Fragmentation, metalinguistic ignorance, and logical omniscience

Jens Christian Bjerring[1] · Weng Hong Tang[2]

Abstract

To reconcile the standard possible worlds model of knowledge with the intuition that ordinary agents fall far short of logical omniscience, *a Stalnakerian strategy* appeals to two components. The first is the idea that mathematical and logical knowledge is at bottom metalinguistic knowledge. The second is the idea that non-ideal minds are often fragmented. In this paper, we investigate this Stalnakerian reconciliation strategy and argue, ultimately, that it fails. We are not the first to complain about the Stalnakerian strategy. But in contrast to existing complaints, we want to cause trouble for the strategy directly on its home turf. That is, we will advance our objection while granting both the plausibility of the fragmentation component—save for an extreme version of it—and that of the metalinguistic component. Once our central objection to the Stalnakerian strategy is in place, we will show how it negatively affects Adam Elga and Augustín Rayo's recent attempt to apply the Stalnakerian strategy in the context of Bayesian decision theory.

Keywords Logical omniscience · Fragmentation · Metalinguistic ignorance · Bayesian decision theory

1 Introduction

We are used to conducting philosophical thinking in terms of possible worlds. We know what it means to *analyze* various modal notions in terms of quantification over possible worlds. To believe *P*, for instance, is to have *P* be true at all possible worlds

✉ Jens Christian Bjerring
jcbjerring@gmail.com

Weng Hong Tang
wenghong@nus.edu.sg

[1] Department of Philosophy, School of Culture and Society, Aarhus University, Jens Chr. Skous Vej 7, 8000 Aarhus C, Denmark

[2] Department of Philosophy, Faculty of Arts and Social Sciences, National University of Singapore, 3 Arts Link, Block AS3, Level 5, Singapore 117570, Singapore

that are doxastically possible for the cognizer. We also know what it means to *identify* central philosophical concepts such as propositions and mental and linguistic content with sets of possible worlds. The proposition that *P*, for instance, is often identified with the set of possible worlds where *P* is true.

Thinking philosophically in terms of possible worlds is attractive. First, the possible worlds framework is formally elegant: the Boolean structure underlying the framework is mathematically and logically very well-behaved. Second, there is no alternative general framework that has received nearly the same amount of scrutiny, motivation and development as the possible worlds framework; just think about the status of, say, situation semantics, truthmaker semantics, or impossible worlds semantics. Third, and perhaps most importantly, the possible worlds framework seems to be strongly motivated by certain philosophical views about the nature of language, information, and the mind. Concerning the latter, we will focus on the philosophy of mind, and, in particular, on Robert Stalnaker's *causal-pragmatic* picture of the nature of belief.[1]

On Stalnaker's picture, "[w]e believe that *P* just because we are in a state that, under optimal conditions, we are in only if *P*, and under optimal conditions, we are in that state because *P*, or because of something that entails *P*." (Stalnaker, 1987, p. 18.) If the state of believing in this way systematically depends on the environment being in certain specific states, it follows immediately that beliefs are closed under logical consequence. When *P* logically entails *Q*, every environmental state in which P is a state in which Q, and hence, on the causal-pragmatic picture, every state of believing P is a state of believing Q. Moreover, when P is necessary, any state of the environment will be a state in which P, and when P is impossible, no state of the environment will be a state in which P. Accordingly, the necessary is always believed, whereas the impossible is never believed.

According to Stalnaker, if we think of beliefs in this causal way, it is natural to identify belief content with possible worlds propositions. When P logically entails Q, the set of possible worlds that verify P is a subset of the set of worlds that verify Q. So if the objects of beliefs are possible worlds propositions, any agent who stands in the belief relation to P automatically stands in the belief relation to Q—as required by the causal-pragmatic picture of belief. Moreover, when P is necessary, the set of worlds that verify P is the universal set, and when P is impossible, the set of worlds that verify P is empty. So if the objects of beliefs are possible worlds propositions, every agent stands in the belief relation to P when P is necessary, and no agent stands in the belief relation to P when P is impossible—again, as required by the causal-pragmatic picture. So if this approach to the nature of belief is on the right track, we seem to have a strong motivation for subscribing to the possible worlds individuation of propositions. Indeed, as Stalnaker puts it, the causal-pragmatic approach shows that "the possible worlds analysis of propositions … [has] a deeper philosophical motivation than has sometimes been supposed" (Stalnaker,

---

[1] The focus here is on the *causal* part of the causal-pragmatic picture. The *pragmatic* part has to do with how certain representational states get to count as beliefs—rather than, say, imaginings or hopes—in virtue of being closely connected with desires and actions.

1987, p. 24). A motivation, that is, which is philosophically deeper than, say, the mere formal and mathematical elegance of the framework.

Yet, despite being well-motivated and formally elegant, the possible worlds account of propositional content is not without serious problems. Central here is the *problem of logical omniscience*. Whenever Q follows logically from P, as we have seen, every agent who believes—or knows—the possible worlds proposition that P believes the possible worlds proposition that Q, irrespective of how complicated and logically complex the logical entailment from P to Q is. As a special case: since any logical truth follows logically from the empty set, every agent believes—or knows—every logical truth. But agents of such logical sophistication are *logically omniscient*.

As Stalnaker himself acknowledges, logical omniscience is a *problem* because it conflicts with clear intuitions about the cognitive and computational capacities of ordinary people. Intuitively, it is simply *not* the case that we believe Q just because we believe P and because P logically entails Q. For instance, first year arithmetic students can happily believe the Peano axioms without also believing that Fermat's Last Theorem is true, although the axioms (plausibly) entail the theorem. Also, intuitively, it is simply *not* the case that we believe P just because P is a necessary truth. Have your pick of any sufficiently complex truth of mathematics or logic, and chances are that we will not believe it. So even if one has good philosophical reasons to adopt the possible worlds account of propositional content, one still needs to explain the striking fact that it at least *appears* to us ordinary agents *as if* we fall short—indeed, far short—of logical omniscience.

In this paper, we investigate a *Stalnakerian strategy* for reconciling the possible worlds account of propositional content with our intuitions about our non-omniscience. Ultimately, we will argue, there is reason to be doubtful about the strategy. To be sure, we are not the first to complain about the Stalnakerian strategy.[2] But our objection differs from the usual ones in the sense that it has bite even when we grant the Stalnakerian all the conceptual and formal tools that he or she wields. If our central arguments are successful, it is a cause for worry not just for Stalnaker, but also for those, such as Lewis (1982, 1986), Braddon-Mitchell and Jackson (1996), Greco (2021), and Elga and Rayo (2022) who explicitly address logical omniscience along Stalnakerian lines. To justify this latter claim, we will show in detail how our arguments cause trouble for Elga and Rayo's recent attempt to apply the Stalnakerian strategy in pursuit of a "fragmented decision theory" suitable for logically non-omniscient agents.

Here is how we proceed. In Sect. 2, we recap briefly the Stalnakerian reconciliation strategy. In Sect. 3, we unfold and discuss our central objection to the strategy. In Sect. 4, we show how the objection applies to Elga and Rayo's Stalnakerian approach to decision theory. In Sect. 5, we conclude.

---

[2] For critical discussions of the Stalnakerian strategy, see for example Borgoni et al. (2021), Jago (2014a), Field (2001), Forbes (1989), Robbins (2004), and Stanley (2010). For alternative proposals on how to deal with the problem of logical omniscience, see, for example, Berto and Jago (2019), Bjerring and Skipper (2019), Jago (2014a), and Dogramaci (2018).

2 The Stalnakerian strategy

There are two central components in the Stalnakerian strategy for reconciling the possible worlds account of propositions with our intuitions about our non-omniscience. The first *metalinguistic* component appeals to the idea that mathematical and logical knowledge is at bottom metalinguistic knowledge, whereas the second *fragmentation* component appeals to the idea that the non-ideal mind is fragmented. Let us briefly consider each in turn.

