Tackling Logical Omniscience: A Probabilistic Extension of the Logic of Fast and Slow Thinking

Candidate Number: 27008
Supervisor: Dr. Laurenz Hudetz

August 17, 2022

PH499- MSc Philosophy of Science Dissertations (2021-2022)

Word Count: 9946
(Excluding the Title Page, Abstract, Keywords, Table of Contents and Bibliography.)

Dissertation Submitted to the Department of Philosophy, Logic and Scientific Method, LSE for the partial fulfillment of the degree of MSc in Philosophy of Science in August, 2022.

Acknowledgements

I would like to thank my supervisor Dr. Laurenz Hudetz, for his guidance and support throughout the year. I would also like to thank Prof. Miklös Rédei for his advice and guidance. I benefited a lot from their office hours and they also helped shape my future research interests. Also, I would like to thank Prof. Roman Frigg, Prof. Richard Bradley and Dr. Anna Mahtani for their support and advice during the office hours. They gave me extra motivations to come up with new ideas so that I could discuss with them during the office hours. I would like to thank Prof. Christian List for his detailed feedback on my draft. I would also like to thank the organisers, reviewers and audience of the conferences AAL2022, the organisers and reviewers of AiML2022, DaLi2022 for their valuable feedback.

Finally, I would like to express gratitude to my parents, who encourage me to pursue my interests and try to not worry about life. I have not been home for three years due to the pandemic but I know that my family always stand by me.

Abstract

Traditional epistemic logic and orthodox Bayesianism assume that agents are logically omniscient. This assumption is problematic because it sets an overly idealized standard of rationality for agents with limited computational resources. In an attempt to model non-logically-omniscient, but moderately rational agents, Solaki et al. (2021) proposed a framework for epistemic logic that can model the logical aspects of System 1 ("fast") and System 2 ("slow") cognitive processes. This paper proposes a framework for the probabilistic extension of this logic and discusses the philosophical implications of the proposed framework.

1 The Logical Omniscience Problem

Epistemic logic uses logical approaches to study epistemic concepts such as knowledge and belief and the orthodox epistemic logic at present is modal logic (Rendsvig and Symons, 2021). In this particular epistemic logic, the operator '$\Box$' stands for the epistemic modality 'the agent[1] knows that' instead of the commonly-known metaphysical modality 'it is necessary that'. This approach of using normal modal logic as epistemic logic brings the benefit of enabling us to use possible worlds semantics to study knowledge and belief. In modal logic, $\langle M, w\rangle \vDash \Box\phi$ iff $(\forall w' \in W)((w, w') \in R \rightarrow \langle M, w'\rangle \vDash \phi)$. In the context of epistemic logic, it says that an agent knows ϕ in world w iff ϕ is true in all worlds accessible from w (the worlds considered possible by the agent). However, since it also shares the axioms and rules of normal modal logic, some unwanted features may be generated. In particular, the problem of logical omniscience arises from the Distribution Axiom and the Necessitation Rule of normal modal logic (Artemov and Kuznets, 2014).

Distribution Axiom: $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

Necessitation Rule: $\vdash A \Rightarrow \vdash \Box A$

The Distribution Axiom could be written as: $(\Box(A \rightarrow B) \wedge \Box A) \rightarrow \Box B$, which states that an agent knows B if she knows both A and $(A \rightarrow B)$. In other words, the agent's knowledge is closed under *modus ponens*. The Necessitation Rule states that an agent knows all valid facts (Tautologies). In the case when $(A \rightarrow B)$ is actually valid, the agent would know it by the Necessitation Rule. She would then know B once she knows A. To generalize this property:

Whenever an agent c knows all of the formulas in a set Γ and A follows logically from Γ, then c also knows A.

This property is against our understanding of knowledge since it would unrealistically idealize the agent's cognitive abilities. For example, any rational agent who knows the rule of chess should know whether White has a non-losing strategy (Artemov and Kuznets, 2009). Given a formula ϕ in propositional logic, the agent should

[1] This can be easily generalised to multi-agent cases by including multiple modal operators '$\Box_i$' to stand for 'agent$_i$ knows that'.

know whether there exists a satisfying assignment for ϕ. This is obviously not the case for many resource-bounded agents because the deductive reasoning process could be too complicated for them to gain the conclusion as knowledge.

