A dynamic solution to the problem of logical omniscience

Jens Christian Bjerring · Mattias Skipper
Penultimate draft, forthcoming in *The Journal of Philosophical Logic*

Abstract. The traditional possible-worlds model of belief describes agents as 'logically omniscient' in the sense that they believe all logical consequences of what they believe, including all logical truths. This is widely considered a problem if we want to reason about the epistemic lives of non-ideal agents who—much like ordinary human beings—are logically competent, but not logically omniscient. A popular strategy for avoiding logical omniscience centers around the use of *impossible worlds*: worlds that, in one way or another, violate the laws of logic. In this paper, we argue that existing impossible-worlds models of belief fail to describe agents who are both logically non-omniscient and logically competent. To model such agents, we argue, we need to 'dynamize' the impossible-worlds framework in a way that allows us to capture not only what agents *believe*, but also what they are able to *infer* from what they believe. In light of this diagnosis, we go on to develop the formal details of a dynamic impossible-worlds framework, and show that it successfully models agents who are both logically non-omniscient and logically competent.

Keywords Logical omniscience · Resource-bounded reasoning · Bounded rationality · Impossible worlds · Doxastic logic · Epistemic logic

1 Introduction

Consider the standard possible-worlds model of belief:[1]

[1]For early developments of this approach to the semantics of belief (and related notions), see Hintikka (1962) and von Wright (1951).

(Belief) An agent believes a proposition p iff p is true at all possible worlds that are doxastically possible for the agent.

Whatever else might be said about the nature of possible worlds, they are usually assumed to respect the laws of classical logic. That is, the truths at any given possible world form a deductively closed and consistent set. As a consequence, the possible-worlds model of belief describes agents as 'logically omniscient' in the sense that they believe all logical consequences of what they believe, including all logical truths. To see why, suppose that an agent believes a proposition p, and let q be any logical consequence of p. Since the agent believes p, p is true at all possible worlds that are doxastically possible for the agent. And since p entails q, all possible worlds that verify p also verify q. Hence q is true at all doxastically possible worlds for the agent, which means that the agent believes q. So if the agent believes p, she believes all logical consequences of p.

The assumption of logical omniscience is widely considered a problem if we want to reason about the epistemic lives of agents who—much like ordinary human beings—are logically competent, but not logically omniscient. It may well be that Goldbach's Conjecture follows from the Peano Axioms. Yet, just because I believe that the Peano Axioms are true, I need not believe that Goldbach's Conjecture is true. The same goes for artificial agents such as computers or robots that are subject to computational limitations. While such agents are able to compute *some* of the logical consequences of what is stored in their memory, they are not able to compute *all* such consequences. Relatedly, the assumption of logical omniscience is a problem if we want to develop a normative theory of belief that is sensitive to the cognitive limitations of ordinary agents. As Weirich (2004) points out, while such agents fail to live up to the idealized rationality standards of logically omniscient agents, they need not "be irrational in any way. They may fully conform to all standards for agents with their limitations" (Weirich 2004, p. 100). So if we want to reason about bounded rationality, logical omniscience should be avoided.

If our sole goal is to model agents who fall short of logical omniscience,

our job is done once we have a model of agents who fail to believe all logical consequences of what they believe. As we shall see in §2, there are worked out formal theories of belief that meet this objective. But if we—like Cherniak (1986), Jago (2013; 2014), Weirich (2004), and many others—are interested in logically non-omniscient agents who are nevertheless (to some extent) logically *competent*, there is more work to be done. For, as we will shall see, it is a non-trivial task to avoid logical omniscience without sacrificing all traits of logical competence.

What do we mean by 'logically competence'? As a first pass, we will say that an agent is logically competent when she at least does not miss out on any *trivial* logical consequences of what she believes. To get an intuitive handle on this idea, consider the following test: suppose an agent believes p, and let q be any trivial consequence of p. We can then ask: upon being asked whether q is the case, is the agent immediately able to answer "yes"? If she is, she passes the test and counts as logically competent. For example, suppose you believe that it rains and that it rains only if the streets are wet. We can then ask: are you able to immediately answer "yes" when asked whether the streets are wet? Assuming that you are attentive, mentally well-functioning, and so on, it surely seems so. So you do not miss out on this trivial logical consequence of your beliefs, and hence count as logically competent in the relevant sense.

Why should we care about agents who are logically competent in roughly this sense? Because many of the reasons for being interested in logically *non-omniscient* agents are also reasons for being interested in logically *competent* agents. Suppose we aim to model real-world agents, such as humans or computers, who have the ability to perform at least fairly simple chains of logical reasoning. You are about to leave your house and notice that it snows heavily outside. Well aware of your fragile health, you believe that if it snows heavily, you should wear a winter jacket when outside. Surely, in the normal run of things, you will wear the winter jacket when leaving your house. To explain this behavior, we must appeal not only to your desire not to get sick, but also to your ability to engage in simple logical reasoning. For based on your beliefs that it snows heavily and that if it snows heavily, you should wear

a winter jacket, you can infer that you should wear a winter jacket. Had you missed out on this trivial consequence of your beliefs, you would not have been able to act in ways that would satisfy your desire not to get sick. This is not to say that ordinary humans never make mistakes even in simple logical reasoning. But we take it that most ordinary people have at least a basic ability—albeit a fallible one—to engage in simple logical reasoning; and it is this ability that we want to capture by developing a model of agents who do not miss out on any trivial logical consequences of what they believe.

Obviously, what counts as a 'trivial' logical consequence depends on the cognitive resources that agents have available for logical reasoning. If you are an experienced logician, it might be trivial for you to see that '$\neg q \rightarrow \neg p$' entails '$\neg(p \wedge \neg q)$', whereas this inference may be non-trivial for a first-year philosophy student. To capture this agent-relativity, we will adopt a simple *step-based* picture of what it means to reason with limited cognitive resources.[2] On this picture, agents reason logically by applying rules from a set $\mathcal{R}$ of inference rules, where one 'step' of reasoning corresponds to one application of a rule in $\mathcal{R}$. This allows us to model an agent's cognitive resources in terms of the number n of steps of reasoning in $\mathcal{R}$ that the agent can easily or trivially perform. In the limit where $n = 0$, agents have no cognitive resources available, and hence no logical consequences will count as trivial (assuming that nothing is provable in zero steps of reasoning). In the opposite limit where n approaches infinity, agents have unlimited cognitive resources, in which case all logical consequences, however complicated, will count as trivial. In-between these extremes, we find a spectrum of agents with different levels of cognitive resources. For such agents, *some* but not *all* logical consequences will count as trivial.

We can then give the following precisification of what it means for a logical consequence to be trivial:

> **(Trivial consequence)** A proposition q is a trivial logical consequence of a set Γ of propositions iff q can be inferred from Γ within n applications of the inference rules in $\mathcal{R}$.

[2] We are here inspired by work in *active logic* (or *step logic*) by Drapkin et al. (1999), Drapkin & Perlis (1986; 1990), and others.

Accordingly, we will say that an agent counts as 'logically competent' just in case she has the ability to infer at least the trivial logical consequences of what she believes. Since such agents have the ability to tease out what follows within n steps of logical reasoning from what they believe, they do not miss out on any trivial logical consequences of what they believe.