It seems obvious that computationally bounded agents like us often fail to believe certain necessary truths like those expressed by complex tautological formulas such as '$\neg(S_1 \rightarrow S_2) \rightarrow \neg((\neg S_1 \rightarrow \neg S_3) \rightarrow \neg(\neg S_1 \rightarrow \neg S_2))$', where '$S_1$', '$S_2$', and '$S_3$', here, as elsewhere, stand for sentences in English. Yet, if we identify propositional content with possible worlds propositions, this cannot happen: the universal proposition is always known. Rather, according to Stalnaker, when it comes to knowledge of mathematics and logic, what we often fail to know is the contingent metalinguistic proposition *that* a certain string of symbols expresses the necessary proposition. More generally,

> "the apparent failure to see that a proposition is necessarily true, or that propositions are necessarily equivalent, is to be explained as the failure to see what propositions are expressed by the expressions in question." (Stalnaker, 1987, p. 84)

So, in the case at hand, what an agent may fail to believe is the *contingent* proposition that the string '$\neg(S_1 \rightarrow S_2) \rightarrow \neg((\neg S_1 \rightarrow \neg S_3) \rightarrow \neg(\neg S_1 \rightarrow \neg S_2))$', which standardly expresses the necessary proposition, in fact does so.

The metalinguistic strategy is intended to explain away apparent failures of knowing the necessary proposition, whether this knowledge is obtained in a way that is usually thought to be a priori—say, via reasoning—or in a way that is usually thought to be a posteriori—say, via testimony. For current purposes, we will restrict our attention to supposed cases of a priori logical and mathematical knowledge where the metalinguistic strategy is arguably most promising.[3] To be clear, though, we should not be interpreted as endorsing the metalinguistic strategy. Rather, we want to argue that the Stalnakerian strategy faces a serious objection even when the plausibility of the metalinguistic component is taken for granted.

On its own, however, the metalinguistic component is inadequate. As Stalnaker notes, metalinguistic ignorance cannot help us explain how agents can seemingly fail to know the logical consequences of what they already know:

> "[C]onsider a particular axiomatic formulation of first order logic with which, suppose, I am familiar. While it is a contingent fact that each axiom sentence expresses a necessary truth (however the descriptive terms are interpreted), this is a contingent truth which I know to be a fact. It may also be only contingently true that the rules of inference of the system when applied to sentences

---

[3] For motivation of this thought, see Stalnaker (1990), and for a critique, see Field (2001).

which express necessary truths always yield sentences which express necessary truths, but this fact too is known to me. Now consider any sentence of the system in question which happens to be a theorem. It is only a contingent truth that that sentence expresses a necessary truth, but this contingent fact follows deductively from propositions that I know to be true. Hence if my knowledge is deductively closed, as seems to be implied by the conception of states of knowledge and belief that I have been defending, it follows that I know of every theorem sentence of the system in question that it expresses a necessary truth. But of course I know no such thing." (Stalnaker, 1987, p. 76)

Let us illustrate the idea behind Stalnaker's thinking with a simple example.

Say, as above, that an agent knows P just in case P is true at all worlds that are epistemically possible for the agent. Suppose then that the following three propositions are all true at all possible worlds that are epistemically possible for the agent:

the proposition that *the sentence 'it rains' expresses a truth*;
the proposition that *the sentence 'if it rains, then AC Milan's game will be canceled' expresses a truth*; and
the proposition that *if 'A' and 'If A, then B' both express a truth, then 'B' expresses a truth*.

We can think of the latter proposition as encoding information about the inference rule modus ponens, and we can think of the variables 'A' and 'B' as placeholders for arbitrary sentences in English. When an agent knows modus ponens in this sense—in this *schema* sense as we shall say later—he thus knows that he needs to instantiate the variables 'A' and 'B' with English sentences in order to apply the rule to specific cases. We will be more precise about this kind of knowledge of inference rules in Sect. 3.2, but let us for now simply assume that the agent knows of the relevant instantiations. Given that each possible world is a maximal, logically consistent entity, it then follows deductively from the three propositions above that the following proposition is also true at all epistemically possible worlds for the agent:

the proposition that *the sentence 'AC Milan's game will be canceled' expresses a truth*.

In this sense, the agent's metalinguistic knowledge is deductively closed: if the agent knows the first three propositions above, then the agent also knows the fourth proposition that *the sentence 'AC Milan's game will be canceled' expresses a truth*.

In light of this example, it is now easy to appreciate Stalnaker's reasoning in the quote above. Suppose an agent knows the basic metalinguistic truths about a particular (axiomatic) proof system: he knows that 'AX_1', 'AX_2', …, 'AX_n' express the necessary proposition, where 'AX_1' to 'AX_n' are axiom sentences in the proof system, and he knows what modus ponens is and that it is the only rule in the system. Consider then any theorem sentence of the system: that is, any sentence that can be derived from 'AX_1', 'AX_2', …, 'AX_n' by (repeated) applications of modus ponens. Since the agent, on Stalnaker's view, knows every logical consequence of what he already knows, it follows that he knows, of every theorem sentence in the system, that it expresses the necessary proposition. This result is unacceptable to the

Stalnakerian. To be able to explain why it appears as if ordinary agents like us fall far short of logical omniscience, it should be possible for an agent to know that the axiom sentences in a given proof system express the necessary proposition without knowing of each theorem sentence in the system that it does.

As Stalnaker acknowledges, Powers (1976), Kripke (in conversation), and Field (1978) have made variations of the above objection to the metalinguistic strategy (p. 174). To deal with the objection, Stalnaker tries to avoid having to hold that an agent's belief and knowledge are automatically closed under logical consequence in the sense described above. To do this, Stalnaker introduces the idea that the non-ideal mind can be divided into various fragments or into different belief systems. He writes:

> "A person may be disposed, in one kind of context, or with respect to one kind of action, to behave in ways that are correctly explained by one belief state, and at the same time be disposed in another kind of context or with respect to another kind of action to behave in ways that would be explained by a different belief state. This need not be a matter of shifting from one state to another or vacillating between states; the agent might, at the same time, be in two stable belief states, be in two different dispositional states which are displayed in different kinds of situations." (Stalnaker, 1987, p. 83)

On Stalnaker's view, a fragmented agent is thus an agent who has several distinct belief systems encoding distinct bodies of information, each of which helps to explain the agent's behavior in different circumstances.[4] Formally, fragments correspond to sets of possible worlds. So each belief system within an agent is both deductively closed and logically consistent. While an agent may believe or know a proposition in a given fragment of his mind without believing or knowing that proposition in another fragment, the agent can be said to believe or know a proposition P *simpliciter* just in case P is true at all worlds that are doxastically or epistemically possible for the agent relative to at least one fragment.

Since belief and knowledge are relativized to fragments, it is now easy to see how Stalnaker avoids closing an agent's beliefs and knowledge under logical consequence. Suppose, for instance, that the proposition that *it rains* is true at all epistemically possible worlds for the agent relative to fragment F_1, whereas the proposition that *if it rains, AC Milan's game will be canceled* is only true at all epistemically possible worlds for the agent relative to fragment F_2. Insofar as the agent fails to put fragments F_1 and F_2 together—for whatever reason—the agent can know that it rains and that if it rains, AC Milan's game will be canceled without knowing that AC Milan's game will be canceled. Likewise, we can appeal to fragmentation to explain how the agent from above can know each axiom sentence and inference rule

---

[4] To be sure, there are more questions that one could ask about the nature of fragments; for some of these questions, see Borgoni et al. (2021). But in line with most other people in the philosophical literature appealing to fragments, we will settle with the rough characterization above. Technically, what matters for our argument is that fragments correspond to sets of possible worlds that are complete and deductively closed, and everyone in the Stalnakerian camp would agree with this.

in the proof system without knowing, of every theorem sentence in the system, that it expresses the necessary proposition. For the agent's logical knowledge of the system—say, the logical information he has about the axiom sentences—may be scattered across different fragments.

