Logical omniscience · Bounded rationality · Logical learning

1 INTRODUCTION

Keep your degrees of belief probabilistically coherent at all times, and update them by conditionalization as new information comes in. So the orthodox

Bayesian story goes. The story is picture-perfect. It draws the contours of an ideal epistemic life. A life where all tautologies are believed with certainty, and where an agent's confidence never drops across entailment; one where logical perfection is just a matter of good epistemic housekeeping.

But for ordinary humans like you and me, life isn't perfect. For us, and other imperfect beings like us, probabilistic coherence remains an unattainable ideal. As hard as we may try, we will never unveil all tautologies or recognize all entailment relations. Inevitably, we will fall short of logical perfection. Not by choice, to be sure. Most of us would take great pride in being able to prove the Riemann hypothesis or decide whether $P = NP$. We simply can't. After all, we are only human; logical imperfection is part of our condition as cognitively limited beings.

Does this elementary fact about our epistemic predicament spell doom for orthodox Bayesianism? Not straightforwardly. Epistemic ideals are interesting in their own right, and deserve our attention no less than do moral or political ideals.¹ But if we want to reason about bounded rationality, logical learning, or other aspects of non-ideal epistemic agency, the message seems clear: 'probabilistic coherence must go!' The message has not gone unheard. Many formally inclined epistemologists have viewed the commitment to probabilistic coherence as one of the most serious problems for orthodox Bayesianism—one that has come to be known as *the problem of logical omniscience*.²

Don't get us wrong: Ordinary humans need not be careless or epistemically irresponsible. We often engage successfully in logical reasoning, not only when we sweat over a logic exam, but also when we deliberate about which decisions to make in day-to-day life. Suppose you ponder whether to ask your boss for a pay raise today. You know that your boss is in a generous mood only if she took the bike to work. Yet, you don't see her bike in the bike rack. Thus, you decide to defer your request for another day. This sort

¹Although it is a matter of contention whether probabilistic coherence is indeed a rational ideal. For more on this point, see Christensen (2007), Smithies (2015), and Titelbaum (2015).

²Cf. Easwaran (2011) and Talbott (2016). See also Fagin et al. (1995) for a discussion of logical omniscience as it arises in epistemic and doxastic logic.

of basic ability to engage in logical reasoning should, we submit, feature in a solution to the problem of logical omniscience. It is not enough to model agents who fall short of *logical omniscience*; we also need to capture the sense in which such agents nevertheless display some level of *logical competence*.

Over the past half century, several attempts have been made at solving the problem of logical omniscience. But none has gained widespread acceptance. Some proposals merely replace logical omniscience with a different logical idealization; others sacrifice all traits of logical competence on the altar of logical non-omniscience (§2). We think a better strategy is available: by enriching the Bayesian framework with tools that allow us to capture what agents can and cannot *infer* given their limited cognitive resources, we can avoid logical omniscience while retaining the idea that rational degrees of belief are in an important way constrained by the laws of probability (§3). As we will see, this ‘dynamic’ approach to the problem is not without substantive commitments; it forces us to reconsider some fundamental principles of orthodox Bayesianism (§4). Yet it offers what we take to be a principled and intuitive solution to a long-standing problem in Bayesian epistemology.

2 EXISTING APPROACHES TO LOGICAL OMNISCIENCE

Let us begin by taking a closer look at the two main sources of logical omniscience within orthodox Bayesianism:³

Classical Preservation: For any two propositions, A and B , if A logically entails B , then $Cr(A) \leq Cr(B)$.

Classical Normality: For any tautology T , $Cr(T) = 1$.

According to Classical Preservation, an agent’s credences never drop across entailment. For example, someone who is 80% confident that “it’s raining” will be at least 80% confident that “it’s raining or snowing.” More generally, whenever a proposition, A , entails another proposition, B , an agent will be at least as confident of B as of A , however difficult it may be to see that

³Both principles follow from the Kolmogorov axioms; see Earman (1992) and Titelbaum (forthcoming) for relevant background.

B follows from A . According to Classical Normality, agents are certain of all tautologies. For example, they will be certain that “it’s either raining or not.” More generally, for any tautology, T , an agent will be certain of T , however difficult it may be to see that T is tautological.

Anyone who seeks to avoid logical omniscience must find a way to drop (or relax) Classical Preservation and Classical Normality. But how? In the rest of this section, we survey and criticize three prominent answers.

2.1 GARBER ON ‘LOCAL BAYESIANISM’

In his “Old Evidence and Logical Omniscience in Bayesian Confirmation Theory” (1983), Daniel Garber attempts to solve the problem of logical omniscience by making room for *logical learning* in the Bayesian framework.⁴ Logical learning is not possible within orthodox Bayesianism, since Classical Preservation and Classical Normality tell us that agents already know everything there is to know about logic; and you cannot learn what you already know. Thus, while orthodox Bayesianism can accommodate *empirical* learning (in terms of conditionalization on incoming evidence), it rules out the possibility of *logical* learning.

Garber seeks to rectify this situation by developing what he calls a *local* version of the Bayesian framework. The idea is to think of an agent’s credence function as being defined over a *problem-relative* language, which consists of (the truth-functional closure of) those sentences that the agent is concerned with when being engaged in a given “inferential problem” (Garber 1983, p. 111). For example, the local language of a 17th-century physicist might contain vocabulary from Newtonian mechanics, but will presumably not contain vocabulary from quantum mechanics.

In addition to these problem-relative sentences, Garber also includes sentences of the form ‘ $A \vdash B$ ’ in the agent’s local language. Informally, these sentences are to be interpreted as ‘ A entails B ’. But they are treated as atomic

⁴In a broader perspective, Garber’s approach to logical omniscience forms part of an attempt to solve the so-called ‘problem of old evidence’ introduced by Glymour (1980). Related approaches to this problem can be found in Gaifman (2004), Good (1968), and Jeffrey (1983). See also Fitelson and Hartmann (2015) and Sprenger (2015) for more recent developments in this direction.

sentences in the formalism, and may therefore be assigned non-extreme credences.

The idea, then, is to make room for logical learning by allowing agents to *update* on sentences of the form ‘ $A \vdash B$ ’ that they are not already certain of. For example, a physicist might at one point have a non-extreme credence that quantum mechanics entails such-and-such evidence, but might later discover that this entailment relation in fact obtains.

For present purposes, let us say that A *locally entails* B iff B can be derived from A in the agent’s local language, and let us say that T is a *local tautology* iff T is a tautology in that language. We can then think of Garber’s proposal as replacing Classical Preservation and Classical Normality with the following two principles:

Local Preservation: For any two sentences, A and B , if A locally entails B , then $Cr(A) \leq Cr(B)$.

Local Normality: For any local tautology T , $Cr(T) = 1$.

Assuming that all *local* entailments are *logical* entailments, but not *vice versa*, Local Preservation is strictly weaker than Classical Preservation. Likewise, assuming that all *local* tautologies are *logical* tautologies, but not *vice versa*, Local Normality is strictly weaker than Classical Normality. Thus, Garber’s proposal offers a way around the classical assumption of logical omniscience.

But although Local Bayesianism does not give rise to full-blown logical omniscience as we know it from orthodox Bayesianism, it still carries a commitment to a problematic kind of logical idealization. Ellery Eells puts the point as follows: “[T]here will always be extremely complex logically true sentences of the local language [for which] it will be inappropriate to insist on probability 1” (Eells 1985, p. 241). The observation here is that the local language is still closed under truth-functional operations, which means that even a very sparse characterization of the local language will give rise to very complicated logical relations that go far beyond the cognitive reach of ordinary agents. As such, Local Bayesianism merely replaces the classical assumption of logical omniscience with a different logical idealization.

It is also worth pointing out that even if we could define a local language

without any overly complex logical relations, there are general grounds for doubting that the resulting model would adequately capture the sense in which ordinary humans fall short of logical omniscience. The central reason why such agents fail to be logically omniscient, we take it, is not that they operate with a restricted language, but rather that they have limited cognitive resources available to reason logically in that language. Thus, it seems to us that the strategy of defining an agent’s credence function over a restricted language does not ultimately get at the heart of the matter.