Arguably, the normative standard of rationality should not be set by this idealized model. For it is intuitively too harsh to accuse an agent of being irrational just because she could not meet this standard of ideal rationality due to her constraint of computation resources, which is a constraint that applies to all agents in real-life. Also, the ideal notion of rationality becomes vacuous when we try to model the reasoning of agents with bounded resources in the context of AI, game theory, decision theory, etc. As pointed out by Stalnaker (1991), we would be unable to account for contemplation of facts already known to us, which is the main activity and a main subject of philosophy, if we assumed logical omniscience. As a result, there is a strong incentive to propose a more plausible model of reasoning, one that can model non-logically-omniscient, but still moderately rational agents.

1.1 The Awareness Approach

Different approaches have been proposed to remove the unwanted property and many of them involve modifications of the normal modal logic as epistemic logic. We would now introduce two main approaches[2]:

The addition of awareness to the standard model captures the intuition that an agent must be aware of a concept before she knows it (Fagin and Halpern, 1987). Specifically, $A(w)$ is the set of formulas that the agent is aware of at world w. This can be seen as adding a filter on the set of knowledge in the standard model:

$$(M, w) \models \Box\phi \text{ iff } (M, w') \models \phi \text{ for all } w' \in W \text{ and } \phi \in A(w),$$

This approach marks the difference between implicit knowledge and explicit knowledge. Specifically, some 'knowledge' in the standard approach can be seen as implicit in the sense that the agent is not aware of them even though they are the logical consequences of what the agent knows explicitly. This also reveals some problems with the standard epistemic logic since it cannot make distinction between implicit and explicit knowledge and would include some formulas that the agents are not even aware of as knowledge. In contrast, the awareness approach refines the notion of knowledge by excluding such formulas as explicit knowledge and removes the logical omniscience property. This approach could capture the intuition of situations when an agent makes a mistake that she could have avoided with more careful thinking. The overlooked formulas that the agent could have been aware of are described as implicit but not explicit knowledge in the awareness approach.

[2] A survey of more approaches of tackling the logical omniscience problem can be found in Moreno (1997)

1.2 The Impossible Worlds Approach

Another approach to avoid the problem is to distort the knowledge of ideal agents. Impossible worlds are worlds in which tautologies may not be true and inconsistent formulas may be true (Cresswell, 2016; Hintikka, 1979; Rantala, 1982). In the standard model, an agent knows A at world w if A is true in all worlds that are epistemically accessible from w. For example, in the middle of a chess game, a chess player knows that she will win the game if she will win in every outcome that are considered possible by her. However, only possible worlds are considered as epistemically accessible by the agent and an important source of the logical omniscience problem is captured by this idealization: Ideal agents do not mistakenly consider impossible worlds as possible.

In response, the impossible worlds approach tries to avoid the problem by distorting the knowledge of ideal agents. By adding impossible worlds to the structure, the agent would not take all tautologies as truth and would have inconsistent knowledge. Therefore, the agent would not be logically omniscient. For example, the chess player in the previous example may falsely consider an impossible situation (one in which her opponent will make a move against the rule and defeat her) as epistemically accessible from her current state and fail to see that she has a sure win [3].

The impossible world structure is a tuple $M = (W^P, W^I, \pi, R, C)$, where W^P is the set of possible worlds, W^I is the set of impossible worlds which are mistakenly considered possible by the agent, π is a valuation that assigns truth values to each primitive propositions at each world in W^P, $R \subseteq W^P \cup W^I \times W^P \cup W^I$, and C associates a set of formulas to each world in W^I. Different from the possible world structure, a set of impossible worlds W^I is included and this means that the agent would also consider worlds that are not in W^P as possible. In other words, the agent would falsely consider some impossible worlds as possible. The truth value of ϕ at an impossible world w is determined by $C(w)$. Specifically, ϕ is true at $w \in W^I$ if and only if $\phi \in C(w)$. Thus,

$$\text{if } w \in W^P, \text{ then } (M, w) \models p \text{ iff } \pi(w)(p) = \mathit{true}$$

$$\text{if } w \in W^I, \text{ then } (M, w) \models \phi \text{ iff } \phi \in C(w)$$

$$\text{if } w \in W^P \cup W^I, \text{ then } (M, w) \models \Box\phi \text{ iff } (M, w') \models \phi \text{ for all } w' \text{ such that } wRw'$$