On the Stalnakerian account, then, metalinguistic ignorance yields an explanation of why ordinary agents can seemingly fail to know the necessary proposition, and fragmentation yields an explanation of why ordinary agents can fail to know the logical consequences of what they already know. In particular, the Stalnakerian strategy puts us in a position to explain how an agent can know, of each sentence in some premise set that the sentence expresses a truth, without knowing of every sentence, which follows deductively from the sentences in the premise set, that it expresses a truth.

The question remains: how plausible is the Stalnakerian reconciliation strategy?

3 The local omniscience problem

Ultimately, we will argue that the answer to this question is: "not very". As mentioned, we are of course not the first to argue that the Stalnakerian strategy faces difficulties. But in contrast to existing complaints, we want to cause trouble for the strategy on its home turf. That is, we will advance our objection while granting both the plausibility of the fragmentation component—save for an extreme version of it—and that of the metalinguistic component.[5] But obviously, this is not to say that we endorse either component.

To state our central argument against the Stalnakerian strategy, let $\Gamma \vdash$ 'C' be any sufficiently complex entailment from Γ to 'C', where $\Gamma = \{$ 'S$_1$', 'S$_2$', … 'S$_n$' $\}$ is a set of sentences (the premises) and 'C' is a single sentence (the conclusion). Let R be the set of inference rules that are needed to derive 'C' from Γ. For now we can think of R as containing simple inference rules such as conjunction introduction and modus ponens. The entailment from Γ to 'C' can be understood as a sequence of sentences 'T$_1$', 'T$_2$', … 'T$_n$' ending in 'T$_n$' = 'C', each member of which is either a member of Γ or inferable from one or two earlier elements in the sequence by applications of the rules in R. As above, let us say that an agent knows an inference rule

---

[5] Here are two examples that illustrate how existing critiques of the Stalnakerian strategy—in contrast to our approach—directly attack either the metalinguistic component or the fragmentation component. In Jago (2014a), the strategy is criticized for its characterization of mathematical and logical knowledge as, essentially, linguistic knowledge. For instance, failing to spot a particular winning strategy in chess seems hardly to reduce to a pure lack of linguistic knowledge. In Field (2001), it is argued that neither metalinguistic ignorance nor fragmentation can help account for the kind of behavior that is typically displayed when agents believe the impossible. For example, while the belief that a 60 degree angle can be trisected and the belief that a specific map requires more than four colors to color are both impossible, the types of behavior associated with these beliefs can be very different for mathematically untrained agents: one type involves the use of a compass while the other involves the use of color pencils. Yet, as Field argues, "[i]t does not seem […] that one can plausibly explain this difference in behavior in terms of different attitudes to *sentences* like 'I will trisect a 60 degree angle'; and invoking 'compartmentalized belief' does not seem substantially more promising" (Field 2001, p. 103).

like modus ponens whenever the agent knows a proposition with a content like *if 'A' and 'If A, then B' both express a truth, then 'B' expresses a truth*. Finally, to save breath, let us say that an agent *knows a sentence* 'S' whenever the agent understands 'S' and knows that the proposition that 'S' expresses is the proposition that it standardly expresses.[6]

We can now state our main argument against the Stalnakerian strategy. Central to the argument is the following 'local omniscience' (LoOm) result:

> (**LoOm**)
> When $\Gamma \vdash$ 'C', *if*
> (1) an agent knows in fragment F sentence 'S_i', for each 'S_i' in the premise set Γ; and
> (2) the agent knows in fragment F each inference rule in R,
> *then* the agent knows 'C' in F.

It is not hard to see why (LoOm) holds. Suppose that $\Gamma \vdash$ 'C' and that an agent knows in some fragment F each premise sentence 'S_i' in Γ. It then follows that each 'S_i' is true at all possible worlds that are epistemically possible for the agent relative to fragment F. Suppose 'T_1' is the first sentence in the sequence—eventually leading to 'C'—which follows from the premise sentences in Γ by application of the inference rule R_1 in R. If 'T_1' is to be false at some epistemically possible world relative to F, it must be because the agent fails to know the relevant inference rule R_1 in the fragment F. But, by condition (2), the agent knows in F rule R_1. So 'T_1' must be true at all possible worlds that are epistemically possible for the agent relative to F. So the agent knows 'T_1' in F. Since it is obvious how to repeat this line of reasoning, for each sentence 'T_i' in the sequence leading to 'C', the consequent in (LoOm) follows: the agent knows 'C' in F.

(LoOm) is problematic for a proponent of the Stalnakerian strategy. To see this, note that (LoOm), even for very sparse characterizations of the set Γ of premises and the set R of inference rules, entails a degree of logical omniscience that is unacceptable to a Stalnakerian. Suppose, for instance, that Γ only contains two sentences, and that R only contains standard inferential rules for two of the connectives—say, negation and implication. Even in that case, extremely complex theorem sentences can be expressed and hence derived in the corresponding system. To insist that ordinary agents must know of such complicated theorem sentences is very implausible. More generally, while it is true that we can express a more *limited* range of theorem sentences when we severely restrict the sets Γ and R, it seems wrongfooted to try to explain the appearance of logical non-omniscience by limiting the range of metalinguistic beliefs that agents hold about logical consequence. Rather, the real problem is that (LoOm) tells us that agents can *effortlessly* come to know sentences that follow only by very complicated logical reasoning from initially known premises—reasoning that goes far beyond the cognitive resources of ordinary agents. Severely restricting the range of such logical consequences does not address this problem.

---

[6] By holding that an agent understands 'S', we intend to rule out cases such as one in which an agent does not grasp 'S' but still knows, solely on the basis of testimony, that it expresses a necessary truth.

The worry posed by (LoOm) bears some similarities to the worry that Stalnaker himself raises for the purely metalinguistic component of his strategy—and that Stalnaker acknowledges is essentially a version of worries raised by Powers, Kripke, and Field. As we saw, the purely metalinguistic approach leads to an unwanted degree of logical omniscience for the Stalnakerian, and the appeal to fragmentation is meant to address this worry. But as (LoOm) shows, this appeal does not work. For even if it helps us to avoid full blown logical omniscience, we are still left with a degree of logical omniscience that is unacceptable to a Stalnakerian. For the purposes of reconciliation, it should be possible for an agent to know in a fragment that some (premise) sentences express a truth without knowing in that fragment of arbitrary logical consequences of these (premise) sentences that they do. Yet, given (LoOm), the Stalnakerian strategy seems incapable of delivering this result, even when we grant the fragmentation and metalinguistic components. So if the Stalnakerian strategy is to help us deal with logical omniscience, we must have a response to (LoOm). We discuss two such responses next.

3.1 First response to (LoOm)

One response to (LoOm) appeals to a sort of *extreme fragmentation*. As seen, (LoOm) entails a worrisome degree of logical omniscience even when Γ contains only two premise sentences and R only a few inference rules. However, as made clear by (1) and (2) in (LoOm), this conclusion presupposes that the agent in question simultaneously knows the relevant premise sentences and rules within a fragment. If we deny that this is possible, we can obviously block the derivation of 'C' in (LoOm). But denying that an agent can simultaneously hold information about just a few premise sentences and rules within a single fragment of his mind is tantamount to accepting that the non-ideal mind can be extremely fragmented.