What counts as a trivial consequence of a given set of propositions depends not only on the value of n, but also on the rules included in $\mathcal{R}$. If $\mathcal{R}$ contains only modus tollens, $\neg p$ will be the only trivial consequence of $\{p \to q, \neg q\}$, no matter how high n goes. By contrast, if $\mathcal{R}$ is a complete proof system for classical propositional logic, the number of trivial consequences of $\{p \to q, \neg q\}$ can be much larger, depending on the value of n. We will deliberately leave the specification of $\mathcal{R}$ and n open in what follows. In doing so, our framework will be applicable in a wide range of theoretical contexts that may call for different sets of inference rules and different levels of cognitive resources. For instance, someone who is interested in relatively complex agents with powerful reasoning mechanisms and high levels of cognitive resources may let $\mathcal{R}$ be a rich proof system and choose a relatively high value of n. By contrast, someone who is interested in relatively simple agents with weak reasoning mechanisms and low levels of cognitive resources may give a sparse characterization of $\mathcal{R}$ and choose a relatively low value of n.

With these preliminaries in place, our aim in the remainder of the paper is to develop a model of agents who are logically non-omniscient, yet logically competent in the sense specified above. We will build our model on a version of the impossible-worlds framework developed by Cresswell (1970; 1972; 1973), Fagin et al. (1995), Hintikka (1975), Rantala (1982), Wansing (1990), and others. Subtleties aside, such models retain (Belief) as the semantics for belief, but extend the underlying space of worlds to include *impossible worlds*: worlds that, in one way or another, violate the laws of classical logic (we return to the details below). To set the stage for our proposal, we will hence focus exclusively on the impossible-worlds framework and leave a discussion of other approaches to logical omniscience for another occasion.[3]

[3]For an overview of classical approaches to logical omniscience, see Fagin et al. (1995).

Here is an overview of the rest of the paper. In §2, we argue that existing impossible-worlds models of belief fail to describe agents who are both logically non-omniscient and logically competent. To model such agents, we argue in §3, we need to 'dynamize' the impossible-worlds framework in a way that allows us to model not only what agents *believe*, but also what they are able to *infer* from what they believe. In light of this diagnosis, we go on to develop the formal details of a dynamic impossible-worlds framework, and show that it successfully models agents who are both logically non-omniscient and logically competent. In §4, we argue that the proposed framework has a number of advantages over a related impossible-words model recently developed by Mark Jago (2013; 2014). Finally, in §5, we provide some concluding remarks.

2 Impossible worlds and logical omniscience

The central idea behind the impossible-worlds model of belief is to extend the space of possible worlds with what Hintikka called *impossible possible 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). The motivating thought is that, for agents with limited cognitive resources, the space of doxastic possibilities is larger than the space of logical possibilities. Since such agents may well believe each of the Peano Axioms but fail to believe Goldbach's Conjecture, the Peano Axioms can be true in all their doxastic alternatives, even if the Conjecture is not. Yet, if the Conjecture in fact follows from the Axioms, no logically possible world verifies the Axioms but falsifies the Conjecture. So doxastic possibility seems to outstrip logical possibility.

Consider then the following impossible-worlds model of belief:

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

Thus formulated, the impossible worlds model of belief says nothing about

the nature of impossible worlds. Corresponding to different conceptions of what impossible worlds are, we get different versions of the impossible-worlds model. On one conception—what Berto (2013) calls the "American stance"—impossible worlds are allowed to be arbitrarily logically ill-behaved: for any set of propositions, however blatantly inconsistent, some impossible world verifies just those propositions. On a second conception—what Berto (2013) calls the "Australasian stance"—impossible worlds are required to respect the laws of some non-classical logic. Yet, as we shall argue now, neither stance allows us to model agents who are both logically non-omniscient and logically competent.

The American stance: Suppose that impossible worlds are not closed under any notion of logical consequence. Impossible worlds then satisfy the following principle:

> **(Non-closure)** For any two propositions p and q, some impossible world verifies p but not q.

It is easily seen that logical omniscience is avoided if we accept (Non-closure): if some impossible world verifies p but not q, for any p and q, nothing prevents it from being the case that some doxastically possible worlds verify p but not q. So agents may believe p without believing q, and hence they need not believe all logical consequences of what they believe. At the same time, however, it is also easy to see that (Non-closure) does not give us the tools to model logically *competent* agents: if agents may believe p but not q, for *any* p and q, they need not believe *any* logical consequences of what they believe. As such, nothing in the formalism allows us to capture the sense in which ordinary agents are logically competent.

The Australasian stance: Suppose next that impossible worlds are closed under logical consequence in some non-classical logic L (e.g. intuitionistic or paraconsistent logic).[4] Impossible worlds then satisfy the following principle:

> **(Non-classical closure)** For any two propositions p and q such that q is a logical consequence of p in L, any impossible world that verifies p

[4]See Fagin et al. (1995), Levesque (1984), and Lakemeyer (1987) for approaches to logical omniscience that appeal to a non-classical logic.

verifies q.

Given that L is weaker than classical logic, this closure principle allows us to avoid logical omniscience with respect to *classical* logic: even if q follows from p in classical logic, agents may believe p without believing q. Hence they need not believe all classical consequences of what they believe.

However, agents are still characterized as logically omniscient with respect to the non-classical logic L. For instance, if we understand L as a paraconsistent logic, agents will believe all paraconsistent consequences of what they believe, including all paraconsistent tautologies. But just as it is implausible that cognitively limited agents believe all *classical* consequences of what they believe, so it is implausible that they believe all *paraconsistent* consequences of what they believe. Even supposing that such agents reason paraconsistently, they clearly cannot reason unlimited in that logic. So the Australasian stance still commits us to an undesirable kind of logical omniscience. Moreover, it does not adequately capture the sense in which agents are logically competent. After all, not all bounded agents have a non-classical reasoning mechanism, let alone the *same* reasoning mechanism. Consider, for instance, a proof generator for classical propositional logic. Such a generator reasons purely classically but nevertheless falls short of logical omniscience due to its computational limitations. So the strategy of describing this proof generator as omniscient with respect to some non-classical logic simply misses the target.

In light of these problems with the American and Australasian stances, one might naturally consider a closure principle on impossible worlds that closely reflects the characterization of trivial logical consequence from above:

> **(Trivial closure)** For any two propositions p and q such that q is a trivial logical consequence of p (that is, such that q is n-step inferable in $\mathcal{R}$ from p), any impossible world that verifies p also verifies q.

By closing impossible worlds under trivial logical consequence, we ensure that agents believe all trivial consequences of what they believe. Consider an agent who believes p, and let q be any trivial logical consequence of p. By (Belief-impossible), p is true at all worlds that are doxastically possible for

the agent. By (Trivial closure), q must then also be true at all such worlds. So, by (Belief-impossible), the agent believes q.

Needless to say, if we accept (Trivial closure), we must at the same time ensure that impossible worlds are not fully deductively closed with respect to $\mathcal{R}$. If they were, we would end up describing agents as logically omniscient with respect to $\mathcal{R}$. So the challenge is to satisfy both (Trivial closure) and the following principle:

> **(Deductive openness)** For some impossible world w and proposition p, the set of true propositions at w entails p in some number of steps in $\mathcal{R}$, but w does not verify p.

However, it turns out, perhaps surprisingly, to be impossible to satisfy both (Trivial closure) and (Deductive openness).[5] To see this, consider any inference in $\mathcal{R}$ from a set Γ of premises to a conclusion q. In order for (Trivial closure) and (Deductive openness) to be jointly satisfiable, there must be at least *one* impossible world w such that:

(i) w verifies all the premises in Γ;
(ii) w verifies all trivial logical consequences of the truths at w; and
(iii) w does not verify the conclusion q.