Different from those in a possible world, the true formulas in an impossible world are not determined recursively. Instead, the true formulas, atomic or composite, were stipulated by the function C. As a result, the set of true formulas at each impossible world is not logically closed and the agent's knowledge is not necessarily closed under logical consequence since it could well be the case that there exists an impossible world w' accessible from the world w such that the formulas $\varphi_1, \varphi_2, ..., \varphi_k$ but not their logical consequence ψ is true in w'. Therefore, there is no guarantee that we could infer $\Box\psi$ solely based on $\Box\varphi_1, \Box\varphi_2, ..., \Box\varphi_k$. To illustrate this with an example, see the contrast between the possible worlds structure and the impossible worlds structure in Figure 1.

[3] This example assumes that a move against the rule is impossible in this game. Imagine as if the agent is playing with a rule-binding computer.

Example 1. *In the possible worlds structure, p is true in w_1 and q is true in w_2. Since the possible worlds are logically closed, $p \vee q$ is true in both worlds and we could conclude that $(M, w_1) \vDash \Box(p \vee q)$ given the relations depicted in Figure 1. In the impossible worlds structure in Figure 1, we include an impossible world w_3 in which the true formulas are $p, q, p \wedge q$. The formula $p \vee q$, though being a logical consequence of P, is not assigned to w_3 as a true formula. Therefore, $(M, w_1) \nvDash \Box(p \vee q)$ and the agent fails to know some formulas that she would have known has she been omniscient.*