Extreme fragmentation, however, is not a very attractive option in our opinion.[7] First, one might think that extreme fragmentation is psychologically unrealistic. After all, there are no findings in cognitive science—as far as we are aware—that suggest such an extreme degree of fragmentation or such compartmentalized cognitive architecture, and neither introspection nor intuitions suggest it either. Yet, one might deny that there is any perspicuous correspondence between a fragmentation-based model and details about human cognitive psychology. Elga and Rayo, for instance, whose view we will discuss in detail below, deny that their fragmentation based-model is "intended to map cleanly onto components of a realistic cognitive psychology" (Elga & Rayo, 2021, p. 43). It is a bit more unclear what other fragmentation-friendly philosophers such as Stalnaker and Lewis would think about the psychological plausibility of extreme fragmentation. Certainly, the kinds of cases that Stalnaker and Lewis use to motivate the idea of fragmentation do not suggest extreme fragmentation. Lewis gives the example of a double thinker who is

---

[7] Note: questioning *extreme* fragmentation is still compatible with engaging Stalnakerians on their home turf. For, as we argue later, Stalnakerians will want to capture the thought that agents are logically competent, but if extreme fragmentation is permitted, they will not be able to do so.

simultaneously disposed to act as if he is deadly sick and to act as if he is completely healthy: the hypochondriac fragment, for instance, might manifest in the morning while the cheerful one might manifest in the evening (Lewis, 1986, pp. 31–32). Similarly, Lewis proclaims that he used to believe both that Nassau St. in Princeton ran roughly east–west, that the nearby railroad and Nassau St. were roughly parallel, and that the railroad ran roughly north–south (Lewis, 1982, p. 432). While it is intuitively clear in such cases how we can explain the agent's conflicting dispositions to act by appealing to the idea of a fragmented mind, these explanations do not by any reasonable standards suggest the sort of extreme fragmentation that is seemingly needed to avoid (LoOm). But of course, one might hold, such cases are only the thin edge of the wedge.[8]

But secondly, and more worrisome, extreme fragmentation threatens to undermine the very common idea that ordinary agents are *minimally rational*.[9] While the notion of minimal rationality is multifaceted—we will explore Elga and Rayo's way of capturing a notion of minimal rationality in Sect. 4—we do not need any fancy theoretical groundwork to appreciate why extreme fragmentation can undermine most non-trivial standards for minimally rational beliefs. For if an agent can be arbitrarily fragmented, there is no guarantee that the agent will believe any sentence that logically follows from sentences he already believes, however obvious or trivial such logical consequences might be. Put differently, if extreme fragmentation is allowed, then anything goes, logically speaking, when it comes to metalinguistic reasoning. Even if an agent knows in a fragment that the propositions expressed by 'Rome is hot' and 'If Rome is hot, there are many fountains in Rome' are true, we cannot make any predictions about whether the agent will also know in that fragment that the proposition expressed by 'There are many fountains in Rome' is true too. For if the required contingent metalinguistic information about modus ponens is not known in the relevant fragment, the agent might fail to infer the latter proposition from the former two.[10]

To be sure, it is open to a proponent of a fragmentation-based strategy to deny that knowledge, beliefs, and credences satisfy any minimal standards of rationality. But such a move does not not seem very plausible. In fact, it is reasonable to assume that proponents of a fragmentation-based strategy do want to say that certain minimal standards of rationality should be in place. Indeed, Elga and Rayo (2022) uses fragmentation to ensure that logically non-omniscient agents nevertheless remain

---

[8] Thanks to an anonymous reviewer for suggesting how a proponent of extreme fragmentation might respond to some of the issues we raise for them.

[9] In addition to Lewis (1982), Stalnaker (1991), and Elga and Rayo (2022), the list of people who accept minimal standards on rationality is long; for some of the people on that long list, see Smithies et al. (2022).

[10] There is a trivial sense in which agents are minimally rational—and far more—in the Stalnakerian framework. After all, in this framework, agents are logically omniscient with respect to each fragment: they know every logical truth and every logical consequence of what they know in every fragment. But for our purposes, the interesting notion of minimal rationality is one that relates to mathematical and logical knowledge and reasoning (understood metalinguistically).

capable of performing logically competent deductions.[11] Stalnaker and Lewis are not as explicit as Elga and Rayo about the need to account for logically competent agents who fall short of logical omniscience. But as we pointed out earlier, even though Stalnakerians maintain that we have good reasons to adopt the possible worlds account of propositional content, they would grant that one still needs to explain the striking fact that it at least *appears* to us ordinary agents as if we fall far short of logical omniscience. Given that it also clearly appears to us as if we have some minimal level of logical competence—as indeed it seems plausible that we do—it is reasonable to presume that Stalnakerians would want to account for this intuition as well.[12] Since we will struggle, as argued, to meet such minimal standards of rationality if we allow extreme fragmentation, extreme fragmentation is undesirable for proponents of a fragmentation-based strategy.

But, in any case, even if extreme fragmentation in some form is acceptable, a proponent of the Stalnakerian strategy owes us an explanation of its grounds and nature. For now, we can hold that one way to avoid (LoOm) is to accept extreme fragmentation. Since we have argued that extreme fragmentation is unappealing, we do not find this reply to (LoOm) promising.

3.2 Second response to (LoOm)

Condition (2) in (LoOm) requires that the agent knows in the fragment the relevant inference rules that are needed to infer 'C' from Γ. But on the face of it, there are two ways in which we can understand what it means to know an inference rule: one can know a rule in a *schematic* sense and know a rule in an *instance* sense. Let us illustrate the distinction with modus ponens as an example—we trust that it is easy to see how it generalizes to other rules of propositional logic. Let us say that an agent has *schema* knowledge of modus ponens when the agent knows the proposition that *if 'A' expresses a truth and 'if A, then B' expresses a truth, then 'B' expresses a truth*, where 'A' and 'B' are placeholders for arbitrary English sentences. By contrast, let us say that an agent has *instance* knowledge of modus ponens when the agent knows the proposition that *if 'S_1' expresses a truth and 'If S_1, then S_2' expresses a truth, then 'S_2' expresses a truth*, where 'S_1' and 'S_2' are specific sentences in English. So essentially, in having instance knowledge of an inference rule, the variables in the inference schema are systematically replaced by sentences in English. Although we will not aim for a precise definition of the difference between having schema and instance knowledge of an inference rule, the above characterization should convey the central idea that is already familiar from logical and mathematical proof

---

[11] We will return to Elga and Rayo (2022) on logically non-omniscient but logically competent agents in Sect. 4. For further motivation of why fragmentation-based strategies do not sit well with everything-goes views on rationality, see Kinderman and Onofri (2021).

[12] In addition to Elga and Rayo (2022), there are many in the literature who share the thought that models of ordinary, non-omniscient agents should be able to capture the intuitive sense in which such agents remain capable of performing logically competent deductions; see, for instance, Bjerring and Skipper (2019, 2020), Cherniak (1986), Jago (2014a, b), Smets and Solaki (2018), Solaki (2021), and Weirich (2004).

contexts: just as an agent may have knowledge of an axiom schema, the agent may have—or may lack for that sake—corresponding knowledge of a specific instance of the schema. For instance, an agent may know in the schema sense that the formula '$A \rightarrow (B \rightarrow A)$' expresses a truth while not knowing that a specific instance such as '$(S_1 \rightarrow \neg S_2) \rightarrow (S_3 \rightarrow (S_1 \rightarrow \neg S_2))$' does.

In light of this distinction, suppose the derivation of 'C' from Γ requires applications of modus ponens—whatever we say here about deductive steps involving modus ponens applies to any deductive step involving the rules in R. A proponent of the Stalnakerian strategy might then respond to (LoOm) as follows. The derivation of 'C' from Γ requires that the agent knows in fragment F modus ponens in the *instance* sense. Yet, we can deny that the agent has this instance knowledge, for any step in the deduction requiring modus ponens, without losing the intuition that the agent knows in F what modus ponens is: namely in virtue of the agent having *schema* knowledge of the rule. So it is compatible with everything we have said that there is a particular instance of modus ponens in the derivation leading from Γ to 'C' that the agent fails to realize is such an instance. As such, against (LoOm), we can explain how an agent can fail to know in F the conclusion 'C' even when he knows in F the premise sentences in Γ and the relevant inference rules (in the schema sense).