2.2 HACKING ON ‘PERSONAL PROBABILITY’

In his “Slightly More Realistic Personal Probability” (1967), Ian Hacking attempts to solve the problem of logical omniscience by developing a ‘personalized’ version of the Bayesian framework. The idea is to replace the classical entailment relation with a ‘personal’ entailment relation, which is strictly weaker (that is, all *personal* entailments are *logical* entailments, but not *vice versa*). Personal entailment is defined in terms of a notion of *personal possibility*: just as A *logically* entails B iff $A \wedge \neg B$ is not *logically* possible, A *personally* entails B (for a given agent) iff $A \wedge \neg B$ is not *personally* possible (for that agent).

When is a proposition personally possible for a given agent? Intuitively, says Hacking, when the agent does not know that the proposition is false (Hacking 1967, p. 318). In other words, A is personally possible for an agent iff A cannot be ruled out by the agent given his or her empirical information and deductive abilities. For example, suppose that B is a highly complex logical consequence of A , which lies far beyond the cognitive reach of any human. Given this, the notion of personal possibility will come apart from logical possibility: although $A \wedge \neg B$ is not logically possible, it will be personally possible for agents who are unable to recognize that A and $\neg B$ are in fact logically incompatible. Accordingly, personal entailment comes apart from logical entailment in this case: while A logically entails B , it does not personally entail B .

For present purposes, let us supply the notion of personal entailment with

a notion of a *personal tautology*: T is a personal tautology (for an agent) iff $\neg T$ is not personally possible (for that agent). We can then think of Hacking's proposal as replacing Classical Preservation and Classical Normality with the following two principles:

Personal Preservation: For any two sentences, A and B , if A personally entails B , then $Cr(A) \leq Cr(B)$.

Personal Normality: For any personal tautology T , $Cr(T) = 1$.

Assuming that all *personal* entailments are *logical* entailments, but not *vice versa*, Personal Preservation is strictly weaker than Classical Preservation. Likewise, given that all *personal* tautologies are *logical* tautologies, Personal Normality is strictly weaker than Classical Normality. Thus, like Garber's proposal, Hacking's proposal offers a way around the classical assumption of logical omniscience.

Just *how* weak are Hacking's principles? The answer is: extremely weak! The reason for this is that Hacking treats personal entailment as a "degenerate concept with no closure conditions" since such closure conditions (say, closure under modus ponens) tend to "lead disastrously near the divine sense of knowledge" (Hacking 1967, p. 319). That is, to eliminate all traits of logical omniscience, Hacking assumes that *no* logical entailment, however trivial, need count as a personal entailment. The result is a model that shows no sign of logical competence.

In response to this sort of worry, Hacking submits that the notion of personal probability is not, in fact, as 'laissez faire' as one might think. He writes:

[D]oes not the slightly more realistic theory excuse a man from any cogent reasoning whatsoever? No. In the classical [Bayesian] theory, the Dutch Book argument is used to club a man into reasoning. There may be a better club to hand. (Hacking 1967, p. 322)

The thought here is that, even if ordinary humans cannot be faulted for being susceptible to a dutch book made by a logically omniscient Bookie, they *can*

be faulted for being susceptible to a dutch book made by a Bookie who is epistemically *on a par* with them. If so, we can still rely on Dutch book considerations to justify certain non-ideal rationality constraints.

Unfortunately, we don't think that the appeal to non-ideal Bookies provides us with the necessary resources to model agents who are logically competent in the relevant sense. As Ellery Eells rightly points out:

[Hacking's proposal] has the consequence that if the agent is *unaware of an incoherence* in his subjective probabilities, then so must be an appropriate betting opponent. But this means that a person will turn out to be rational in the Bayesian sense as long as the person is *not aware of an incoherence*. (Eells 1985, p. 217)

In other words, as long as there are no general constraints on what can be personally possible for an agent, there is no incoherence so obvious that an agent must be aware of it to be rational. As such, nothing in Hacking's framework reflects the sense in which ordinary humans are logically competent.

Nevertheless, we think that Hacking is halfway right. He is wise not to impose any substantive closure constraints on epistemic states, since these give rise to a problematic kind of logical omniscience. What is missing from his picture is a way of capturing what ordinary humans can and cannot infer given their limited cognitive resources.

2.3 HINTIKKA ON 'IMPOSSIBLE POSSIBLE WORLDS'

In his "Impossible Possible Worlds Vindicated" (1975), Jaakko Hintikka tries to solve the problem of logical omniscience by developing an *impossible-worlds* model of belief, which extends his earlier work on doxastic and epistemic logic.⁵ While Hintikka's own exposition centers around "all-or-nothing" belief rather than credences, we can straightforwardly translate it into Bayesian terms.

To put Hintikka's proposal in its proper context, let us begin by considering his original possible-worlds model for belief:

⁵Hintikka (1962). See also von Wright (1951) for an early precursor to Hintikka.

Belief: An agent believes a proposition, A , iff A is true at all possible worlds that are doxastically possible for the agent.

As Hintikka himself recognized, this model carries a commitment to logical omniscience: it implies that any agent believes all logical consequences of what they believe, including all tautologies. To see this, suppose that you believe A , and let B be any logical consequence of A . Since you believe A , A is true at all possible worlds that are doxastically possible for you. And since A entails B , all possible worlds that verify A also verify B . Thus, B must be true at all doxastically possible worlds for you, which means that you believe B .

To avoid this result, Hintikka suggests that we extend the set of possible worlds with a set of *impossible worlds*: worlds that “look possible and hence must be admissible as epistemic alternatives but which none the less are not logically possible” (Hintikka 1975, p. 477).⁶ The basic idea is that, for limited agents, the space of *doxastic* possibilities is larger than the space of *logical* possibilities. For example, even if you believe each of the Peano Axioms, you might well fail to believe Goldbach’s Conjecture, even if the former in fact entail the latter. In other words, even if the Axioms are true at all doxastically possible worlds for you, the Conjecture need not be.

Motivated by this thought, Hintikka suggests the following modification of the original possible-worlds model of belief:

Belief-impossible: An agent believes a proposition, A , iff A is true at all worlds (whether possible or impossible) that are doxastically possible for the agent.

By quantifying over impossible as well as possible worlds, we can make room for agents who believe the Peano Axioms, but fail to believe Goldbach’s Conjecture. For even if the Axioms are true in all doxastically possible worlds, the Conjecture need not be, as long as impossible worlds are allowed to verify the Axioms without verifying the Conjecture.

When translated into a credence framework, Hintikka’s proposal becomes:

⁶For other early discussions of impossible worlds, see Cresswell (1973) and Rantala (1982). See also Berto and Jago (2018; 2019) for relevant background.

Credence-impossible: An agent’s credence in a proposition, A , is x iff the probabilities of those A -worlds (whether possible or impossible) that are doxastically possible for the agent sum up to x .

Like Belief-impossible, Credence-impossible offers a way around the assumption of logical omniscience. Even if you are certain that the Peano Axioms are true, you need not, according to Credence-impossible, be certain that Goldbach’s Conjecture is true. After all, there might be doxastically possible (but logically impossible) worlds that verify the Axioms, but not the Conjecture.

As stated, Credence-impossible says nothing about the nature of impossible worlds. Hence, we can get different versions of the impossible-worlds model by adopting different underlying conceptions of impossible worlds. Following Berto and Jago (2018), we can distinguish between two general such conceptions—what they call the ‘Australasian stance’ and the ‘American stance’ on impossible worlds. According to the Australasian stance, impossible worlds are allowed to violate the laws of *classical* logic, but must still respect the laws of some *non-classical* logic (say, intuitionistic or paraconsistent logic).⁷ According to the American stance, impossible worlds are not subject to any closure constraints whatsoever, but may be arbitrarily logically ill-behaved.⁸

Both these stances give rise to problems that are structurally very similar to those that led us to reject the proposals by Garber and Hacking. Consider first the Australasian stance. Suppose that all impossible worlds must respect the laws of some non-classical logic, \mathcal{L} . Given this, we can replace Classical Preservation and Classical Normality by the following two principles:

Non-classical Preservation. For any two propositions A and B , if A entails B in \mathcal{L} , then $Cr(A) \leq Cr(B)$.