2 More about the Impossible Worlds Approach

2.1 The Plausibility Model

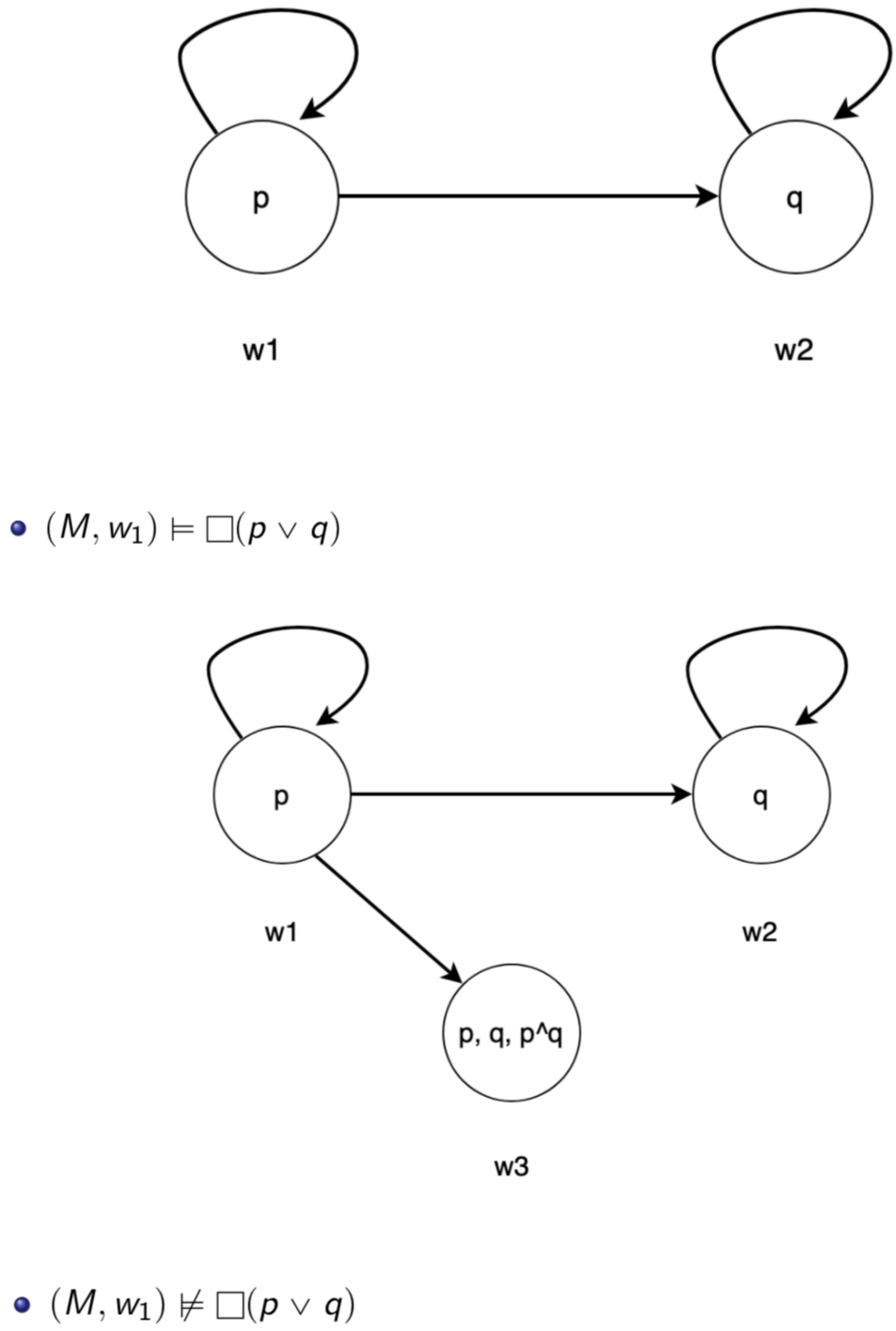
These two approaches have the advantage of keeping the possible worlds structure but one main source of logical omniscience is the lack of computation resources for boundedly rational agents. The quantitative measures of resources are missing in these models. In order to solve this problem, Solaki et al. (2021) use techniques from dynamic epistemic logic subject to cognitive cost constraint and combine them with the impossible worlds semantics[4]. By adding modal operators that describe model-transforming actions, dynamic epistemic logic enriches the static epistemic logic and enables us to study belief revision (Baltag and Renne, 2016). We could then model epistemic updates: from one plausibility model to a new plausibility model of an agent. Specifically the definition of a plausibility model is as follows:

Definition 1 (Plausibility model of possible worlds(Solaki et al., 2021))**.** *A plausibility model M is a structure $\langle W, \geqslant, V\rangle$ where:*

- *W is a non-empty set of possible worlds.*
- *$\geqslant$ is a locally well-preordered (plausibility) relation on W, such that $w \geqslant u$ reads "w is considered no more plausible than u".*
- *V is a valuation such that each propositional atom from a given set Ψ is assigned to the set of worlds where it is true.*

The plausibility relation $\geqslant$ is transitive and complete. Each world is at least as plausible as itself and for any two possible worlds, at least one of them is at least as plausible as the other. The plausibility relation is also transitive. If $w_0 \geqslant w_1$ and $w_1 \geqslant w_2$, then $w_0 \geqslant w_2$. The plausibility relation is also well-founded, which means that there does not exist an infinite ascending chain. This shows that a set of the most plausible worlds always exists (Baltag and Renne, 2016). A pointed model is an ordered pair (M, w) such that its first element is a model and the second element a designated

[4] Some other attempts to incorporate the bounded resources features include proposing an $n-$ *entailment* relation or a series of knowledge operators to reflect n steps of reasoning (Skipper and Bjerring, 2020; Duc, 2001). Artemov and Kuznets (2006) proposed a Logical Omniscience Test and showed that justification systems are not logically omniscient w.r.t. the evidence-based knowledge.



- $(M, w_1) \models \Box(p \vee q)$

- $(M, w_1) \not\models \Box(p \vee q)$

Figure 1: Examples of the Possible and Impossible World Structures.

world in the model. In our context, the possible world that represents the actual world is usually chosen as the designated world.