While it is undoubtedly correct that it can often be hard to see that a particular instance of an inference schema is indeed such an instance—think about the last time you attempted to do an axiomatic proof—it is not clear that the distinction really helps a proponent of the Stalnakerian strategy. In (LoOm), by assumption, 'C' follows deductively from sentences that the agent already *knows* in fragment F. So the objection above requires that there is a specific step in the derivation of 'C' from Γ—here illustrated by a step involving modus ponens—such that the agent knows in F the sentences 'S_i' and 'If S_i, then S_{i+1}', and yet fails to know in F the sentence 'S_{i+1}'. But what can explain the agent's failure to see that such an instance is indeed an instance of the inferential schema for modus ponens? Typically, when people struggle to see that some logical formula is an instance of some axiom schema, it is because they do not fully grasp the logical form of the instance formula. But we cannot use this explanation in the context of (LoOm). Since the agent *knows* in F both the sentences 'S_i' and 'If S_i, then S_{i+1}', the agent knows that the propositions expressed by 'S_i' and 'If S_i, then S_{i+1}' are the propositions that these sentences standardly express—and he knows this because he *understands* the sentences, and not because, say, he has learned their truth through mere testimony. So we cannot explain why the agent fails to see that 'S_i' and 'If S_i, then S_{i+1}' are instances of the schema variables 'A' and 'if A, then B' by citing the agent's failure to somehow grasp or comprehend the logical form of the premise sentences. But then it is not quite clear what else could explain why the agent would lack the relevant instance knowledge.

A proponent of the Stalnakerian strategy might reply that people typically do not have schema knowledge of inference rules. Only logicians, they might argue, have this kind of knowledge. So the reason agents can fail to know the conclusion of a rule whose premise instances they know is to be explained in terms of a lack of knowledge of the corresponding rule schema. On its own, however, this kind of

reply is clearly unsatisfactory: trying to avoid (LoOm) by denying agents a type of contingent knowledge they often have seems clearly ad hoc. Also, troublesome logical omniscience should not be within your reach just because you are a first year logic student who has learned about the basic rules of propositional logic in terms of schemas and metavariables.

Thus our response to the second reply to (LoOm): while we acknowledge that an agent may know an inference rule in the schema sense without knowing it in the instance sense, the distinction does not help block the reasoning underlying (LoOm). Or, if it is to help, a proponent of the Stalnakerian strategy owes us an explanation of how an agent can know in a fragment the relevant inference schema, the relevant instances of the premises in the schema, and yet fail to know the relevant instance of the conclusion.

4 Elga and Rayo's appeal to fragmentation

To avoid (LoOm) and thus troublesome logical omniscience, we have so far argued that a proponent of the Stalnakerian strategy must *either* accept that agents can be extremely fragmented *or* that agents within fragments can know the relevant inference schema, the relevant instances of the premises in the schema, and yet fail to know the relevant instance of the conclusion. Both options seem implausible to us. At the least, the Stalnakerian owes us a further explanation as to why either option should be acceptable.

The objection that we have discussed also applies to Adam Elga and Augustín Rayo's recent attempt to avoid logical omniscience in Bayesian decision theory by Stalnakerian means. As we know, standard Bayesian decision theory requires that an agent's credences be represented by a standard probability function Cr—that they satisfy the well-known Kolmogorov axioms. From these axioms it is easy to derive the following "logical omniscience" theorem:

> (**Omni**) For any propositions P and Q such that P logically entails Q, Cr(Q) ≥ Cr(P).

According to (Omni), an agent's credences never drop across entailments—irrespective of how complicated the entailments are. So if we suppose that Fermat's Last Theorem follows logically from the Peano Axioms, and if we consider an agent who is certain of the conjunction of the Peano Axioms, then (Omni) tells us that this agent is also certain of Fermat's Last Theorem. But, as we know, there are many logically non-omniscient agents who at least seemingly can be certain of the Peano Axioms without being certain of Fermat's Last Theorem; Andrew Wiles was one such person back in the 1990s. So (Omni) clearly seems to fail for ordinary agents.

Accordingly, to make Bayesian decision theory applicable to ordinary agents, Elga and Rayo set out to devise a framework in which (Omni) fails. Yet, avoiding logical omniscience is only part of the challenge:[13]

> "For although everyday standards of rationality allow for some failures of logical omniscience, not just anything goes. For example, assuming Watson understands the logical connectives, it would be irrational for him to assign more credence to "it is sunny and windy" than to "it is sunny". And the same would go for assignments that violate other obvious logical entailments. But were we to discard the standard probabilistic coherence assumptions altogether, nothing would rule out such assignments." (Elga & Rayo, 2022, p. 717)

Let us say that a *logically competent agent* is an agent whose credence function respects obvious logical entailments. Extrapolating from the quote above, let us for concreteness say that an entailment from P to Q is *obvious* whenever Q can be inferred from P by at most one application of a basic rule of propositional logic. While Elga and Rayo never explicitly define what it means for an entailment to be obvious, this characterization seems faithful to their underlying ideas—as witnessed by their example of an obvious entailment involving a single application of conjunction elimination from "It is sunny and windy" to "It is sunny". For our overall argument, though, nothing hangs on the finer details here.

One might think that an agent who is logically competent but not logically omniscient is simply one whose credence function respects all obvious logical entailments without respecting all the non-obvious ones. But as Elga and Rayo point out, "on any interesting way of spelling out 'obvious', chaining together obvious entailments can result in a non-obvious one" (Elga & Rayo, 2022, p. 718).[14] That is, if we understand an entailment from 'S_1' to 'S_n' as a series of obvious entailments from 'S_1' to 'S_2', from 'S_2' to 'S_3', and so on, it can be shown that a credence function that respects each obvious entailment from 'S_i' to 'S_{i+1}' will thereby respect a single entailment from 'S_1' to 'S_n', even if this entailment is far from obvious.

So there has to be another way to model logically non-omniscient, yet logically competent agents. To develop a Bayesian framework that can do this job, Elga and Rayo begin with the idea that an agent's decision-theoretic state can be broken into different fragments relative to the information that is accessible to an agent in a given choice situation. Although we are not given a systematic explanation as to what *accessibility* amounts to, it is closely connected to what is *salient* to an agent, or to what an agent *attends* to or is *aware* of in a given context. Yet, what Elga and Rayo do say suggests that accessibility, salience, and attention are rather *fine-grained* notions:

> "[T]ake a condition in which only sentences 'S' and '($W\&R$)' are salient. Relative to such a condition, the entailment from '$S\&(W\&R)$' to '($W\&R$)' counts

---

[13] For similar considerations of what a proper solution to the problem of logical omniscience requires, see Bjerring and Skipper (2019), Jago (2014a, b), and Weirich (2004).

[14] For earlier statements of this result, see Bjerring and Schwarz (2017).

as obvious because it is guaranteed by the meaning of '&', as it applies to '*S*' and '(*W&R*)'. In contrast, relative to that same condition the entailment from '(*W&R*)' to '*R*' does *not* count as obvious, since *R* is not salient." (Elga & Rayo, 2022, p. 721.)

So according to Elga and Rayo, a complex sentence can be salient to an agent without each component of the sentence being so—just as one can apparently attend "to a forest without simultaneously attending to each of its trees" (Elga & Rayo, 2022, p. 727).

On Elga and Rayo's view, the fragmentation of the non-ideal mind is thus tied to the information that is accessible or salient to an agent in a given choice condition. To borrow one of their own examples, consider two people who are trying to solve a crossword puzzle.[15] Both are tasked with completing the blanks to generate a word in English: "_ _ _ _ MT". Both puzzle solvers, Elga and Rayo assume, know that *dreamt* is a word of English, and both know how to spell it. Yet, while the first person solves the puzzle by filling in just the right letters to generate the word "DREAMT", the second person fills in nothing. Why?