However, it can be shown that no world can jointly satisfy (i) to (iii). As a first step, note that since Γ entails q in $\mathcal{R}$, there exists a sequence of propositions $\langle \Gamma, q_1, q_2, \ldots, q \rangle$ corresponding to an inference from Γ to q by some number of applications of the rules in $\mathcal{R}$.[6] Given that $n \geq 1$ (that is, given that agents meet a minimal level of logical competence), it follows by

[5] Earlier versions of this argument can be found in Bjerring (2013; 2014) and Jago (2014).

[6] While the details of the inference from Γ to q depends on the rules in $\mathcal{R}$, nothing of importance hinges on whether we think of $\mathcal{R}$ as a natural deduction system, a sequent calculus, or some other proof system. To establish that (i) to (iii) are not jointly satisfiable, we only need the assumption that there exists an inference in $\mathcal{R}$ from Γ to q such that each step in the inference is trivial. And at least for standard rules such as conjunction introduction, modus ponens, and double negation elimination, it is plausible that each such rule is cognitively or computationally trivial to apply. For further motivation of these thoughts, see also Bjerring (2013; 2014), Bjerring and Schwarz (2017), and Jago (2014).

(Trivial consequence) that each q_i is a trivial consequence of q_{i-1}. Consider now a world w that satisfies (i) and (ii): w verifies each premise in Γ as well as every trivial logical consequence of the truths at w. It then follows that w verifies q_1. If it did not, it would fail to verify a trivial consequence of Γ and hence fail to satisfy (ii). But given that w verifies q_1, it must also verify q_2 since q_2 is a trivial logical consequence of q_1. Continuing this line of reasoning, it follows that w must verify q, and hence fail to satisfy (iii). So if w satisfies (i) and (ii), it cannot satisfy (iii). As such, (Trivial closure) and (Deductive openness) cannot be satisfied simultaneously: as soon as we attempt to close a world under trivial logical consequence, we end up closing it under full logical consequence. Intuitively, that is, a world that is closed under *trivial* logical consequence "collapses" under its own deductive weight to a world that is closed under *full* logical consequence.

This "collapse result" can in fact be established without appeal to any particular formal theory of belief. In line with the reasoning above, it is easy to see that no agent can satisfy the following conditions:

(1) The agent believes all the premises in Γ;
(2) The agent believes all trivial consequences of what she believes; and
(3) The agent does not believe the conclusion q.

When both (1) and (2) are satisfied, (3) cannot be. That is, if an agent believes every proposition that follows trivially from her beliefs, she ends up believing all logical consequences of what she believes. Hence we cannot model agents who are both logically non-omniscient and logically competent by closing beliefs under trivial logical consequence.

How then *can* we model such agents? We propose an answer in the next section.

3 A dynamic model of belief

Consider a logically non-omniscient agent who passes the test for logical competence: for any p and q such that q is a trivial consequence of p, if the agent believes p, then, upon being asked whether q is the case, she can

answer "yes" immediately. Given the collapse result, we know that we cannot model this ability by saying that the agent believes q prior to being asked about it. If she did, she would have to believe all logical consequences of her beliefs, and hence qualify as logically omniscient. Instead we can model how she passes the test by citing her ability to engage in logical reasoning. When asked about q, the question primes her to infer q from p and thereby move from a belief state that contains p to one that contains q as well. This enables the agent to answer the question about q positively despite not believing q to begin with.

We thus suggest that a proper solution to the problem of logical omniscience should appeal to an appropriate relation between doxastic states. This relation should be understood "dynamically" as a reasoning process that issues a transition from a doxastic state containing the premises of a given inference to an updated doxastic state that contains the conclusion as well.[7] To formalize this idea, we will develop a dynamic version of the impossible-worlds model of belief that will allow us to capture not only what agents *believe*, but also what they can *infer* from what they believe. We will approach matters from a purely model-theoretic point of view and leave a proof-theoretic investigation for another occasion.[8]

We begin by recalling that a chain of logical reasoning from p to q counts as trivial just in case q can be inferred from p by at most n steps of logical reasoning using the rules in $\mathcal{R}$. More generally, if Γ and Γ' are sets of sentences, we will write '$\Gamma \vdash^n_{\mathcal{R}} \Gamma'$' to say that Γ' is n-step inferable from Γ using the rules in $\mathcal{R}$. For the central results below, we assume that the relation $\vdash^n_{\mathcal{R}}$ is monotonic:

($\boldsymbol{\mathcal{R}}$-monotonicity) If $\Gamma \subseteq \Gamma'$ and $\Gamma \vdash^n_{\mathcal{R}} p$, then $\Gamma' \vdash^n_{\mathcal{R}} p$.

This ensures that logical inferences are never defeated by the addition of further assumptions.

We will base our model on the following object-language:

[7] The idea of modeling belief change in terms of transitions or relations between doxastic states is well-known from dynamic epistemic logic; see, e.g., van Ditmarsch et al. (2008) and Duc (1997).

[8] For some preliminary work in this direction, see Duc (1997) and Rasmussen (2015).

Definition 1. ***(Language)*** *The language $\mathcal{L}$ is defined in the usual inductive way from a set Φ of atomic sentences, a set $\{\neg, \wedge\}$ of connectives, a belief operator B, and a countably infinite set of dynamic operators $\langle n \rangle$ and $[n]$:*

$$p \ ::= \ \varphi \mid \neg p \mid p \wedge q \mid Bp \mid \langle n \rangle p \mid [n]p,$$

where $n = 0, 1, 2, \ldots$ and $\varphi \in \Phi$.

The operators in $\mathcal{L}$ have the following intended readings:

Bp: p is believed.
$\langle n \rangle p$: p is the case after some n steps of logical reasoning.
$[n]p$: p is the case after any n steps of logical reasoning.

For example, the sentence '$\langle n \rangle Bp$' should be interpreted as "the agent believes p after some n steps of logical reasoning" or, equivalently, that "the agent believes p after some trivial chain of logical reasoning." Likewise, the sentence '$[n]Bp$' says that "the agent believes p after any n steps of logical reasoning" or, equivalently, that "the agent believes p after any trivial chain of logical reasoning."

Next, we can define our doxastic models:

Definition 2. ***(Doxastic model)*** *Let W^P and W^I be non-empty sets (of possible and impossible worlds, respectively), and let $W = W^P \cup W^I$. A doxastic model for a single agent is a structure:*

$$M = \langle W^P, W^I, f, V \rangle,$$

where $f : W \mapsto 2^W$ is an accessibility function that assigns a set of worlds in W to each world in W, and $V : W \mapsto 2^{\mathcal{L}}$ is a function that assigns a set of sentences in $\mathcal{L}$ to each world in W.

The function f associates each world with a set of doxastically accessible worlds, where the accessible worlds may be either possible or impossible. We can think of possible worlds as complete and deductively closed entities, whereas impossible worlds need neither be complete nor subject to any

closure constraints (we will make this informal characterization precise below). The function V will play a somewhat unusual role in our framework since we will evaluate sentences for truth and falsity differently at possible and impossible worlds. More specifically, as we shall see, V is going to deliver truth-values to *atomic* sentences only at possible worlds, but is going to deliver truth-values to *all* sentences at impossible worlds.