Non-classical Normality. For any tautology T in \mathcal{L} , $Cr(T) = 1$.

Since \mathcal{L} is assumed to be strictly weaker than classical logic, these principles do not imply that agents are omniscient within *classical* logic. But they *do*

⁷See, e.g., Fagin et al. (1995), Levesque (1984), and Lakemeyer (1987).

⁸See, e.g., Nolan (1997).

imply that agents are omniscient within the chosen *non-classical* logic, \mathcal{L} . For example, if we understand \mathcal{L} as an intuitionistic logic, agents are assumed to be certain of all intuitionistic tautologies and entailment relations. Yet, just as ordinary humans fall short of *classical* omniscience, they fall short of *intuitionistic* omniscience as well. Indeed, even if we grant that a particular non-ideal agent reasons intuitionistically rather than classically, the agent obviously cannot reason *unlimitedly* in that logic. Thus, the Australasian stance faces much the same problem as Garber’s Local Bayesianism: it merely replaces logical omniscience with a different logical idealization.

To add fuel to the fire, we do not think the Australasian stance provides us with the resources to capture the sense in which ordinary humans are logically *competent*. The central reason why we fall short of logical omniscience, it seems, is not that we operate with a non-classical notion of entailment, but rather that we have limited cognitive resources available for logical reasoning. Thus, it seems that the Australasian stance, like Garber’s approach, does not ultimately get at the heart of the problem.

What about the American stance? Suppose that impossible worlds are not subject to any closure constraints whatsoever. That is, impossible worlds are allowed to verify A without verifying B , for any A and B . Given this, we can eliminate all traits of logical omniscience: an agent’s confidence need not be preserved across *any* entailments, however trivial; and the agent need not be certain (or even moderately confident) of *any* tautologies, however obvious.

The problem, of course, is that logical anarchy not only eliminates logical omniscience, but also breeds logical incompetence. More specifically, if our model allows an agent’s credences to be arbitrarily logically ill-behaved, we are left with no way of capturing the logical reasoning abilities of ordinary humans. Thus, the American stance faces much the same problem as Hacking’s approach: it sacrifices all traits of logical competence on the altar of logical omniscience.

In light of these problems, it is very natural to search for a ‘middle way’ between the Australasian stance and the American stance on impossible worlds. In particular, it is natural to think that we should try to close

an agent’s doxastic state under a notion of logical entailment that—unlike classical entailment and, say, intuitionistic entailment—reflects the agent’s limited cognitive resources. At the level of worlds, this amounts to saying (roughly) that impossible worlds should be closed under a notion of *partial* logical consequence: they should verify everything that lies within the agent’s cognitive reach, and nothing more. The hope is to thereby be able to replace Classical Preservation and Classical Normality with the following two principles:

Partial Preservation. For any propositions A and B , if A partially entails B , then $Cr(A) \leq Cr(B)$.

Partial Normality. For any partial tautology T , $Cr(T) = 1$.

As stated, these principles are obviously not fully precise, since we have not given a precise definition of partial entailment. But the idea is clear enough: the principles are supposed to avoid the assumption of logical omniscience while retaining an appropriate level of logical competence.

Unfortunately, the natural idea cannot be made to work. The problem, in a nutshell, is that any notion of *partial* entailment ‘collapses’ into *full* entailment. Here is a way of illustrating the mechanism behind this sort of collapse: Consider a very minimal partial closure constraint, which says that impossible worlds should at least obey those entailment relations that are *trivial* or *obvious* for ordinary humans to recognize. That is, if A trivially entails B , every impossible world that verifies A must verify B as well. What counts as ‘trivial’ or ‘obvious’ is clearly a vague matter. But nothing turns on the vagueness: regardless of how we make the notion of a ‘trivial logical entailment’ precise, it turns out that we cannot close impossible worlds under *trivial* logical consequence without closing them under *full* logical consequence.

Here is why: let w be any world that is not closed under full logical consequence. That is to say, there exists two propositions, A and B , such that the following three conditions are met: (i) A entails B ; (ii) w verifies A ; and (iii) w does not verify B . By (i), we can consider a sequence of propositions A, S_1, S_2, \dots, B corresponding to a step-by-step inference from A to B in

terms of simple logical rules such as conjunction elimination, modus ponens, and the like. By (ii) and (iii), we know that w must violate at least one step in this inference. Yet, each step in the inference is exceedingly simple, and so must count as trivial for ordinary humans, if anything does. It follows, then, that w cannot be closed under trivial logical consequence. Upshot: if w is closed under *trivial* logical consequence, w must be closed under *full* logical consequence. Intuitively, it *collapses* under its own deductive weight.⁹

In sum, there is no stable ‘middle way’ between the Australasian stance and the American stance. Even the most minimal closure constraints on impossible worlds are too strong. What to do about this?

3 A DYNAMIC BAYESIAN FRAMEWORK

The foregoing considerations urge us to choose between two evils: *logical omniscience or logical incompetence*—*which shall it be?* Neither! Or so we submit. The dilemma arises when we try to model logical competence in terms of *closure constraints* on doxastic states. But we need not restrict ourselves to this ‘static’ way of modeling logical competence. A better strategy is available: if we enrich the Bayesian framework with tools that allow us to model how an agent’s doxastic state can *change* as a result of engaging in logical reasoning, we can steer clear of logical omniscience and logical incompetence at the same time. In previous work, we have used this ‘dynamic’ strategy to solve the problem of logical omniscience as it arises within doxastic and epistemic logic (Bjerring & Skipper forthcoming). Here we want to show how the same basic strategy can be extended to a Bayesian context.

Before we get into the details, let us clarify how the positive proposal outlined below should be seen in relation to the collapse result discussed in the previous section. Our claim is not that we can *avoid* the collapse result by moving into a dynamic framework. Indeed, if the foregoing considerations are right, this cannot be done. Rather, our dynamic framework will offer a

⁹For related discussions of similar collapse results that arise in the contexts of epistemic logic, decision theory, and formal semantics, see Bjerring (2013), Bjerring and Skipper (forthcoming), Bjerring and Schwarz (2017), Elga and Rayo (ms.), Jago (2013), and Rasmussen (2015).

way out of the more generic dilemma between logical omniscience and logical incompetence, which is *compatible* with the collapse result. That is the claim, anyway. The rest of the section makes the case.

Our first task is to define a quantitative measure of an agent's cognitive resources, which can be implemented in a Bayesian framework. The exact choice of measure is of little importance for present purposes. Our aim is not to give an empirically accurate representation of how human beings reason logically. Rather, we are looking to provide a general, empirically non-committal way of capturing the elementary fact that some logical inferences are more complex or difficult to perform than others. That is, we need a way of distinguishing the Kurt Gödels of this world from his less resourceful fellow earthlings.

In principle, many different measures could do this job, whether they appeal to time consumption, neural activity, or some third quantity. However, in line with previous work, we will use a simple *step-based* model of bounded logical reasoning.¹⁰ The basic idea is to represent an agent's cognitive resources by a number, n , which corresponds to the number of inference steps that the agent is able to perform. By varying the value of n , we can then generate a whole *spectrum* of agents with different levels of cognitive resources. When $n = 0$, no chain of logical reasoning, however simple, lies within the agent's cognitive reach. Intuitively, the agent has no cognitive resources whatsoever. When n approaches infinity, no chain of logical reasoning, however complex, lies beyond the agent's cognitive reach. Intuitively, the agent has unlimited cognitive resources. For intermediate values of n , *some* but not *all* chains of logical reasoning lie within the agent's cognitive reach. Intuitively, the agent is neither logically omniscient nor logically incompetent. This is the part of the spectrum that we will mainly be interested in.