2.2 Incomplete vs. Trivially Inconsistent Worlds

We now review some challenges faced by the impossible worlds approach. Bjerring and Schwarz (2017) argue we would face the dilemma between having worlds that are not maximally specific and worlds that are trivially inconsistent when we try to use the impossible worlds approach to model the epistemic status of logically non-omniscient but competent agents.

Specifically, Bjerring and Schwarz (2017) reason as follows: Assume that all the possible and impossible worlds considered by the logically-competent agent are complete. This means that for each sentence $S_i \in \mathcal{L}$, either S_i or $\neg S_i$ is true in $w \in W^P \cup W^I$. It is impossible for such a world to verify only a complex contradiction C but not some trivial contradictions. Some examples of trivial contradictions include when both ϕ and $\neg\phi$ are considered to be true in an impossible world w or when both the premises and the negation of the conclusion of some simple inference rule are considered to be true in the impossible world w. Since C is a contradiction, there is a proof for $\neg C$, which is a finite sequence of sentences, each of which is an axiom (a simple tautology), or follows from the preceding sentences in the sequence by a simple rule of inference: $S_1, S_2, S_3, ..., S_k(= \neg C)$. It is now plausible to assume that the logically-competent agent knows the axioms and can apply the simple rules of inference like *Modus Ponens* as this captures our intuition about the capability of such an agent. Since the world w is complete, for each S_i, either S_i or $\neg S_i$ is true in w. In order for w to verify C, there are only three cases:

1. The negation of some simple tautology is true in w.
2. Both the premises and the negation of the conclusion of some simple inference rule are true in w.
3. Both C and $\neg C$ are true in w.

In all three cases, w has to be a trivially inconsistent world. This concludes that maximally specific impossible worlds cannot model the epistemic status of logically competent but non-omniscient agent.

It seems that a potential solution is to drop the assumption that these worlds have to be maximally specific. By excluding some of the sentences and its negations from w, we could have a world that does not fall into any of the three cases. For example, let w_1 be an impossible world such that $C(w_1) = \{A, A \rightarrow \neg B, A \wedge B\}$. Then w_1 avoids the problem by not including B or $\neg B$ in $C(w_1)$. In contrast, an example of its maximally specific counterpart could be a world w_1' such that $C(w_1') = \{A, A \rightarrow \neg B, A \wedge B, B...\}$ which is trivially inconsistent and cannot be considered possible by a logically-competent agent. However, Bjerring and Schwarz (2017) argue that a non-complete world like w_1 should also not be considered as possible by the agent. Specifically, they argue that it goes too far to assume that the agent is skeptical about bivalence just because she is non-omniscient. In other words, we should still assume that the agent, though may be uncertain about which one of the sentences $S_i, \neg S_i$ is true, is still certain that one and only one of them is true for any sentence S_i. If so, the agent should not consider possible any incomplete world like w_1 since it verifies neither S_i nor $\neg S_i$ for some S_i.

In response, I would argue that we do not have to interpret the agent as being skeptical about bivalence just because she assigns some credence to an incomplete world like w_1. Here I take insight from the awareness approach of dealing with logical omniscience problems and claim that a more plausible interpretation for a logically competent agent believing in an incomplete world like w_1 is that the agent is simply not aware of some sentences given her bounded resources. If we assume that the agent has finite memory space, it would be too demanding for the agent to consider each of the element of an infinite enumeration of sentences. In epistemic logic that uses normal modal logic, besides the implausible deductive closure property, it is also implausible to assume that each of the worlds has to be maximally specific and it is another property that we want to avoid while modelling the epistemic status of boundedly rational agent.

Therefore, the decision to drop the assumption that the impossible worlds need to be maximally specific is not only acceptable but also required by our assumptions about the limited cognitive resources of boundedly rational agent. Since the incompleteness of impossible worlds makes complex contradictions possible, we could have impossible worlds that exclude trivial inconsistency but not complex contradiction. This assures us that we could continue on our approach to use impossible worlds to model logically non-omniscient but competent agents.