For the central results below, we will assume the following comprehension principle for impossible worlds (cf. Nolan 1997):

> **(Comprehension principle)** For any incomplete and/or inconsistent set of sentences $\Gamma \subseteq \mathcal{L}$, there is an impossible world $w \in W^I$ such that $V(w) = \Gamma$.

This principle ensures that our models will be rich enough to represent all the different ways the world could not possibly be. Since (Non-closure) follows from (Comprehension principle), note that we effectively take an American stance on impossible worlds.

We can now go on to develop our semantics for $\mathcal{L}$. Since we will retain (Belief-impossible) as our semantics for belief, we only need to develop a semantics for the dynamic operators $\langle n \rangle$ and $[n]$. To flag the core idea behind our semantics for $\langle n \rangle$, begin by considering the special sentence type $\langle n \rangle Bp$. The semantics below will tell us that $\langle n \rangle Bp$ is true at a world w just in case p follows within n steps of logical reasoning in $\mathcal{R}$ from each doxastically accessible world from w. The truth conditions for $\langle n \rangle Bp$ will thus be weaker than those for Bp: while the semantics for Bp requires that p is *true* at all doxastically accessible worlds, the semantics for $\langle n \rangle Bp$ merely requires that p *follows within* n *steps* from the truths at each doxastically accessible world.

To spell out the details of this semantics, we first need a formal device that can tell us what follows within n steps of reasoning from the truths at a given world. The notion of an 'n-radius' of a world will play this role:

Definition 3. ***(n-radius)*** *The n-radius of a world $w \in W$ is written 'w^n'*

and is defined as follows:

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

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

That is, the n-radius of a world w is the set of n-expansions of w, where the conditions for being an n-expansion of w depends on whether w is possible or impossible. If w is possible, w is simply its own unique n-expansion. This reflects the fact that possible worlds are deductively closed entities that already verify everything that follows within n steps from what they verify (for any value of n). Given this, it might seem odd to define the n-radius of possible worlds in the first place. Indeed, one could in principle spell out the formalism below in terms of a restricted version of (n-radius) that applies to impossible worlds only. But for reasons of formal simplicity, it will be convenient to have the means to talk about the n-radius of an arbitrary world, whether possible or impossible. So, going forward, we will treat each possible world as its own unique n-expansion.

More interestingly, if w is an impossible world, w' is an n-expansion of w just in case the following three conditions are satisfied: first, w' is impossible; second, w' verifies everything that w verifies; and, third, w' does not verify anything that cannot be derived from what w verifies within n steps of reasoning in $\mathcal{R}$. Consequently, every impossible world is a member of its own n-radius, just like every possible world is a member of its own n-radius. But in contrast to possible worlds, impossible worlds need not be deductively closed, and hence they may have more than one n-expansion. For example, if $V(w) = \{p \to q, \neg q\}$, $V(w_1) = \{p \to q, \neg q, \neg p\}$, and $V(w_2) = \{p \to q, \neg q, \neg q \wedge \neg q\}$, then w_1 and w_2 are both in the 1-radius of w, assuming that $\mathcal{R}$ contains modus tollens and conjunction introduction. So, while the n-radius of a possible world is always a singleton set, this is not the case for impossible worlds.

For the semantics of $\langle n \rangle Bp$, we want to require that *at least one* n-expansion of each doxastically accessible world verifies p. So we need a

formal device that can pick out exactly one n-expansion of each doxastically accessible world:

Definition 4. ***(Choice function)*** *Let $\mathcal{C} : 2^{2^W} \mapsto 2^{2^W}$ be a function that takes a set $\mathcal{W} = \{W_1, \ldots, W_n\}$ of sets of worlds as input and returns the set $\mathcal{C}(\mathcal{W})$ of sets of worlds which results from all the ways in which exactly one element can be picked from each $W_i \in \mathcal{W}$. Each member of $\mathcal{C}(\mathcal{W})$ is called a 'choice' of $\mathcal{W}$.*

To illustrate this definition, let $\mathcal{W} = \{\{w_1, w_2\}, \{w_3\}\}$. Since each choice of $\mathcal{W}$ corresponds to one way of picking out exactly one world from each member of $\mathcal{W}$, there are two choices of $\mathcal{W}$: we can either pick $\{w_1, w_3\}$ or $\{w_2, w_3\}$. So $\mathcal{C}(\mathcal{W}) = \{\{w_1, w_3\}, \{w_2, w_3\}\}$.

We can now use (n-radius) and (Choice function) to define a relation '$\overset{n}{\sim}$' between pointed models. When 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). Informally, if (M, w) characterizes an agent's current belief state, we want to say that (M', w') is n-accessible from (M, w) just in case (M', w') characterizes a belief state that the agent can enter from (M, w) by performing up to n steps of logical reasoning. To capture this idea, (M', w') should be n-accessible from (M, w) just in case the set of doxastically accessible worlds from w in M is replaced in M' by a choice of n-expansions of the accessible worlds from w in M. So we need a device that can help us to replace a set accessible worlds with a choice of n-expansions of those worlds. The notion of an 'n-variation' of an accessibility function will serve this purpose:

Definition 5. ***(n-variation)*** *Let $M = \langle W^P, W^I, f, V\rangle$ be a model. We define $\mathcal{F}^n$ (for $n = 0, 1, 2, \ldots$) as a function from pointed models to sets of accessibility functions:*

$$\mathcal{F}^n(M, w) = \left\{ g \mid g(v) = \begin{cases} c & \text{for } v = w \\ f(v) & \text{for } v \neq w \end{cases} \right\},$$

where $c \in \mathcal{C}(\{w'^n | w' \in f(w)\})$. If g is a member of $\mathcal{F}^n(M, w)$, we say that g

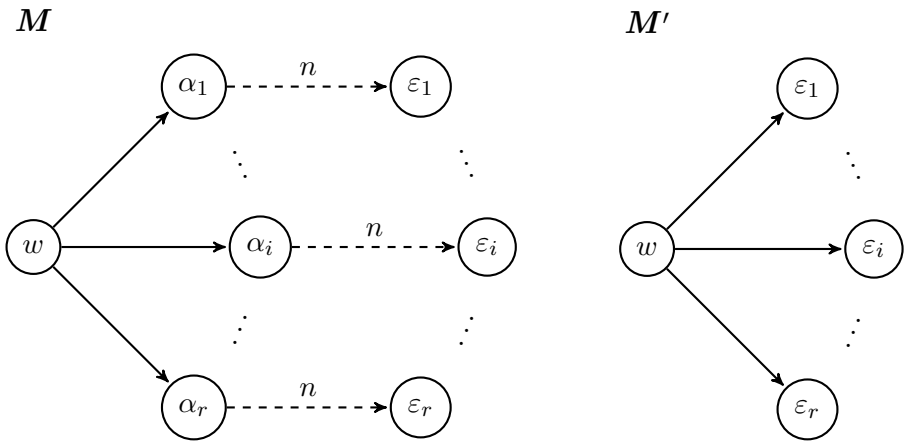

Figure 1: Illustration of (n-accessibility). A solid arrow from w to w' represents that w' is doxastically accessible from w, and a dashed arrow labelled 'n' from w to w' represents that w' is an n-expansion of w. (M', w) is n-accessible from (M, w), since the set $\{\alpha_1, \ldots, \alpha_r\}$ of accessible worlds from w in M is replaced in M' by a choice $\{\varepsilon_1, \ldots, \varepsilon_r\}$ of n-expansions of these accessible worlds.

is an 'n-variation' of f.