The step-based model goes along with a broadly *rule-based* picture of logical reasoning. On such a picture, agents reason logically by applying rules from a designated set, R , of available inference rules. For illustrative

¹⁰Bjerring & Skipper (forthcoming). The step-based model was initially inspired by work in *active logic*; see Elgot-Drapkin and Perlis (1990) for background.

purposes, we will think of R as containing familiar inference rules such as conjunction elimination, modus ponens, disjunction introduction, and the like. But on the official story, we do not presuppose any particular specification of R . There are two reasons for this. First, by leaving the specification of R open, our framework will be applicable to different contexts, which call out for different specifications of R . In particular, our framework will not be limited to contexts that call out for a sound and complete proof system of classical logic. Second, our framework is not going to stick its neck out with respect to substantive, empirical questions about human cognition. This is not to say that such questions are entirely orthogonal to the present project. Indeed, we suspect that certain applications of our framework may require an empirically informed reasoning mechanism. But for the purposes of laying out the basic framework, there is no reason to sacrifice generality for empirical accuracy.

Henceforth, then, we will think of logical competence as the ability to perform up to n steps of reasoning using the rules in R . In the formalism, we will implement this idea by defining a ‘non-ideal’ notion of logical entailment, which we call *n-entailment*:

***n-entailment*:** A set of sentences, Γ , *n-entails* a sentence, A , (written ‘ $\Gamma \vdash_R^n A$ ’) iff A can be inferred from Γ within n applications of the inference rules in R .

The role of the \vdash_R^n -relation is to capture what logical entailments lie within an agent’s cognitive reach; and which do *not*. To get a feel for the definition, suppose that R contains just a single rule: conjunction elimination. Given this, A is 1-entailed by $A \wedge B$, and $A \wedge B$ is 1-entailed by $(A \wedge B) \wedge (A \wedge B)$, but A is *not* 1-entailed by $(A \wedge B) \wedge (A \wedge B)$, since it takes *two* applications of conjunction elimination to infer A from $(A \wedge B) \wedge (A \wedge B)$.

Three properties of the \vdash_R^n -relation are worth noting. First, as illustrated by the example above, *n-entailment* is not transitive (in contrast to classical entailment): even if A *n-entails* B , and B *n-entails* C , A need not *n-entail* C . Such transitivity failures are going to play an important role in our solution to the problem of logical omniscience.

Second, the \vdash_R^n -relation is monotonic in n : if A i -entails B , A also j -entails B , for $i \leq j$. The reason is trivial: any inference that can be carried out in i steps or less can also be carried out in j steps or less, for $i \leq j$. The opposite is obviously not the case. So, even if A j -entails B , A need not i -entail B .

Third, n -entailment is equivalent to classical entailment in the special case where n goes to infinity and R is a sound and complete proof system of classical logic. Although not our main target, this effectively means that our framework is general enough to model agents who are logically omniscient in the classical sense.

Our next task is to define the formal language over which we will define our models. So, let L be a probabilistic modal language with atomic sentences p_1, p_2, \dots , negation \neg , conjunction \wedge , weak inequality \leq , a credence function Cr , and a countably infinite set of ‘dynamic’ operators $\langle n \rangle$ and $[n]$ (for $n = 0, 1, 2, \dots$).¹¹ In addition to the atomic sentences, L contains the following sentence types:

$$\neg A \mid A \wedge B \mid Cr(A) \leq x \mid \langle n \rangle A \mid [n] A,$$

where x is a real number in the unity interval $[0, 1]$, and A and B are arbitrary sentences of L . We will help ourselves to other familiar connectives (\vee, \rightarrow, \dots) and (in)equalities ($=, <, \dots$), which can be defined in the usual way from the primitive language.

The dynamic operators have the following intended readings:

$\langle n \rangle A$: after some n -step reasoning process, A is the case.

$[n] A$: after any n -step reasoning process, A is the case.

By combining the dynamic operators with the credence function, we can write things like ‘ $\langle n \rangle (Cr(A) < .7)$ ’ to say that the agent is less than 70 % confident of A after some n -step reasoning process, or ‘ $[n] (Cr(A) = 1)$ ’ to say that the agent is certain of A after any n -step reasoning process.

¹¹Slightly abusing notation, we will use ‘ L ’ both as the name of our object language and as a variable that ranges over all sentences of that language.

To develop a semantics for L , we will combine some familiar tools from probabilistic modal logic with some more recent developments in dynamic epistemic logic.¹²

We begin with the notion of a *subjective probability space*:

Subjective probability space: Let W^P and W^I be finite, non-empty sets of possible and impossible worlds, and let $W = W^P \cup W^I$. A subjective probability space is a pair, (S, Pr) , where $S \subseteq W$ is a non-empty set of worlds, and $Pr : S \mapsto [0, 1]$ is a distribution over S such that $\sum_S Pr(S) = 1$. The set of all subjective probability spaces is denoted by \mathcal{S} .

As usual, we think of a subjective probability space as a representation of an agent's doxastic situation at a given time. The worlds in S are those that are *doxastically possible* for the agent. That is, for each $w \in S$, w might be the actual world for all the agent can tell given his or her cognitive resources and empirical information. The distribution, Pr , encodes information about how probable the agent takes it to be that a given world is actual. For example, if $Pr(w) < Pr(w')$, the agent takes it to be more probable that w' is actual than that w is. Since the agent is certain that *some* world in S is actual, we require that $\sum_S Pr(S) = 1$.

We extend the notion of a subjective probability space to a full probabilistic model as follows:

Probabilistic model: A probabilistic model is a tuple,

$$M = (W^P, W^I, f, V),$$

where $f : W \mapsto \mathcal{S}$ assigns a subjective probability space to each world in W , and $V : W \mapsto 2^L$ assigns a set of sentences in L to each world in W .

The function f tells us what the agent's subjective probability space looks like at different worlds. In general, f will assign different subjective probability spaces to different worlds, since an agent's doxastic situation differs

¹²For background on dynamic epistemic logic, see Ditmarsch et al. (2008) and Baltag and Renne (2016).

from world to world. The function V serves as a ‘labeling device’ that associates each world in W with a set of sentences in L . As we will see, V is going to behave as a standard valuation function at possible worlds, but will behave non-standardly at impossible worlds. For ease of exposition, we will henceforth say that w n -entails A iff $V(w)$ n -entails A .

In light of the collapse result, we do not want to impose any closure constraints on impossible worlds. Doing so would result in a problematic kind of logical omniscience. To avoid any such problems, we will instead adopt a highly liberal *comprehension principle*, as Nolan (1997) calls it, according to which no set of sentences is too logically ill-behaved to count as an impossible world. More precisely:

Comprehension principle: For any incomplete and/or inconsistent set of sentences $\Gamma \subseteq L$, there is a world $w \in W^I$ such that $V(w) = \Gamma$.

This effectively means that we will take an ‘American stance’ on impossible worlds. However, while the American-style approach discussed in the previous section fails to retain a proper measure of logical competence, our approach is going to avoid this pitfall.

We now turn to the dynamic part of our semantic framework. To make it easier to parse the definitions below, it will be helpful to have the intuitive picture in mind. So, consider a simple sentence of the form ‘ $\langle n \rangle(Cr(A) = 1)$ ’. The semantics below will tell us that this sentence is true iff A is n -step inferable from every doxastically possible world. This is meant to capture the idea that the agent is in a position to become certain of A after having performed some n -step reasoning process. But note that even if the agent is capable of performing n steps of reasoning, she need not be in a position to become *certain* of A after some n -step reasoning process. After all, she might be uncertain of one or more of the *premises* involved in the relevant reasoning process. At the level of worlds, this amounts to saying that one or more of the premises might fail to be true at one or more doxastically possible worlds. More generally, then, our semantics will say that ‘ $\langle n \rangle(Cr(A) = x)$ ’ is true iff the probabilities of those doxastically possible worlds that n -entail A sum up to x .

This informal characterization of our semantics already shows that the truth-conditions for ‘ $\langle n \rangle(Cr(A) = x)$ ’ will be weaker than those for ‘ $Cr(A) = x$ ’: while the semantics for ‘ $Cr(A) = x$ ’ requires that the probabilities of those doxastically possible worlds that *verify* A sum up to x , the semantics for ‘ $\langle n \rangle(Cr(A) = x)$ ’ merely requires that the probabilities of those doxastically possible worlds that *n-entail* A sum up to x . So, for example, the truth-conditions for ‘ $\langle n \rangle(Cr(A) = 1)$ ’ will be weaker than those for ‘ $Cr(A) = 1$ ’: while the semantics for ‘ $Cr(A) = 1$ ’ requires that A is *true* at every doxastically possible world, the semantics for ‘ $\langle n \rangle(Cr(A) = 1)$ ’ merely requires that every doxastically possible world *n-entails* A .