Elga and Rayo write:
"We suggest that both puzzlists possess the information they need to fill in the blanks, but that the conditions relative to which they have access to this information are different. Let *D* be the set of worlds in which *dreamt* is a word of English spelled D-R-E-A-M-T. Both puzzlists have access to *D* for the purpose of using "dreamt" in a written essay. And they both have access to *D* for the purpose of answering the question "Is 'dreamt' a word of English ending in MT?". But for the purpose of filling in the blanks in "M T", only the first puzzlist has access to *D*." (Elga & Rayo, 2022, p. 718)

The information that resides in each fragment is modelled by a set of possible worlds. So the information that *dreamt* is a word of English spelled D-R-E-A-M-T is modeled by the set of possible worlds in which *dreamt* is a word of English spelled D-R-E-A-M-T. But whether an agent can access this information—whether the fragment that stores this information is active—depends on the agent's choice condition. Relative to the conditions of solving the crossword puzzle, only the first person can access the fragment.

Following Elga and Rayo, we can associate a probability function Pr with each choice condition and represent each puzzle solver's decision-theoretic state by means of a so-called *access table*. The following table represents the second puzzlist's decision-theoretic state.[16]

Choice condition	Accessible information
Working on puzzle; *dreamt* salient	Pr_1

Choice condition	Accessible information
Working on puzzle, *dreamt* not salient	Pr_2
[Further conditions]	[Information accessible relative to those conditions]

Each probability function encodes the information that is accessible relative to the specific choice condition. Since information is represented by a set of possible worlds, it follows that each probability function encoding that information is probabilistically coherent. According to Elga and Rayo, while we should not think of access tables as directly corresponding to families of propositional attitudes, it is "appropriate to ascribe A [a family of propositional attitudes] to a fragmented subject if and only if the dispositions predicted by A are sufficiently similar to the dispositions predicted by the subject's access table" (Elga & Rayo, 2022, p. 733).

In the case of the two puzzle solvers, we know that different dispositions are manifested in the context of solving the puzzle. Let the probability function Pr_1 associated with the first row in the access table assign a high probability to the proposition DREAMT—the set of possible worlds in which *dreamt* is a word of English spelled D-R-E-A-M-T—and let the probability function Pr_2 associated with the second row assign a low probability to that proposition. In the context of trying to solve the puzzle, the first row is inactive for the second puzzlist since Pr_1 predicts that he will be disposed to fill in the blanks with the correct letters. By contrast, the second row is active for him since Pr_2 predicts that he will not be disposed to fill in the blank with the correct letters. Of course, in a context in which *dreamt* is salient to the second puzzlist, the first row will be active for him. Thus we have a fragmentation-based explanation of why the second puzzlist can still be said to know or to have a high credence in DREAMT despite not manifesting that knowledge or high credence in the particular context of solving the puzzle: he lacks the relevant disposition in that context but manifests it in others.

However, while it is appropriate to ascribe to the second puzzle solver a family of propositional attitudes that includes a high credence in DREAMT, it is not appropriate to include a low credence in DREAMT in that family. For having a low credence in DREAMT would mean also having a high credence in its negation—that is, having a high credence in the proposition that 'dreamt' is not a word of English spelled D-R-E-A-M-T. But since the second puzzlist presumably does not have any disposition associated with such a credence, it would be strange to ascribe such a propositional attitude to him. Intuitively, if an agent cannot recall a word like 'dreamt', it is not because he is confident that it is not a word in English, or that it is not spelled D-R-E-A-M-T. Rather, it is because he assigns neither DREAMT nor not-DREAMT a credence at all. But of course, on the standard interpretation of a credence function that Elga and Rayo adopt, this is not an option. So they are explicit in denying that an agent's credence in a proposition can be read off directly from the probability functions associated with rows in an access table.[17] Nonetheless, for our purposes,

---

[17] Elga & Rayo (2022) explicitly deny that "an agent is confident in a claim if and only if some row of her access table assigns high probability to that claim", for the biconditional "entails that a puzzlist with

there is often no harm in assuming that each probability function in an access table does as a matter of fact map on to the agent's credence function in some fragment of his mind. We just have to remember that this does not hold in general on Elga and Rayo's view.

We now have the tools to appreciate Elga and Rayo's attempt to avoid logical omniscience while retaining logical competence. Using as an illustration Elga and Rayo's own example from the quote above, suppose 'S' and '(W & R)' are salient in fragment F_1, but that 'R' is not, and suppose that 'S', '(W & R)', and 'R' are all salient in fragment F_2. Suppose also that the relevant logical information about the rule of conjunction elimination is accessible in both F_1 and F_2—bracket for now whether this information consists in instance or schema knowledge of the rule. We can use the following access table to model this fragmented state of mind:

Choice condition	Accessible information
'S' and '(W & R)' are salient; 'R' is not salient; logical information about conjunction elimination is salient	Pr_1
'S', '(W & R)', and 'R' are salient; logical information about conjunction elimination is salient	Pr_2

Suppose $Pr_1('S') \geq Pr_1('S \ \& \ (W \ \& \ R)')$ and $Pr_1('W \ \& \ R') \geq Pr_1('S \ \& \ (W \ \& \ R)')$. Assuming that the probability function Pr_1 maps on to an agent's credences in fragment F_1, the first row in the access table suggests a credence distribution that respects the obvious entailment from 'S & (W & R)' to 'S' and '(W & R)'. Yet, since 'R' is not salient relative to the choice condition in the first row, we need not suppose that $Pr_1('R') \geq Pr_1('S \ \& \ (W \ \& \ R)')$. So neither do we need to assume that the corresponding credence distribution, contrary to what is demanded by (Omni), respects the entailment from '(W & R)' to 'R'. On the other hand, since $Pr_2('R') \geq Pr_2('W \ \& \ R')$ when 'R' is salient, the second row in the access table does suggest that there is a fragment F_2 relative to which the agent's credence in 'R' is at least as high as his credence in '(W & R)'. As such, the access table above can be used to characterize a logically non-omniscient, yet logically competent agent: non-omniscient because there are certain conditions—there is a certain fragment F_1—in which the agent's credences fail to respect (Omni), and competent because there are certain conditions—there is a certain fragment F_2—in which the agent's credences do respect the obvious entailment from '(W & R)' to 'R'.

Generalizing this idea, Elga and Rayo want to capture an agent's *logical competence* by claiming that, for each obvious logical entailment, there is at least one fragment within the agent's mind such that the credence function associated with that fragment respects that entailment. At the same time, they want to avoid *logical omniscience* by claiming that the pieces of information required to deductively

---

Footnote 17 (continued)

the access table [above] is confident not just in the claim that *dreamt* is a word of English spelled D-R-E-A-M-T, but also in the negation of that claim" (p. 733).

infer a conclusion from a set of premises need not be contained within a single fragment.[18] So Elga and Rayo have a model of how ordinary agents can display logical competence while simultaneously failing to display logical perfection.

In view of (LoOm), how plausible is Elga and Rayo's Stalnakerian inspired approach to logical omniscience? Consider first an obvious entailment from 'S_1' and 'S_2' to 'S_3', and consider an agent who knows in the schema sense the relevant inferential rule R_1 that permits inferring the conclusion instance 'S_3' from the known premise instances 'S_1' and 'S_2'. Since the entailment is obvious, there is a fragment F_1 and an associated credence function Cr_1 that respects that entailment: $Cr_1('S_3') \geq Cr_1('S_1 \& S_2')$. Suppose now that '$S_4$' follows obviously from 'S_2' and 'S_3' by another instantiation of the rule R_1. To avoid collapsing obvious entailment into non-obvious entailment, and, eventually, full entailment—thereby creating a situation in which, with respect to the rules in F_1, (Omni) is satisfied and logical omniscience is restored—Elga and Rayo need the following to be possible in F_1: $Cr_1('S_4')$ is *not* greater than or equal to $Cr_1('S_2 \& S_3')$.[19] The question arises: can they get this result?