This definition says that an accessibility function g counts as an n-variation of the accessibility function f just in case $g(w)$ is a choice of n-expansions of $f(w)$. For example, if $f(w) = \{\alpha_1, \alpha_2\}$ and $g(w) = \{\varepsilon_1, \varepsilon_2\}$, then g is an n-variation of f, if $\varepsilon_1 \in \alpha_1^n$ and $\varepsilon_2 \in \alpha_2^n$. In general, there will be many different n-variations of f, insofar as there are many different available choices of n-expansions of $f(w)$.

We can use the notion of an n-variation to give the following definition of the n-accessibility relation '$\overset{n}{\sim}$' between pointed models:

Definition 6. ***(n-accessibility)*** *Let $M = \langle W^P, W^I, f, V\rangle$ and $M' = \langle W^{P\prime}, W^{I\prime}, f', V'\rangle$ be any two models. Then $(M, w) \overset{n}{\sim} (M', w')$ iff $w' = w$, $W' = W$, $V' = V$, and $f' \in \mathcal{F}^n(M, w)$.*

According to this definition, (M', w') is n-accessible from (M, w) just in case the set of doxastically accessible worlds from w in M is replaced in M' by a choice of n-expansions of the accessible worlds from w in M (see figure 1 for an illustration). We can think of the set of n-accessible pointed models as representing all the different ways in which an agent's doxastic state can change as a result of performing up to n steps of logical reasoning.

Given (n-accessibility), we can now complete our semantics:

Definition 7. ***(Satisfaction)*** *Sentences are evaluated for truth and falsity on pointed models, where a pointed model is a pair consisting of a model and a world. We write '$M, w \vDash p$' to say that p is true (or satisfied) at w in M, and we write '$M, w \dashv p$' to say that p is false (or dissatisfied) at w in M.*

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

(P1) $M, w \vDash \varphi$ *iff* $\varphi \in V(w)$, *where* $\varphi \in \Phi$.
(P2) $M, w \vDash \neg p$ *iff* $M, w \nvDash p$.
(P3) $M, w \vDash p \wedge q$ *iff* $M, w \vDash p$ *and* $M, w \vDash q$.
(P4) $M, w \vDash Bp$ *iff* $M, w' \vDash p$ *for all* $w' \in f(w)$.
(P5) $M, w \vDash \langle n \rangle p$ *iff* $M', w' \vDash p$ *for some* $(M', w') : (M, w) \overset{n}{\sim} (M', w')$.
(P6) $M, w \vDash [n]p$ *iff* $M', w' \vDash p$ *for all* $(M', w') : (M, w) \overset{n}{\sim} (M', w')$.
(P7) $M, w \dashv p$ *iff* $M, w \nvDash p$.

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

(I1) $M, w \vDash p$ *iff* $p \in V(w)$.
(I2) $M, w \dashv p$ *iff* $\neg p \in V(w)$.

A few comments about this semantics are in order. First, note that it holds for both possible and impossible worlds that p is false just in case $\neg p$ is true. However, at impossible worlds, p may be *neither* true nor false (that is, p may constitute a 'truth-value gap'), and p may be *both* true and false (that is, p may constitute a 'truth-value glut'). By contrast, possible worlds never contain any truth-value gaps or gluts.

Second, (P4) is simply a formalization of (Belief-impossible). Since we have not imposed any logical constraints on impossible worlds, this means that agents may believe both a proposition and its negation. This obviously raises some questions about how agents should react when they discover that they have contradictory beliefs. While we do not want to enter a discussion of this question here, it is worth noting that our framework is compatible with a number of different answers. Classically inclined philosophers may supply $\mathcal{R}$ with an 'explosion rule' that allows agents to infer any proposition from a contradiction. Non-classically inclined philosophers may instead specify $\mathcal{R}$ in

accordance with a suitable paraconsistent logic that deals with contradictions in a non-explosive way. So our proposal is general enough to accommodate various views of how one might reason rationally with contradictions.[9]

Third, (P5) says that $\langle n \rangle p$ is true at w in M just in case p is satisfied by at least one n-accessible pointed model from (M, w). In particular, $\langle n \rangle Bp$ is satisfied by (M, w) just in case Bp is satisfied by some n-accessible pointed model from (M, w) (see figure 2 for an illustration). Hence (P5) captures the central idea that an agent can come to believe p after a trivial chain of logical reasoning whenever there is a transition from the agent's doxastic state through n applications of the rules in $\mathcal{R}$ to a state in which she believes p. Likewise, (P6) gives the conditions under which an agent believes p after *any* trivial chain of logical reasoning. Note that the conditions under which an agent believes p after *any* trivial chain of logical reasoning are the same as the conditions under which the agent believes p. While this makes the semantics for $[n]Bp$ somewhat uninteresting, it captures the intended idea that among all the possible chains of trivial reasoning that an agent can perform is the chain that merely infers what is already believed.

Finally, for the central results below, we define validity with respect to possible worlds only. That is, if Γ is a set of sentences and q is a sentence, $\Gamma \vDash q$ just in case, for all models, q is true at all possible worlds that verify all sentences in Γ.

With our semantics on the table, we can now establish the main result of the paper (see figure 3 for a diagrammatic representation of the proof):

Theorem 1. *If* $\{p_1, \ldots, p_k\} \vdash^n_{\mathcal{R}} q$, *then* $\{\langle m_1 \rangle Bp_1, \ldots, \langle m_k \rangle Bp_k\} \vDash \langle \omega + n \rangle Bq$, *where* $\omega = m_1 + \cdots + m_k$.

Proof. Let $M = \langle W^P, W^I, f, V \rangle$ be any model, and suppose $\{p_1, \ldots, p_k\} \vdash^n_{\mathcal{R}} q$ and $M, w \vDash \langle m_i \rangle Bp_i$, for $1 \leq i \leq k$, where $w \in W^P$. We must show that $M, w \vDash \langle \omega + n \rangle Bq$, where $\omega = m_1 + \cdots + m_k$. By (P5), $M_i, w_i \vDash Bp_i$ for some $(M_i, w_i) : (M, w) \overset{m_i}{\sim} (M_i, w_i)$, for $1 \leq i \leq k$. By (n-accessibility), $M_i, w \vDash Bp_i$ for some $M_i = \langle W^P, W^I, f_i, V \rangle$, where $f_i \in \mathcal{F}^{m_i}(M, w)$. By ($n$-variation), $f_i(w) = c_i$ for some choice $c_i \in \mathcal{C}(\{v^{m_i} | v \in f(w)\})$. By (P4), $M_i, \varepsilon \vDash p_i$ for all $\varepsilon \in c_i$.

[9] For discussions of paraconsistent reasoning, see Andreas and Verdee (2016).

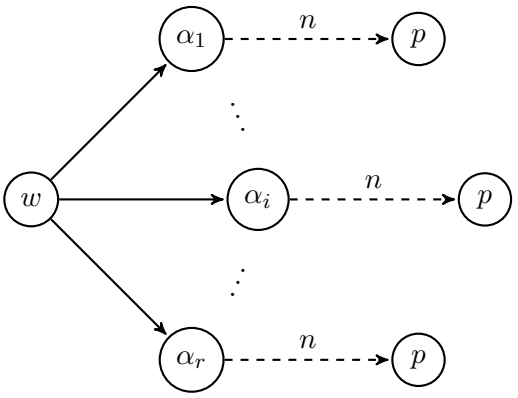

Figure 2: Illustration of the semantics for $\langle n\rangle Bp$. $\langle n\rangle Bp$ is true at w since p follows within n steps of reasoning from the truths at each accessible world from w. The ps in the n-expansions of the accessible worlds $\{\alpha_1, ..., \alpha_r\}$ are not names, but indicate that p is true at the relevant worlds.