That’s the intuitive picture; now for the formal details. We begin by defining a formal device that allows us to capture what is *n-step inferable* from a given world:

n-radius: The *n-radius* of a world, w , is written ‘ w^n ’ and is defined as follows:

$$w^n = \begin{cases} \{w\} & \text{for } w \in W^P. \\ \{w' \in W^I : V(w) \subseteq V(w') \text{ and } V(w) \vdash_R^n V(w')\} & \text{for } w \in W^I. \end{cases}$$

Each member of w^n is called an *n-expansion* of w .

Let us unpack this definition a bit. The idea is that any given world is associated with an ‘*n-radius*’, which is the set of ‘*n-expansions*’ of that world. Each *n-expansion* is itself a world; so the *n-radius* of a world is a *set* of worlds. Now, *which* worlds count as an *n-expansion* of w depends on whether w is possible or impossible. If w is possible, w is its own unique *n-expansion*, and so the *n-radius* of w is a singleton set: $w^n = \{w\}$, for any n . This reflects the fact that possible worlds are deductively closed entities that already verify everything that follows from them in any number of steps. More interestingly, if w is impossible, w' is an *n-expansion* of w iff the following three conditions are satisfied: (i) w' is impossible; (ii) w' verifies everything that w verifies; and (iii) everything that w' verifies is *n-step inferable* from w . It follows that every impossible world is an *n-expansion* of itself, just as every

possible world is an n -expansion of itself. However, in contrast to possible worlds, impossible worlds generally have more than one n -expansion. For example, suppose that $V(w) = \{A \rightarrow B, \neg B\}$, $V(w_1) = \{A \rightarrow B, \neg B, \neg A\}$ and $V(w_2) = \{A \rightarrow B, \neg B, \neg B \vee C\}$. Here w_1 and w_2 both count as 1-expansions of w (assuming that R contains modus tollens and disjunction introduction).

Since the \vdash_R^n -relation is monotonic in n , any i -expansion of w is also a j -expansion of w , for $i \leq j$. The opposite is obviously not the case. Thus, we can think of w^0, w^1, w^2, \dots as a sequence of concentric circles that stand in the following subset relations:

$$w^0 \subseteq w^1 \subseteq w^2, \dots$$

Assuming, as we will henceforth do, that no inference can be carried out in zero steps (except for the trivial inference ‘ A , therefore A ’), the 0-expansion of a world contains just the world itself. That is: $w^0 = \{w\}$, for any $w \in W$.

In addition to the n -radius of a world, we will also need a way to pick out exactly one n -expansion from the n -radius of each doxastically possible world. The following *choice function* allows us to do just that:

Choice function: Let $C : 2^{2^W} \mapsto 2^{2^W}$ be a function that takes a set $\mathbf{W} = \{W_1, \dots, W_m\}$ of sets of worlds as input and returns the set $C(\mathbf{W})$ of sets of worlds that results from all the ways in which exactly one element can be picked from each $W_i \in \mathbf{W}$. Each member of $C(\mathbf{W})$ is called a *choice* of \mathbf{W} .

An example will help illustrate this somewhat cumbersome definition. Let $\mathbf{W} = \{\{w_1\}, \{w_2, w_3\}\}$. A *choice* of \mathbf{W} is a set of worlds formed by picking exactly one world from each member of \mathbf{W} . In the case at hand, there are precisely two ways of doing so: we can either pick w_1 and w_2 or w_1 and w_3 . Accordingly, the choice function maps \mathbf{W} to a set containing those two choices: $C(\mathbf{W}) = \{\{w_1, w_2\}, \{w_1, w_3\}\}$.

While the choice function is defined in a highly general way, it will serve a much more specific purpose in what follows: it will allow us to capture all the different ways in which one can pick exactly one n -expansion of each

doxastically possible world from a given world. For present purposes, then, we can think of a choice as a set of worlds formed by picking exactly one n -expansion of each world in S .

The reason why choices are important is that they will allow us to distinguish semantically between the dynamic operators, $\langle n \rangle$ and $[n]$. Here is the rough idea: consider again a simple sentence of the form ' $\langle n \rangle(Cr(A) = 1)$ '. For such a sentence to be true, we do not want to require that *every* n -expansion of each doxastically possible world verifies A . We only want to require that *at least one* n -expansion of each doxastically possible world verifies A . By contrast, for ' $[n](Cr(A) = 1)$ ' to be true, we *do* want to require that every n -expansion of each doxastically possible world verifies A . The same goes for all sentences of the form ' $\langle n \rangle(Cr(A) = x)$ ' and ' $[n](Cr(A) = x)$ '.

Next up is the most central notion of our semantic framework, which will govern the truth-conditions of the dynamic operators. Let ' $\overset{n}{\sim}$ ' be a binary relation that holds between pairs of pointed models (where, as usual, a pointed model consists of a model and a world). If the relation holds between two pointed models, (M, w) and (M', w') , we write ' $(M, w) \overset{n}{\sim} (M', w')$ ' and say that (M', w') is *n-accessible* from (M, w) . Since the formal definition of n -accessibility will get a bit ugly, it will be helpful to begin with a sketch of the intuitive idea: Suppose that the pointed model (M, w) characterizes an agent's doxastic state at a given time. We then want to say that (M', w') is n -accessible from (M, w) iff (M', w') characterizes a doxastic state that the agent can enter by performing some n -step reasoning process. At the level of worlds, this amounts to saying that (M', w') is n -accessible from (M, w) iff the set of doxastically possible worlds at w' in M' corresponds to a choice of n -expansions of the doxastically possible worlds at w in M .

To ensure that the $\overset{n}{\sim}$ -relation behaves in the desired way, we need a way to replace a set of doxastically possible worlds with a choice of n -expansions of those worlds. The notion of an *n-variation* will allow us to do just that:

n -variation: Let $M = (W^P, W^I, f, V)$ be a model, and let $f(w) = (S, Pr)$ be the subjective probability space associated with (M, w) . The function Var^n (for $n = 0, 1, 2, \dots$) associates (M, w) with a set of subjec-

tive probability spaces:

$$Var^n(M, w) = \left\{ f' \in \mathbf{S} : \begin{cases} f'(w') = f_c(w') & \text{for } w' = w \\ f'(w') = f(w') & \text{for } w' \neq w \end{cases} \right\},$$

where $f_c(w) = (c, Pr_c)$ is a subjective probability space such that $c \in C(\{w'^n | w' \in S\})$, and Pr_c is the unique probability distribution over c such that, for each $w' \in S$, $Pr_c(w'_c) = Pr(w')$, where w'_c is the n -expansion of w' in c . Each member of $Var^n(M, w)$ is called an n -variation of (M, w) .

The way to understand this definition is as follows: We start with a pointed model, (M, w) , which is associated with a subjective probability space, $f(w) = (S, Pr)$. This subjective probability space consists of a set of doxastically possible worlds, S , and a distribution over those worlds, Pr . Now we modify S in a particular way: we replace each doxastically possible world with an n -expansion of that world. That is, we replace S with a choice, c , of n -expansions of the worlds in S . We keep the distribution fixed: each of the chosen n -expansions is assigned the same probability that was assigned to its corresponding world in S . More precisely, for any $v \in S$, if v' is the chosen n -expansion of v , we let Pr_c be a distribution such that $Pr_c(v') = Pr(v)$. We then define a function, f' , such that f' is identical to f except at w , where $f'(w) = (c, Pr_c)$. What we have ended up with is an n -variation of (M, w) .

Three properties of the n -variation of a pointed model are worth keeping in mind. First, there are in general many different n -variations of a given pointed model; indeed, as many as there are ways of forming choices of n -expansions of the doxastically possible worlds. Second, if f' is an i -variation of (M, w) , f' is also a j -variation of (M, w) , for $i \leq j$. Again, this follows from the fact that the \vdash_R^n -relation is monotonic in n . Third, since every world is an n -expansion of itself, f will itself be an n -variation of (M, w) .