We think not. To see this, let us look at the epistemic situation from the perspective of the possible worlds that make up a fragment. Following Elga and Rayo, suppose we encode what it means to have schema knowledge of a rule permitting one to infer 'C' from 'A' and 'B' as involving knowledge of the following sort of proposition:

(R_1) For all situations v, if 'A' and 'B' express truths in v, then 'C' also expresses a truth in v.[20]

So, for instance, if an agent has schema knowledge of conjunction elimination in a fragment, the agent will know the proposition that for all situations v, if 'A & B' expresses a truth at v, then 'A' and 'B' express a truth at v, where, as we saw above, 'A' and 'B' are understood as placeholders for arbitrary sentences. To apply this schema knowledge to specific conjunctions, the agent must know that the relevant conjuncts are instances of the placeholders in the schema. For instance, for an agent to know that certain sentences follow from the known conjunction 'S_1 & S_2', the agent must know that 'S_1 & S_2' is an instance of 'A & B' in the schema for conjunction elimination. But if the agent knows these propositions in a fragment F_i, it is not hard to see that he must then also know 'S_1' in F_i. For in light of the agent's knowledge, *all* epistemically possible worlds in F_i verify the propositions that 'S_1 & S_2' expresses a truth, that every situation in which 'A & B' expresses a truth is a situation in which 'A' expresses a truth, and that 'S_1 & S_2' is an instance of 'A & B'.

---

[18] For further details, see Elga and Rayo's (2022) discussion of how they can "do justice to Frege's logical and semantic competence while respecting his lack of logical and semantic omniscience" (p. 720).

[19] Clearly, the specific claims here are dependent on the characterization of an obvious entailment as involving only a single application of a basic inference rule. Yet, the general point is not. For by chaining together obvious entailments—whether characterized as above or not—we eventually arrive at a non-obvious one. That is, there will have to be a specific step (or steps) in the deduction where we move from an obvious to a non-obvious entailment.

[20] Cf. Elga & Rayo (2022), p. 720.

Since possible worlds are deductively closed, it follows that the proposition that 'S_1' expresses a truth is also true at all epistemically possible worlds in F_i, and hence that the agent knows that 'S_1' expresses a truth in F_i.

We can now apply this line of reasoning to the case above. To ensure that $Cr_1('S_3') \geq Cr_1('S_1 \& S_2')$, the following has to be the case: for *all* epistemically possible worlds w in the relevant fragment F_1, if w verifies the propositions that 'S_1' and 'S_2' express truths, then w also verifies the proposition that 'S_3' expresses a truth. Since the agent, by assumption, knows the sentences 'S_1' and 'S_2' in F_1, all epistemically possible worlds relative to F_1 verify the propositions that 'S_1' and 'S_2' express truths. Hence all epistemically possible worlds relative to F_1 also verify 'S_3'. Given that $Cr_1('S_3')$ is higher than or equal to $Cr_1('S_1 \& S_2')$ *as a result of* the agent's schema knowledge of the rule R_1 taking one from 'A' and 'B' to 'C' in the sense above, the agent must know that 'S_1', 'S_2', and 'S_3' are instances of the schema variables 'A', 'B', and 'C' respectively. Accordingly, all epistemically possible worlds in F_1 verify the propositions that 'S_2' and 'S_3' express a truth, that every situation in which 'A' and 'B' express a truth is a situation in which 'C' expresses a truth, and that 'S_2' and 'S_3' are instances of 'A' and 'B'. Since the sentence 'S_4', by assumption, follows obviously from 'S_2' and 'S_3' by an application of R_1, and since possible worlds are deductively closed, it now follows—by the line of reasoning from above—that the proposition that 'S_4' expresses a truth is also true at all epistemically possible worlds in F_1. If so, it also follows that $Cr_1('S_4') \geq Cr_1('S_2 \& S_3')$, which is contrary to what Elga and Rayo need to avoid logical omniscience. Since it is obvious how to repeat this line of reasoning for every logical consequence of 'S_1', 'S_2', and 'S_3', the reasoning behind (LoOm) thus shows that, with respect to the relevant inference rules, credence functions that respect obvious logical entailment must respect logical entailment simpliciter. Thus a restricted version of (Omni) still holds within fragments: for any P and Q such that P logically entails Q (with respect to the relevant inference rules), $Cr_i(Q) \geq Cr_i(P)$, where Cr_i is the credence function associated with fragment F_i, for each fragment F_i in the non-ideal mind. Hence Elga and Rayo have not, contrary to what they claim, managed to avoid troublesome logical omniscience.

How might Elga and Rayo reply to our argument? They might deny that any individual fragment ever has schema knowledge of inference rules like (R_1). If that is true, the argument above is clearly blocked. But, as touched upon above, we struggle to see why we should ban fragmented agents from having the contingent knowledge that metalinguistic knowledge of inference schemas amounts to. Certainly, there is nothing about the lack of logical omniscience per se that should prevent logically competent agents from having schema knowledge of inference rules. For although not all agents will in fact possess such knowledge, there is nothing that suggests that it is somehow too cognitively demanding for ordinary agents to acquire this knowledge.[21] For instance, we can reasonably expect that first-year students of

---

[21] Note also: when Stalnaker in an earlier quote talks about agents having knowledge of the "contingent fact that each axiom sentence expresses a necessary truth (*however the descriptive terms are interpreted*)", he seems to grant that agents can have *schema* knowledge of axioms and inference rules (Stalnaker 1987, p. 76; our *italics*).

mathematics will know that sentences such as 'For all x, y, and z, if $x > y$, and $y > z$, then $x > z$' expresses the necessary truth, and that they can put this knowledge to work in solving various mathematical problems. And likewise, we can reasonably expect that first-year students of logic will learn about inference schemas such as *if 'A' expresses a truth and 'if A, then B' expresses a truth, then 'B' expresses a truth*, and that they can employ such knowledge in solving various logical problems. Denying that agents—even if fragmented—should be able to come to know these types of contingent propositions seems ad hoc and unmotivated. Accordingly, we claim, if modeling an agent with contingent schema knowledge of inference rules leads to worries about logical omniscience, the problem lies with the model, and not with the assumption that an agent can have such knowledge.[22] Or, to put the point in another way, even if Elga and Rayo only talk about instance knowledge of inference rules, their model of logically competent agents *should* be able to account for schema knowledge of such rules as well. If they are forced to deny that agents can have this schema knowledge—precisely because such knowledge leads to omniscience worries for their model—their model thus faces a problem.

Let us emphasize that, in raising the worry above, we are still engaging with Stalnakerians on their home turf. We are still granting that fragmentation—though not the extreme version—is plausible, and we are still granting that mathematical and logical knowledge can be thought of in metalinguistic terms. We are also not begging the question against Elga and Rayo. We do not merely presume that ordinary agents have schema knowledge. Instead, we have provided evidence that it is implausible to hold that they cannot or never have such knowledge.