By ($\mathcal{R}$-monotonocity), there is a choice $c' \in \mathcal{C}(\{v^\omega | v \in f(w)\})$ such that if $M' = \langle W^P, W^I, f', V\rangle$, where $f'(w) = c'$, then $M', \varepsilon' \vDash p_i$ for all $\varepsilon' \in c'$. Given (Comprehension principle) and the assumption that $\{p_1, \ldots, p_k\} \vdash^n_{\mathcal{R}} q$, there will be a choice $c'' \in \mathcal{C}(\{v^{\omega+n} | v \in f(w)\})$ such that if $M'' = \langle W^P, W^I, f'', V\rangle$, where $f''(w) = c''$, then $M'', \varepsilon'' \vDash q$ for all $\varepsilon'' \in c''$. By (P4), $M'', w \vDash Bq$. By (n-variation), $f'' \in \mathcal{F}^{\omega+n}(M, w)$. By ($n$-accessibility), $(M, w) \overset{\omega+n}{\sim} (M'', w)$. So, for some model $(M'', w'') : (M, w) \overset{\omega+n}{\sim} (M'', w'')$, $M'', w'' \vDash Bq$. Thus, by (P5), $M, w \vDash \langle \omega + n\rangle Bq$. □

Theorem 1 says that if a conclusion q follows within n steps in $\mathcal{R}$ from a set of premises $\{p_1, ..., p_k\}$, and the agent believes the ith premise after some m_i steps of reasoning, for $1 \leq i \leq k$, then the agent believes q after some $n + m_1 + m_2 + \cdots + m_k$ steps of reasoning. For instance, if an agent believes p after 1 step and believes $p \to q$ after 2 steps, then she can apply modus ponens once to infer q and hence come to believe q after $1 + 2 + 1 = 4$ steps of reasoning.

The following result is a special case of Theorem 1:

Corollary 1. ***(n-distribution)*** *If $\{p_1, \ldots, p_k\} \vdash^n_{\mathcal{R}} q$, then $\{Bp_1, \ldots, Bp_k\} \vDash \langle n\rangle Bq$.*

According to (n-distribution), if a conclusion q follows within n steps of

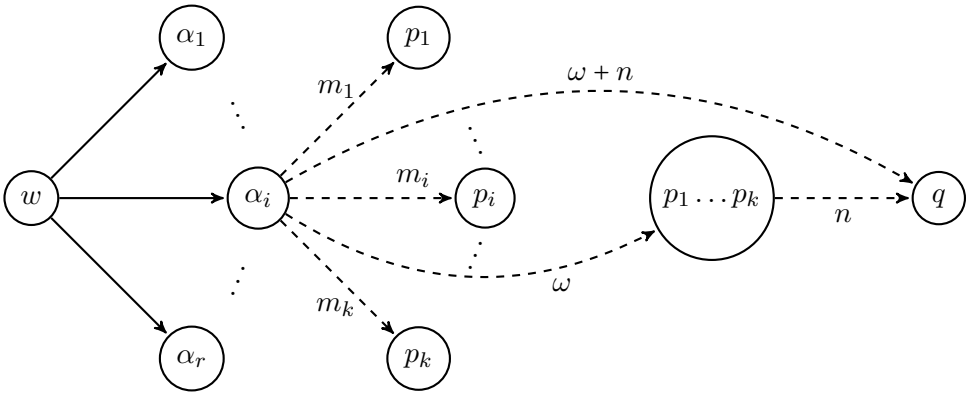

Figure 3: Diagrammatic representation of the proof of Theorem 1. As in Figure 2, the ps and qs are not names of worlds, but indicate that p is true at the relevant worlds.

reasoning from a set $\{p_1, \dots, p_k\}$ of premises, and the agent believes all of the premises, then the agent believes q after some n steps of reasoning. We can understand (n-distribution) as a dynamic counterpart of the distribution axiom **K** from standard doxastic logic:

K $(Bp \wedge B(p \to q)) \to Bq$.

While **K** says that beliefs are closed under believed entailment, (n-distribution) carries no such commitment. It merely says that agents have the ability to immediately form a belief in any proposition that follows within n steps of reasoning from what they already believe.

As a special case of (n-distribution), we get:

Corollary 2. ***(n-necessitation)*** *If* $\vdash^n_{\mathcal{R}} p$, *then* $\vDash \langle n \rangle Bp$.

According to (n-necessitation), if p follows from the empty set within n steps of reasoning, then the agent believes p after some chain of n-step reasoning. We can see (n-necessitation) as a dynamic counterpart of the necessitation rule from standard doxastic logic:

Nec *If* $\vdash p$, *then* $\vdash Bp$.

While **Nec** entails that agents believe all logical truths, (n-necessitation) carries no such commitment. It merely says that agents have the ability to

immediately form a belief in any proposition that is n-step inferable in $\mathcal{R}$ from the empty set.

We are now in a position to show how our framework successfully models agents who are both logically non-omniscient and logically competent. To see how logical omniscience is avoided, suppose Bp is true at w, for some $w \in W^P$, and consider any q that is logically entailed by p. By (P4), p is true at all doxastically accessible worlds for the agent. However, since (P4) quantifies over both possible and impossible worlds, q need not be true at all these doxastically accessible worlds. So Bq need not be true at w, which means that logical omniscience is avoided.

To see how logical competence is secured, suppose Bp is true at w, for some $w \in W^P$, and consider any q that follows from p within n steps of logical reasoning. By (n-distribution), $\langle n \rangle Bq$ is true at w. So it follows that the agent has an ability to immediately form a belief in any trivial logical consequence of what she already believes.[10] Hence our model explains why the agent need never miss out on anything trivial: she can always make any trivial consequence q of her beliefs count in reasoning and action. As such, (n-distribution) helps us explain how logically competent agents pass the test for logical competence. Suppose we ask the agent whether q is the case. While Bq need not be true at w—a desirable result in light of the collapse result—the fact that $\langle n \rangle Bq$ is true at w tells us that the agent can immediately form a belief in q and, as a result, immediately answer "yes" to the question whether q.

By varying the value of n, our framework allows us to model a whole *spectrum* of agents with different levels of cognitive resources. In the limit where $n = 0$, agents have no cognitive resources available, and $\langle n \rangle Bq$ will be false for any q that follows from the agent's beliefs (assuming, as above, that nothing is 0-step inferable in $\mathcal{R}$). In the opposite limit, where n goes towards infinity, agents have unlimited cognitive resources available, and,

[10]Note, though, that our semantics for $\langle n \rangle Bp$ does not commit us to claiming that agents have an ability to infer all logical consequences of what they believe. In general, if Bp is true at w, and q follows from p in more than n steps of logical reasoning, it is easily seen that $\langle n \rangle Bq$ is false at w. So there are plenty of non-trivial logical consequences that lie beyond the cognitive reach of agents.

by (Corollary 1), $\langle n\rangle Bq$ will be true for any q that follows logically from the agent's beliefs. In-between these extremes, we find agents with different intermediate levels of cognitive resources. For such agents, $\langle n\rangle Bq$ will be true for *some*, but not *all* logical consequences q of what they believe.