We can now define the $\overset{n}{\sim}$ -relation:

n -accessibility: Let $M = (W^P, W^I, f, V)$ and $M' = (W'^P, W'^I, f', V')$ be two models. Then $(M, w) \overset{n}{\sim} (M', w')$ iff $w' = w$, $W'^P = W^P$, $W'^I = W^I$, $V' = V$, and $f' \in Var^n(M, w)$.

According to this definition, (M', w') is n -accessible from (M, w) iff the following two conditions are met: (i) f' is an n -variation of (M, w) ; and (ii) (M', w') is otherwise identical to (M, w) .

By construction, the definition reflects the intuitive role that we wanted the n -accessibility relation to play: if (M, w) characterizes an agent's doxastic state at a given time, the n -accessible pointed models from (M, w) represent all the different doxastic states that the agent can enter by performing some n -step reasoning process. As such, the $\overset{n}{\sim}$ -relation gives us a formally precise way of capturing what the agent can and cannot infer given her limited cognitive resources.

The three properties that we highlighted about the notion of an n -variation carry over to the notion of n -accessibility as well. First, there are in general many different n -accessible pointed models from a given pointed model; as many as there are n -variations of that pointed model. Second, if (M', w') is i -accessible from (M, w) , (M', w') is also j -accessible from (M, w) , for $i \leq j$. Third, (M, w) is always n -accessible from itself, since the 'empty' line of reasoning—not performing any inference at all—is always within the agent's cognitive reach, for any value of n .

This puts us in a position to complete our semantics for L . As usual, sentences are evaluated for truth and falsity at pointed models. We write ' \models ' and ' \models ' for verification and falsification, respectively.

For any possible world, $w \in W^P$:

- (P1) $M, w \models p$ iff $p \in V(w)$, where p is an atomic sentence.
- (P2) $M, w \models \neg A$ iff $M, w \not\models A$.
- (P3) $M, w \models A \wedge B$ iff $M, w \models A$ and $M, w \models B$.
- (P4) $M, w \models Cr(A) \leq x$ iff $\sum_Q Pr(Q) \leq x$, where $Q = \{v \in S : M, v \models A\}$.
- (P5) $M, w \models \langle n \rangle A$ iff $M', w' \models A$ for some $(M', w') : (M, w) \overset{n}{\sim} (M', w')$.
- (P6) $M, w \models [n] A$ iff $M', w' \models A$ for all $(M', w') : (M, w) \overset{n}{\sim} (M', w')$.
- (P7) $M, w \models \dashv A$ iff $M, w \not\models A$.

For any impossible world, $w \in W^I$:

- (I1) $M, w \models A$ iff $A \in V(w)$.
- (I2) $M, w \models \dashv A$ iff $\neg A \in V(w)$.

For the central results below, logical validity is defined in terms of truth-preservation across all possible worlds: $\Gamma \models A$ iff every possible world in every model is such that it verifies every sentence in Γ only if it verifies A .

A few remarks about the various satisfaction clauses are in order. First, note that falsehood behaves classically at both possible and impossible worlds: a sentence is false iff its negation is true. But, in contrast to possible worlds, impossible worlds can contain *truth-value gaps* (sentences that are neither true nor false) and *truth-value gluts* (sentences that are both true and false).

Second, note that (P1)-(P4) are identical to a standard possible-worlds semantics for probabilistic modal logic, except the semantics for the credence operator, (P4), which quantifies over both possible and impossible worlds. This is what allows us to steer clear of logical omniscience.

Third, (P5) says that sentences of the form $\langle n \rangle A$ are true at a pointed model, (M, w) , iff A is true at *some* n -accessible pointed model from (M, w) . In particular, $\langle n \rangle (Cr(A) = x)$ is true at (M, w) iff $Cr(A) = x$ is true at some n -accessible pointed model from (M, w) . This reflects the idea that the agent can come to have a credence of x in A by performing some n -step reasoning process *provided that* there is an n -accessible doxastic state from her current doxastic state at which she has a credence of x in A . Since the \vdash_R^n -relation is monotonic in n , any pointed model that verifies $\langle i \rangle A$ will also verify $\langle j \rangle A$, for $i \leq j$.

Finally, (P6) says that sentences of the form $[n]A$ are true at (M, w) iff A is true at *every* n -accessible pointed model from (M, w) . In particular, $[n](Cr(A) = x)$ is true at (M, w) iff $Cr(A) = x$ is true at *every* n -accessible pointed model from (M, w) . This reflects the idea that the agent will come to have a credence of x in A *regardless* of which n -step reasoning process she performs. Since a pointed model is always n -accessible from itself, the semantics for $[n](Cr(A) = x)$ is equivalent to that of $Cr(A) = x$. That is, $[n](Cr(A) = x)$ and $Cr(A) = x$ are true under exactly the same circumstances. This might seem to deprive the $[n]$ -operator of much of its interest. Indeed, our main focus in what follows will be on the $\langle n \rangle$ -operator. But the semantics for the $[n]$ -operator captures the aforementioned idea that the ‘empty’ line of reasoning is always within an agent’s cognitive reach, for any

value of n .

With our semantics in place, we can establish the first of our main results (all proofs can be found in the Appendix):

Theorem 1 (n -preservation)

If $A \vdash_R^n B$, then $Cr(A) = x \models \langle n \rangle (Cr(B) \geq x)$.

This result says that, if A n -entails B , and the agent’s credence in A is x , there is an n -step inference such that, after having performed that inference, the agent’s credence in B is *at least* x . For example, if the agent is 70% confident that “it rains,” she will be at least 70% confident that “it rains or snows” after having performed some 1-step inference (assuming that R contains disjunction introduction).

We can think of n -preservation as a non-ideal analogue of Classical Preservation. In contrast to Classical Preservation, n -preservation does not carry any commitment to logical omniscience: it does not describe an agent’s credences as being preserved across logical entailment. In fact, for all n -preservation says, an agent’s credences need not be preserved across *any* logical entailments. Yet, n -preservation allows us to retain a central trait of logical competence: it describes agents as being in a position to preserve their credences across those entailments that lie within their cognitive reach.

The second of our main results is a non-ideal analogue to Classical Certainty:

Theorem 2 (n -certainty)

If $\vdash_R^n A$, then $\models \langle n \rangle (Cr(A) = 1)$.

According to n -certainty, if A is an ‘ n -step tautology’—that is, if A is n -step inferable from the empty set—then an agent can come to be certain of A after having performed some n -step reasoning process. For example, the agent can come to be certain that “it’s either raining or not” after some 1-step reasoning process (assuming that $A \vee \neg A$ is 1-step provable in R).

In contrast to Classical Certainty, n -certainty does not carry any commitment to logical omniscience: it does not describe agents as being certain of all tautologies. In fact, for all n -certainty says, agents need not be certain of

any tautologies. Yet, n -certainty allows us to retain a central trait of logical competence: it describes agents as being in a position to become certain of any tautology that lies within their cognitive reach.

Together, n -preservation and n -certainty show how our dynamic framework avoids the problems that faced the static approaches discussed in the previous section: it allows us to model agents who are logically competent despite falling short of logical omniscience. But we are not home free yet. The unorthodox nature of our approach gives rise to a number of questions that need to be addressed. That is the task of the next section.

4 DAMAGE CONTROL: BEYOND ORTHODOX BAYESIANISM

It is not cost-free to give up the assumption of logical omniscience. Without it, many fundamental results of orthodox Bayesianism do not go through. But all is not lost. Just as our dynamic framework provides us with non-ideal analogues of Classical Preservation and Classical Certainty, so it provides us with non-ideal analogues of various other centerpieces of orthodox Bayesianism. Here we focus on two in particular.

First up is the notion of a *conditional credence*. Bayesians typically define conditional credences in terms of ratios of unconditional credences:¹³

$$\mathbf{Ratio\ Formula:} \quad Cr(A|B) = \frac{Cr(A \wedge B)}{Cr(B)}$$

This definition is sensible as long as conjunctions relate to their conjuncts in the usual, truth-functional way. But conjunctions do *not* behave in the usual, truth-functional way in our framework: impossible worlds show no respect for classical truth-functional dependencies between conjunctions and their conjuncts. In particular, $A \wedge B$ need not be true just because A and B are. Hence, the Ratio Formula makes little sense in our framework.