Short of denying agents the capacity to have schema knowledge of basic inference rules, it seems that Elga and Rayo can really only avoid our worry by appealing to their concepts of awareness or salience. As we saw above, Elga and Rayo work with a very fine-grained individuation of salience according to which, for instance, the sentences 'S' and '$(W\&R)$' can be salient to an agent although 'R' is not. In the case at hand, since the agent is required to know 'S_3' in F_1 because of the obvious entailment from S_1 and S_2, all these sentences are presumably salient to the agent in F_1 together with information about the rule R_1. Yet, a defender of Elga and Rayo might argue, since 'S_4' need not be salient to the agent in F_1, the entailment from 'S_2' and 'S_3' to 'S_4' need *not* count as obvious. If so, there is no requirement that

---

[22] Instead of capturing schema knowledge of inference rules in terms of knowledge of contingent propositions such as (R_1), one might suggest that we capture it by appealing to a kind of rule-following behavior that generally—but not always—respects the inferential patterns suggested by the rule in question. For instance, as suggested by a referee for this journal, we might say that an agent has schema knowledge of modus ponens when, generally, the following obtains: when 'A' is known relative to some fragment F_1, and when 'If A, then B' is known relative to some other fragment F_2, then 'B' is known relative to some fragment F_3. Setting aside whether such an account of schema knowledge would avoid the problems that we have isolated for Elga and Rayo—we doubt that it will—it requires a story about how different fragments combine and interact to generate the knowledge that 'B' in F_3 *as a result* of applying modus ponens on what is already known in F_1 and F_2. But note: even if this story is available, it still does nothing to suggest that agents cannot also have schema knowledge of inference rules in the way we have suggested. And that is strictly all we need to make our case against Elga and Rayo.

the credence function associated with F_1 respects the entailment, in which case our argument is blocked.

For this reply to have traction, we need to know much more about salience or awareness. Since information about 'S_1', 'S_2', 'S_3', and the rule R_1 is already salient to the agent in fragment F_1, our imagined defender of Elga and Rayo needs an explanation of why 'S_4' does not also count as salient to the agent. After all, 'S_4' follows obviously from sentences and rules that are already salient to the agent in F_1: all the logical and semantic information required to derive 'S_4' is, as it were, at the forefront of the agent's mind. Note, in particular, that we can construct the inferences from 'S_1' and 'S_2' to 'S_3', and from 'S_2' and 'S_3' to 'S_4' as inferences from simple sentences to more complex ones composed of only the simple sentences, where the simple sentences 'S_1' and 'S_2'—and the inference rule generating the more complex sentences—*are* salient to the agent in F_1. For instance, we can let 'S_1' and 'S_2' entail 'S_3', where 'S_3' = '$(S_1 \& S_2)$', and we can let 'S_2' and 'S_3' entail 'S_4', where 'S_4' = '$S_2 \& (S_1 \& S_2)$'. In this case, it is completely unclear why a slight increase in the logical complexity of S_4 should somehow make S_4 cease to be salient when all the simpler components of S_4 are salient. Accordingly, even if we grant Elga and Rayo the quite puzzling idea that a complex sentence like '(*W&R*)' can be salient to an agent without one of its parts being so, the current case is different because all parts of the derived complexes are, by assumption, salient to the agent. Further, Elga and Rayo themselves would seem to be sympathetic to the claim that 'S_4' is salient when information about all the relevant sentences and rules are salient. In a slightly different context in which they present their model of Bayesian *reasoning*—where such reasoning takes place across time—they are happy to hold that "each step in a chain of thought renders a particular set of sentences salient", where one such salient sentence is precisely the conclusion of the step in question (Elga & Rayo, 2022, p. 723).

So to avoid (LoOm) and hence unwanted logical omniscience, it seems that Elga and Rayo must grant that an agent, even *within* a fragment, can fail to perform even the simplest of inferences from premises and rules that the agent both knows and attends to. Indeed, the reasoning behind (LoOm) suggests that Elga and Rayo—contrary to their intentions—struggle to avoid troublesome logical omniscience when they manage to ensure logical competence. For each fragment in the non-ideal mind, that is, if the associated credence function respects obvious entailment, it respects entailment simpliciter. This means that a restricted version of (Omni) still holds relative to each fragment, in which case the non-ideal mind still enjoys an unacceptably high degree of logical omniscience in Elga and Rayo's framework.

5 Conclusion

What are the prospects for the Stalnakerian reconciliation strategy? In an ideal world, the strategy would give us all the benefits of the possible worlds framework while accommodating our intuitions about our logical non-omniscience. Yet, as we have argued, we do not live in an ideal world. The prospects for the Stalnakerian strategy—and accounts that have fragmentation and metalinguistic ignorance at

their core—look dim. So if we want a framework for reasoning about logically non-omniscient agents, fans of, say, situation semantics, truthmaker semantics, or impossible worlds semantics are warranted in continuing to develop these alternatives to the possible worlds framework.

References

Berto, F., & Jago, M. (2019). *Impossible worlds*. Oxford University Press.

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

Bjerring, J. C., & Skipper, M. (2019). A dynamic solution to the problem of logical omniscience. *Journal of Philosophical Logic, 48*(3), 501–521.

Borgoni, C., Kindermann, D., & Onofri, A. (eds.) (2021). *The Fragmented Mind*. Oxford University Press.

Braddon-Mitchell, D., & Jackson, F. (1996). *The philosophy of mind and cognition: An introduction*. Blackwell Publishing.

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

Elga, A., & Rayo, A. (2022). Fragmentation and logical omniscience. *Noûs, 56*(3), 716–741.

Elga, A., & Rayo, A. (2021). Fragmentation and Information Access. In *Borgoni et al. 2021* (pp. 37–53).

Dogramaci, S. (2018). Solving the problem of logical omniscience. *Philosophical Issues, 28*, 107–128.

Easwaran, K. (2011). Bayesianism II: Applications and criticisms. *Philosophy Compass, 6*(5), 321–332.

Field, H. (2001). *Truth and the absence of fact*. Oxford University Press.

Field, H. (1978). Mental representation. *Erkenntnis, 13*, 9–61.

Forbes, G. (1989). Review: Robert Stalnaker, Inquiry, The MIT Press, Cambridge, Massachusetts, 1984. *Synthese, 79*, 171–189.

Greco, D. (2021). Fragmentation and Coarse-Grained Content. In *Borgoni et al. (2021)* (pp. 54–77).

Jago, M. (2014a). T*he impossible: An essay on hyperintensionality*. Oxford University Press.

Jago, M. (2014b). The problem of rational knowledge. *Erkenntnis, 79*(6), 1151–1168.

Kindermann, D., & Onofri, A. (2021). The fragmented mind: An introduction. In *Borgoni et al*. 2021 (pp. 1–33).

Lewis, D. (1982). Logic for equivocators. *Noûs, 16*, 431–441.

Lewis, D. (1986). *On the plurality of worlds*. Blackwell Publishers.

MacKay, A. F., & Merrill, D. D. (Eds.). (1976). *Issues in the philosophy of language*. Yale University Press.

Powers, L. (1976). Comments on Stalnaker, Propositions. *In MacKay and Merrill, 1976*, 93–103.

Robbins, P. (2004). To structure, or not to structure. *Synthese, 139*, 55–80.

Skipper, M., & Bjerring, J. C. (2020). Bayesianism for non-ideal agents. *Erkenntnis, 1*, 1–23.

Smets, S., & Solaki, A. (2018). The effort of reasoning: Modelling the inference steps of boundedly rational agents. In *International workshop on logic, language, information, and computation* (pp. 307–324). Springer.

Smithies, D., Lennon, P., & Samuels, R. (2022). Delusions and madmen: Against rationality constraints on belief. *Synthese, 200*(3), 1–30.

Solaki, A. (2021). *Logical models for bounded reasoners* (Doctoral dissertation, University of Amsterdam).

Stalnaker, R. (1986). Replies to Schiffer and Field. *Pacific Philosophical Quarterly, 67*(2), 113–123.

Stalnaker, R. (1987). *Inquiry (A Bradford Book)*. MIT Press.

Stalnaker, R. (1990). Mental content and linguistic form. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 58*(1/2), 129–214.

Stalnaker, R. (1991). The problem of logical omniscience, I. *Synthese, 1*, 425–440.

Stanley, J. (2010). 'Assertion' and intentionality. *Philosophical Studies, 151*, 87–113.

Titelbaum, M. (forthcoming): *Fundamentals of Bayesian epistemology*. Forthcoming with Oxford University Press.

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