It is worth noting that if we identify $\mathcal{R}$ with a complete proof system for propositional logic, and if we let n approach infinity, we get the following pleasant symmetry between (n-distribution) and the axiom **K**: for any classical consequence q of an agent's beliefs, $\langle n\rangle Bq$ will be true in our logic just in case Bq is true in standard doxastic logic, since q will be n-step inferable in $\mathcal{R}$ whenever q is inferable (simpliciter) in propositional logic. As such, our framework is even able to model agents who are logically omniscient in the sense that they are able to tease out all logical consequences of what they believe, including all logical truths.[11]

4 Jago on logical omniscience

Mark Jago (2013; 2014) has recently proposed an impossible worlds model of belief that, much like ours, promises to steer clear of both logical omniscience and logical incompetence. In this section, we argue that our proposal has a number of advantages over Jago's.

In response to the collapse result, Jago rightly concludes that agents must either be logically omniscient or fail to believe at least some trivial consequences of what they believe—he refers to this dilemma as "the problem of rational knowledge" (Jago 2013, p. 1152). On pain of logical omniscience, Jago accepts that it must be "possible to fail to know or believe trivial truths (and, more generally, trivial consequences of one's beliefs)" (Jago 2014, p. 243). As he notes,

> [t]here is, to be sure, something counter-intuitive in this result. If an agent fails to know or believe that $A \vee \neg A$ for some 'A', then her epistemic or doxastic state misses out on something trivial. Similarly, if an agent knows or believes that such-and-such, from which 'A' triv-

[11] Thanks to an anonymous reviewer for drawing our attention to this point.

> ially follows and yet she does not know or believe that A, then again her epistemic or doxastic state misses out on something trivial. I'll call such cases, in which an agent fails to know (or believe) some trivial consequence of what she knows (or believes), *epistemic oversights*. [...] Epistemic oversights are bizarre, but we know they must exist. For every logically non-omniscient agent, there is some knowledge she has which trivially entails something she does not know. Otherwise, her knowledge would be closed under all trivial inference rules and hence deductively closed. (Jago 2014, p. 206)

Epistemic oversights are "counter-intuitive" or "bizarre", according to Jago, because an agent who suffers from an epistemic oversight seems to "miss out on something trivial" and so seems irrational or logically incompetent.[12] To avoid treating agents as logically incompetent, Jago suggests that epistemic oversights must always be *indeterminate*: we can never rationally assert that an agent has a particular epistemic oversight. For if we do, we thereby treat the agent as logically incompetent (Jago 2013, p. 1152).

To ensure that epistemic oversights are always indeterminate, Jago develops epistemic models that satisfy the following principle:

> **(Indeterminacy)** [I]f 'A' is a trivial consequence of what an agent i knows, then it's never determinate that i fails to know that A. (Jago 2013, pp. 1166–1167)

According to Jago, since it is never rational to assert what is indeterminate, (Indeterminacy) prevents us from attributing particular epistemic oversights to agents and thereby prevents us from treating them as logically incompetent.

While we agree with Jago that logically non-omniscient agents must suffer from epistemic oversights—the collapse result shows that much—we believe that there are several problems involved in using (Indeterminacy) to avoid logical incompetence. Below we raise five such problems.

[12]Since Jago's use of "rationality" and our use of "logical competence" are supposed to do more or less the same work, we will use these terms interchangeably in what follows.

1. (Indeterminacy) is dispensable. We do not agree with Jago's claim that agents who suffer from determinate epistemic oversights must be logically incompetent. For even such agents can have the ability to infer the trivial consequences of what they believe. And such agents, as we have shown, need not "miss out on anything trivial": they can always make any trivial consequence of what they believe count in reasoning and action. Hence (Indeterminacy) is dispensable for modeling agents who are both logically non-omniscient and logically competent.

2. (Indeterminacy) lacks independent motivation. Jago might grant that (Indeterminacy) is dispensable but hold that it nevertheless does the required job: it allows us to treat agents as logically competent. In point 4 below we argue that (Indeterminacy) in fact cannot do this job, but even if it could, we can ask for some *independent* reasons to accept (Indeterminacy)—reasons, that is, that do not derive from the need to avoid logical incompetence. Jago suggests that (Indeterminacy) is motivated by a structural similarity between the problem of rational knowledge and the sorites paradox:

> The problem of rational knowledge can be formulated in terms of a step-by-step deduction D from premises C which the agent in question clearly knows, to a conclusion 'A' that the agent clearly does not know. By assumption, not every step of reasoning in D preserves the agent's knowledge (since we eventually arrive at a conclusion the agent does not know). Yet any attempt to say precisely at which point in the deduction the agent's knowledge gives out is doomed to failure. [...] Formulating the problem in this way brings out its structural similarity with the sorites paradox. In this case, the principle that rational agents know the trivial consequences of what they know plays the role that tolerance conditionals play in the sorites. The tolerance conditionals for 'red', for example, say that (in a sorites series of colour samples), if sample n is red then so is sample $n + 1$. Clearly, not all such conditionals are true; but we cannot say or discover which is false. (Jago 2013, p. 1155)

However, the alleged structural similarity between the problem of rational knowledge and the sorites paradox *presupposes* rather than independently

motivates (Indeterminacy). The similarity is brought out by the claim that "any attempt to say precisely at which point in the deduction the agent's knowledge gives out is doomed to failure" and the claim that "any attempt to say precisely which tolerance conditional fails to hold is doomed to failure." But the former claim is plausible only if we already accept (Indeterminacy). For if we do not, nothing prevents us from pointing out precisely at which point in a deduction the agent's knowledge gives out. So there is only a structural similarity between the problem of rational knowledge and the sorites paradox if (Indeterminacy) is already assumed. As such, Jago has offered no independent reasons to accept (Indeterminacy).

3. (Indeterminacy) faces potential counterexamples. Consider a simple, logically non-omniscient artificial agent who believes just the propositions $p_1, p_1 \to p_2, p_2 \to p_3, \ldots, p_{k-1} \to p_k$, and suppose the agent can apply modus ponens only once. Given this, there is exactly *one* trivial consequence of the agent's beliefs, namely p_2. Since the agent is logically non-omniscient, the collapse result shows that the agent must suffer from at least *one* epistemic oversight. And since p_2 is the *only* trivial consequence of the agent's beliefs, it follows that the agent cannot believe p_2. So we determinately know that the agent fails to believe p_2. But this runs counter to (Indeterminacy): logically competent agents may well suffer from determinate epistemic oversights.

4. (Indeterminacy) lacks explanatory power. Even if we set aside the problems above, we doubt that (Indeterminacy) can adequately capture the sense in which agents are logically competent. Consider again the simple agent from point 3. Due to her logical competence, the agent can be assumed to pass the test for logical competence: she is able to answer "yes" immediately when asked whether p_2 is true. On the face of it, this seems hard to reconcile with the fact that the agent does not believe p_2. What explains why the agent is able to give the correct answer when she does not believe the answer? Our model provides a straightforward explanation: since the agent believes p_1 and $p_1 \to p_2$, (n-distribution) tells us that the agent can come to believe p_2 after one step of reasoning—assuming that $\mathcal{R}$ contains modus ponens. Since, by description, the agent can immediately apply modus ponens once, we know that she can immediately infer p_2 and

hence immediately enter a belief state that contains p_2. This fact explains the agent's ability to immediately answer "yes" when asked about p_2. By contrast, (Indeterminacy) merely tells us that the agent's lack of belief in p_2 is *indeterminate*. We have already argued against this claim, but even if we grant it, we are left without an explanation of why the agent is able to assent to p_2 despite not believing it.