The stakes are high: without the Ratio Formula, the standard derivation of Bayes' theorem is blocked. Suddenly it looks like we are throwing out the baby with the bathwater.

¹³A notable exception is Hájek (2003).

But a fix is available. Instead of the Ratio Formula, we can define conditional credences as follows:

$$(P8) \quad M, w \models Cr(A|B) \leq x \text{ iff } \frac{\sum_Q Pr(Q)}{\sum_{Q'} Pr(Q')} \leq x,$$

where $w \in W^P$, $Q = \{v \in S : M, v \models A \text{ and } M, v \models B\}$, and $Q' = \{v \in S : M, v \models B\}$. This definition captures much the same idea as the Ratio Formula: to determine an agent's credence of A conditional on B , we look at those B -worlds that are doxastically possible for the agent and check which of those worlds verify A . Furthermore, it is easily verified that (P8) and (P4) jointly entail Bayes' theorem. The danger is averted.

Our definition of conditional credence also allows us to establish a third main result:

Theorem 3 (n -conditionality)

If $A \vdash_R^n B$, then $\models \langle n \rangle (Cr(B|A) = 1)$.

According to n -conditionality, if A n -entails B , then an agent can become certain of B conditional on A after having performed some n -step reasoning process. For example, the agent can become certain that “it rains or snows” conditional on “it rains” after some 1-step reasoning process (assuming that R contains disjunction introduction).

We can think of n -conditionality as a non-ideal analogue of the following Bayesian principle:

Classical Conditionality: If A entails B , then $Cr(B|A) = 1$.

This principle captures yet another way in which orthodox Bayesianism gives rise to logical omniscience: intuitively, it describes agents as being certain of all entailment relations. By contrast, n -conditionality carries no such commitment. Indeed, for all n -conditionality says, agents need not be certain of *any* entailment relations. Yet, n -conditionality allows us to retain a central trait of logical competence: it describes agents as being in a position to become certain of those entailment relations that lie within their cognitive reach.

The second aspect of orthodox Bayesianism that we want to focus on is

its *algebraic* structure. The story is familiar: by defining a Boolean algebra on the set of possible worlds, we can understand logical operations in terms of set-theoretic ones. For example, we can understand conjunction and disjunction in terms of intersection and union (where ‘ $|A|$ ’ denotes the set of possible worlds that verify A):

Classical Conjunction: $|A \wedge B| = |A| \cap |B|$

Classical Disjunction: $|A \vee B| = |A| \cup |B|$

These principles do not generally hold in our framework. More specifically, they hold at the level of possible worlds, but fail at the level of impossible worlds. The reason, once again, is that impossible worlds show no respect for classical, truth-functional dependencies between complex sentences and their parts. For example, $A \vee B$ need not be true just because A is.

However, we can still formulate non-ideal analogues to Classical Conjunction and Classical Disjunction (where ‘ $|A|^n$ ’ denotes the set of worlds in W that have at least one n -expansion that verifies A):¹⁴

n -conjunction: $|A \wedge B|^n \subseteq |A|^{n+1} \cap |B|^{n+1}$
 $|A|^n \cap |B|^n \subseteq |A \wedge B|^{n+1}$

n -disjunction: $|A \vee B|^n \cap |A|^n \subseteq |B|^{n+1}$
 $|A|^n \subseteq |A \vee B|^{n+1}$

These principles show that our dynamic framework, while not truth-functional in the classical sense, still allows us to associate set-theoretic properties with various logical connectives; properties that nicely capture the roles that such connectives play in our cognitive lives. This strikes us as an interesting result in its own right.

¹⁴Here is a sketch of a proof of the first part of n -conjunction: suppose $w \in |A \wedge B|^n$, for any $w \in W$. Assuming that R contains standard introduction and elimination rules for conjunction, $|A \wedge B|^n \subseteq |A|^{n+1}$ and $|A \wedge B|^n \subseteq |B|^{n+1}$. Hence, $w \in |A|^{n+1} \cap |B|^{n+1}$. The other subset relations can be established in similar ways.

5 CONCLUDING REMARKS

We began this paper with a critical discussion of three existing approaches to the problem of logical omniscience in the Bayesian literature. Some proposals merely replaced logical omniscience with a different logical idealization; others sacrificed all traits of logical competence on the altar of logical omniscience. The collapse result made the waters hard to navigate. But in diagnosing *why*, a new ‘dynamic’ approach emerged: by enriching the Bayesian framework with tools that allowed us to model what agents can and cannot infer given their limited cognitive resources, hope remained to circumvent collapse. We went on to develop this dynamic approach in formal detail, and showed how the resulting Bayesian framework allows us to model agents who are logically competent despite falling short of logical omniscience.

Let us close by addressing a residual worry about our dynamic approach, due to Berto & Jago (2019, §5.5). The worry goes as follows: while our framework allows us to model what agents *can* infer given his cognitive resources, it does not allow us to model what they *should* infer given those resources (since the semantics for ‘ $[n](Cr(A) = x)$ ’ is equivalent to that of ‘ $Cr(A) = x$ ’). Yet it is the job of a theory of non-ideal rationality to tell us how non-ideal agents *should* live their epistemic lives. After all, rationality is a normative notion; not a descriptive one.

We want to offer two remarks in reply. First, it is worth noting that there is at least a weak sense in which our framework *is* normative. If we accept that ‘ought’ implies ‘can’ in the domain of epistemic rationality (which is obviously a big ‘if’), then agents will not be required to live their epistemic lives in ways that are incompatible with their cognitive abilities. Thus, insofar as our dynamic framework allows us to represent an agent’s cognitive abilities, it will at least arguably place *negative* requirements on how agents ought to live their epistemic lives.

Second, and perhaps more importantly, we are doubtful that a formal theory of non-ideal rationality should indeed place any *positive* demands on which inferences ordinary agents should perform. After all, if an agent performed every inference within her cognitive reach, she would end up ‘clut-

tering her mind with trivialities,' to use a rubric from Harman (1986, p. 12). The situation seems analogous to that of evidence-gathering: if an agent gathered every piece of evidence within her practical reach, she would most likely end up with a massive pile of useless junk. Yet, it is presumably not the task of formal epistemology to say which pieces of evidence, among the practically feasible ones, the agent should gather. Likewise, we do not consider it the task of our dynamic Bayesian framework to say which inferences, among the epistemically feasible ones, agents should perform.

Acknowledgments. An earlier version of this paper was presented at the “Normative Notions Formalized” Workshop in Munich. Many thanks to the audience on that occasion. Thanks also to two anonymous referees from *Erkenntnis* for very detailed and helpful comments.

APPENDIX

This appendix contains proofs of three main results of the paper. All definitions can be found in §3. The results are repeated here for convenience.

Theorem 1 (*n*-preservation)

If $A \vdash_R^n B$, then $Cr(A) = x \models \langle n \rangle (Cr(B) \geq x)$.

Proof. Suppose that $A \vdash_R^n B$ and consider any pointed model, (M, w) , such that $M, w \models Cr(A) = x$, where $M = (W^P, W^I, f, V)$ and $w \in W^P$. We must show that $M, w \models \langle n \rangle (Cr(B) \geq x)$. We proceed by defining a suitable *n*-accessible pointed model from (M, w) . Let $M' = (W'^P, W'^I, f', V')$ be a model such that $W'^P = W^P$, $W'^I = W^I$, and $V' = V$. Since $A \vdash_R^n B$, we can let f' be an *n*-variation of (M, w) for which it holds that $f'(w) = (c, Pr_c)$, where $M', v \models B$, for all $v \in \{v' \in c : M', v' \models A\}$. By the definition of *n*-accessibility, then, $(M, w) \overset{n}{\sim} (M', w)$. Since $M, w \models Cr(A) = x$, (P4) tells us that $\sum_Q Pr(Q) = x$, where $Q = \{v \in S : M, v \models A\}$. Hence, $\sum_{Q'} Pr_c(Q') \geq x$, where $Q' = \{v \in c : M', v \models B\}$. By another application of (P4), $M', w \models Cr(B) \geq x$. So, by (P5), it follows that $M, w \models \langle n \rangle (Cr(B) \geq x)$. \square

Theorem 2 (n -certainty)

If $\vdash_R^n A$, then $\models \langle n \rangle (Cr(A) = 1)$.

Proof. Suppose that $\vdash_R^n A$ and let (M, w) be any pointed model such that $M = (W^P, W^I, f, V)$ and $w \in W^P$. We must show that $M, w \models \langle n \rangle (Cr(A) = 1)$. We proceed by defining a suitable n -accessible pointed model from (M, w) . Let $M' = (W'^P, W'^I, f', V')$ be a model such that $W'^P = W^P$, $W'^I = W^I$, and $V' = V$. Since $\vdash_R^n A$, we can let f' be an n -variation of (M, w) such that $f'(w) = (c, Pr_c)$, where $M', v \models A$, for all $v \in c$. Hence, $\sum_{Q'} Pr_c(Q') = 1$, where $Q' = \{v \in c : M', v \models A\}$. By (P4), $M', w' \models Cr(A) = 1$. By the definition of n -accessibility, $(M, w) \stackrel{n}{\sim} (M', w)$. So, by (P5), it follows that $M, w \models \langle n \rangle (Cr(A) = 1)$. \square

Theorem 3 (n -conditionality)

If $A \vdash_R^n B$, then $\models \langle n \rangle (Cr(B|A) = 1)$.

Proof. Suppose that $A \vdash_R^n B$ and let (M, w) be any pointed model such that $M = (W^P, W^I, f, V)$ and $w \in W^P$. We must show that $M, w \models \langle n \rangle (Cr(B|A) = 1)$. Let $M' = (W'^P, W'^I, f', V')$ be a model such that $W'^P = W^P$, $W'^I = W^I$, and $V' = V$. Since $A \vdash_R^n B$, we can let f' be an n -variation of (M, w) such that $f'(w) = (c, Pr_c)$, where $M', v \models B$, for all $v \in \{v' \in c : M', v' \models A\}$. Hence, $\sum_Q Pr_c(Q) = \sum_{Q'} Pr_c(Q')$, where $Q = \{v \in c : M', v \models A \text{ and } M', v \models B\}$ and $Q' = \{v' \in c : M', v' \models q\}$. By (P8), $M', w \models Cr(B|A) = 1$. By the definition of n -accessibility, $(M, w) \stackrel{n}{\sim} (M', w)$. So, by (P5), it follows that $M, w \models \langle n \rangle (Cr(B|A) = 1)$. \square

REFERENCES

- Baltag, A. and B. Renne (2016). “Dynamic Epistemic Logic”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2016. Metaphysics Research Lab, Stanford University.
- Berto, F. and M. Jago (2018). “Impossible Worlds”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2018. Metaphysics Research Lab, Stanford University.
- (2019). *Impossible Worlds*. Oxford: Oxford University Press.

- Bjerring, J. C. (2013). “Impossible Worlds and Logical Omniscience: An Impossibility Result”. In: *Synthese* 190.13, pp. 2505–2524.
- Bjerring, J. C. and W. Schwarz (2017). “Granularity Problems”. In: *Philosophical Quarterly* 67.266, pp. 22–37.
- Bjerring, J.C. and M. Skipper (forthcoming). “A Dynamic Solution to the Problem of Logical Omniscience”. In: *The Journal of Philosophical Logic*.
- Christensen, D. (2007). “Does Murphy’s Law Apply in Epistemology? Self-Doubt and Rational Ideals”. In: *Oxford Studies in Epistemology* 2, pp. 3–31.
- Cresswell, M. (1973). *Logics and Languages*. London: Methuen.
- Ditmarsch, H. van, W. van der Hoek, and B. Kooi (2008). *Dynamic Epistemic Logic*. Springer.
- Earman, J. (1992). *Bayes or Bust?* Bradford.
- Easwaran, K. (2011). “Bayesianism II: Applications and Criticisms”. In: *Philosophy Compass* 6.5, pp. 321–332.
- Eells, E. (1985). “Problems of Old Evidence”. In: *Pacific Philosophical Quarterly* 66.3, p. 283.
- Elga, A. and A. Rayo (ms.). “Logical omniscience and decision theory: a no-go result”. In: Unpublished manuscript.
- Elgot-Drapkin, J. and D. Perlis (1990). “Reasoning Situated in Time I: Basic Concepts”. In: *Journal of Experimental and Theoretical Artificial Intelligence* 2, pp. 75–98.
- Fagin, R. et al. (1995). *Reasoning About Knowledge*. MIT Press.
- Gaifman, H. (2004). “Reasoning with Limited Resources and Assigning Probabilities to Arithmetical Statements”. In: *Synthese* 140, pp. 97–119.
- Garber, D. (1983). “Old Evidence and Logical Omniscience in Bayesian Confirmation Theory”. In: *Testing Scientific Theories*. Ed. by J. Earman. Minneapolis: University of Minnesota Press.
- Glymour, C. (1980). *Theory and Evidence*. Princeton University Press.
- Good, I. (1968). “Corroboration, Explanation, Evolving Probability, Simplicity and a Sharpened Razor”. In: *British Journal for the Philosophy of Science* 19.2, pp. 123–143.
- Hacking, I. (1967). “Slightly More Realistic Personal Probability”. In: *Philosophy of Science* 34.4, pp. 311–325.
- Hájek, A. (2003). “What Conditional Probability Could Not Be”. In: *Synthese* 137.3, pp. 273–323.
- Harman, G. (1986). *Change in View*. MIT Press.
- Hartmann, S. and B. Fitelson (2015). “A New Garber-Style Solution to the Problem of Old Evidence”. In: *Philosophy of Science* 82.4, pp. 712–717.
- Hintikka, J. (1962). *Knowledge and Belief*. Ithaca, N.Y., Cornell University Press.
- (1975). “Impossible Possible Worlds Vindicated”. In: *Journal of Philosophical Logic* 4, pp. 475–484.

- Jago, M. (2013). “The Problem of Rational Knowledge”. In: *Erkenntnis* 6, pp. 1–18.
- Jeffrey, R. (1983). “Bayesianism With A Human Face”. In: *Testing Scientific Theories*. Ed. by J. Earman. University of Minnesota Press, pp. 133–156.
- Lakemeyer, G. (1987). “Tractable Meta-Reasoning in Propositional Logics of Belief”. In: *Tenth International Joint Conference on Artificial Intelligence*, pp. 402–408.
- Levesque, H. (1984). “A Logic of Implicit and Explicit Belief”. In: *National Conference on Artificial Intelligence*, pp. 198–202.
- Nolan, D. (1997). “Impossible Worlds: A Modest Approach”. In: *Notre Dame Journal of Formal Logic* 38.4, pp. 535–572.
- Rantala, V. (1982). “Impossible Worlds Semantics and Logical Omniscience”. In: *Acta Philosophica Fennica* 35, pp. 106–15.
- Rasmussen, M. Skipper (2015). “Dynamic Epistemic Logic and Logical Omniscience”. In: *Logic and Logical Philosophy* 24, pp. 377–399.
- Smithies, D. (2015). “Ideal Rationality and Logical Omniscience”. In: *Synthese* 192.9, pp. 2769–2793.
- Sprenger, J. (2015). “A Novel Solution to the Problem of Old Evidence”. In: *Philosophy of Science* 82.3, pp. 383–401.
- Talbott, W. (2016). “Bayesian Epistemology”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. Zalta. Winter 2016. Metaphysics Research Lab, Stanford University.
- Titelbaum, M. (2015). “Rationality’s Fixed Point (Or: In Defense of Right Reasons)”. In: *Oxford Studies in Epistemology* 5, pp. 253–94.
- (forthcoming). *Fundamentals of Bayesian Epistemology*. Oxford: Oxford University Press.
- Wright, G. von (1951). *An Essay in Modal Logic*. Amsterdam: NorthHolland Pub. Co.