Of course, the agent above is quite simple: her reasoning mechanism is highly incomplete, her computational resources are very limited, and she only has a small number of beliefs. One might wonder how (Indeterminacy) fares in more complex cases where there are more than one trivial consequence of an agent's beliefs. Consider a logically non-omniscient, yet logically competent agent such that there are m different trivial consequences $q_1, \ldots, q_m$ of the agent's beliefs. Suppose we were to ask this agent a series of questions concerning these trivial consequences: first, we ask whether q_1 is true, then whether q_2 is true, and so on. Due to her logical competence, we can assume that she is able to immediately answer "yes" to each such question. Due to her logical non-omniscience, however, we know that she must suffer from at least one epistemic oversight. So again we face an explanatory challenge. How do we explain the agent's ability to answer "yes" to each question when she fails to believe at least one of the answers? As above, our model gives a straightforward answer. When asked about q_i, (n-distribution) says that the agent can immediately infer q_i, and, as such, immediately enter a belief state that contains q_i. This fact explains the agent's ability to immediately answer "yes" when asked about q_i. (Indeterminacy), by contrast, does not help us explain the agent's logically competent behavior. It prevents us from rationally ascribing any particular epistemic oversight to the agent, but it does not explain how the agent can answer each question correctly despite failing to believe at least one of the answers.

5. Elusive justification for belief ascriptions. The case above also points out a different tension in Jago's proposal. Suppose the logically competent agent from above in fact answers "yes" to the question concerning q_1. This seems to give us good (albeit fallible) justification for saying that the agent believes q_1. By repeating this procedure for $q_2, \ldots, q_m$, we acquire good jus-

tification for saying that the agent believes each q_i. Yet, due to her logical non-omniscience, we determinately know that the agent fails to believe at least one q_i. So we seem to end up with good justification for the claim that the agent believes each of $q_1, \ldots, q_m$ and yet fails to believe at least one of $q_1, \ldots, q_m$. To avoid this absurdity, it seems that Jago must hold that we somehow lose justification for at least one of the belief ascriptions during the process of questioning. It is unclear to us why our justification for belief ascriptions should be "elusive" in this way. But if it is, we should expect an independently plausible story of why our justification is lost during the process of questioning—a story that Jago does not provide.

In sum, since our dynamic framework does not appeal to (Indeterminacy), it avoids the problems and tensions that Jago faces. For this reason, our framework seems to provide a superior model of logically non-omniscient, yet logically competent agents.

5 Summary

We began this paper by motivating the idea that a proper solution to the problem of logical omniscience should allow us to model agents who are both logically non-omniscient and logically competent. We then argued that standard versions of the impossible worlds framework cannot model such agents. Instead, we proposed to dynamize the impossible worlds framework in a way that allows us to capture not only what agents believe, but also what they believe after having performed a chain of logical reasoning. Finally, we developed the formal details of a dynamic impossible worlds model of belief, and showed that it successfully models agents who are logically non-omniscient, yet logically competent.

Acknowledgments. An earlier version of this paper was presented in the Research Unit for Epistemology and Metaphysics at Aarhus University. We thank the audience on that occasion for valuable feedback. Thanks also to Francesco Berto, Mark Jago, and the reviewers from *Journal of Philosophical Logic* for very helpful comments and criticism.

References

Andreas, H. and P. Verdée (Eds.) (2016). *Logical Studies of Paraconsistent Reasoning in Science and Mathematics*. Springer.

Berto, F. (2013). "Impossible Worlds". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2013.

Bjerring, J.C. (2013). "Impossible Worlds and Logical Omniscience: An Impossibility Result". In: *Synthese* 190.13, pp. 2505–2524.

— (2014). "Problems in Epistemic Space". In: *Journal of Philosophical Logic* 43.1, pp. 153–70.

Bjerring, J.C. and W. Schwarz (2017). "Granularity Problems". In: *Philosophical Quarterly* 67.266, pp. 22–37.

Chalmers, D. (2011). "The Nature of Epistemic Space". In: *Epistemic Modality*. Ed. by A. Egan and B. Weatherson. New York: Oxford University Press, pp. 60–107.

Cherniak, C. (1986). *Minimal Rationality*. MIT Press.

Cresswell, M. (1970). "Classical Intensional Logic". In: *Theoria* 36, pp. 347–372.

— (1972). "Intensional Logics and Logical Truth". In: *Journal of Philosophical Logic* 1, pp. 2–15.

— (1973). *Logics and Languages*. Methuen and Co.

Ditmarsch, W., W. Hoek, and B. Kooi (2008). *Dynamic Epistemic Logic*. Springer.

Drapkin, J. and D. Perlis (1986). "A Preliminary Excursion into Step-Logics". In: *Proceedings of the SIGART International Symposium on Methodologies for Intelligent Systems*. Ed. by C. Ghidini, P. Giodini, and W. van der Hoek, pp. 262–269.

— (1990). "Reasoning Situated in Time I: Basic Concepts". In: *Journal of Experimental and Theoretical Artificial Intelligence* 2, pp. 75–98.

Duc, H.N. (1997). "Reasoning about rational, but not logically omniscient, agents". In: *Journal of Logic and Computation* 7.5, pp. 633–648.

Elgot-Drapkin, J. et al. (1999). *Active Logics: A Unified Formal Approach to Episodic Reasoning*. Tech. rep. University of Maryland.

Fagin, R., J. Halpern, and M. Vardi (1995). "A Nonstandard Approach to the Logical Omniscience Problem". In: *Artificial Intelligence* 79, pp. 203–240.

Fagin, R. et al. (1995). *Reasoning About Knowledge*. MIT Press.

Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Two Notions*. 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*, pp. 1–18.

— (2014). *The Impossible: An Essay on Hyperintensionality*. Oxford University Press.

Lakemeyer, G. (1987). "Tractable Meta-Reasoning in Propositional Logics of Belief". In: *Tenth International Joint Conference on Artificial Intelligence*, pp. 198–202.

Levesque, H.J. (1984). "A Logic of Implicit and Explicit Belief". In: *National Conference on Artificial Intelligence*, pp. 198–202.

Nolan, D. (1997). "Impossible Worlds: A Modest aApproach". In: *Notre Dame Journal of Formal Logic* 38, pp. 535–72.

Rantala, V. (1982). "Impossible Worlds Semantics and Logical Omniscience". In: *Acta Philosophica Fennica* 35, pp. 106–115.

Rasmussen, M. Skipper (2015). "Dynamic Epistemic Logic and Logical Omniscience". In: *Logic and Logical Philosophy* 24, pp. 377–99.

Wansing, H. (1990). "A General Possible Worlds Framework for Reasoning about Knowledge and Belief". In: *Studia Logica: An International Journal for Symbolic Logic* 49, pp. 523–39.

Weirich, P. (2004). *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. OUP.

Wright, G. H. von (1951). *An Essay in Modal Logic*. Amsterdam: North-Holland Pub. Co.