3 The Logic of Fast and Slow Thinking and its probabilistic extensions

This part will sketch a framework for the probabilistic extensions of the logic of Fast and Slow thinking. Before we start, I outline this logic and explain why a probabilistic extension of it is desirable. As introduced in section 2.1, a model in the logic of fast and slow thinking is a plausibility model with impossible worlds. Since the plausibility relation $\geqslant$ is transitive, complete and well-founded, an ordinal ranking of the possible worlds can be generated. Let ord be a function that assigns an ordinal number to each world. The smaller the ordinal number for a world is, the more plausible the world. The accessibility relation in modal logic is thus replaced by the plausibility relation. In a pointed model (M, w), the agent knows ϕ ($\Box\phi$) iff ϕ is true in all worlds that are at least as plausible as w. In addition, there are two types of model-changing operators: the fast updater and the slow controller. The employment of the fast updater expresses System 1's way of thinking and is assumed to incur no cost, while the employment of the slow controller expresses System 2's way of thinking and would incur some cognitive cost. Each agent has a tuple $cp \in \mathbb{N}^r$ indicating the amount of each type of resources she has and there is also a cost function indicating the cost for each model-changing action. This would enable us to capture the intuition about the limited computation resources of boundedly rational agent, an important source of non-omniscience [5].

The logic of fast and slow thinking provides us with a promising framework to model the epistemic status of moderately rational agents by impossible worlds and model transformation operators. However, the problem of logical omniscience also haunts probability theory and decision theory. Let $Cr(A)$ be an agent's credence over a proposition A, Skipper and Bjerring (2020) summarised the main sources of logical omniscience in orthodox Bayesianism as the following:

- Classical preservation: For any two propositions, A and B, if A logically entails B, then $Cr(A) \leqslant Cr(B)$.
- Classical normality: For any tautology T, $Cr(T) = 1$.

It can be checked that the problem of logical omniscience in normal modal logic is the special case in which the agent is certain about the propositions. For example, if $Cr(A) = 1$ and A logically entails B, then $Cr(B) = 1$ given some constraints about Cr.

Since the logic is qualitative, it only enables us to talk about the ordinal ranking of the plausibility of worlds. As a result, this logic has limitations when we want to describe the degrees of belief of moderately rational agents. Therefore, we intend to study the probabilistic extension of the logic of fast and slow thinking to model the doxastic status of non-logically omniscient agents. By doing so, we could also seek a way to tackle the logical omniscience problem in probability theory and decision theory. We now introduce the syntax and the semantics of the extended logic.

3.1 Syntax

The language of the probabilistic extensions of the logic of fast and slow thinking can be derived by replacing the formulas $\Box\phi$ (which reads "the agent knows that ϕ") and $B\phi$ (which reads "the agent believes that ϕ") with the formulas $Cr_a(\phi)$ to reflect the credence that the agent assigns to ϕ.

Definition 2 (Language)**.** *Given a set P of propositional atoms and a set of inference rules R available to the agent, the language $\mathcal{L}$ is inductively defined from:*

$\phi ::= p|\neg\phi|\phi \wedge \phi|Cr_a(\phi)|[\alpha]\phi$

where

- $p \in P$
- $a \in [0, 1]$
- $Cr_a(\phi)$ *reads "the agent believes to degree a that ϕ"*

[5] A detailed introduction of the logic can be found in Solaki et al. (2021).

- $[\alpha]$ *is a model-changing update operator that expresses a model-changing action performed by the agent. I will mention the two kinds introduced by Solaki et al. (2021):*

  1. $[\Psi \Uparrow]\phi$ *reads "after upgrading with* Ψ, ϕ *is true". This operator describes the cognitively costless action of fast upgrade given incoming information in System 1 way of thinking.*
  2. $\langle R_k \rangle\phi$ *reads "after some application of inference rule* R_k, ϕ *is true". This operator describes the cognitively costly action of conducting logical reasoning in System 2 way of thinking.*

3.2 Semantics

3.2.1 Dual Process Probabilistic model

In this framework, we get a dual process probabilistic model from a dual process plausibility model[6] by replacing *ord* with a probability distribution Pr.

Definition 3 (Dual Process Probabilistic model)**.** *Let* R *indicate the set of inference rules available to the agent and* Res *indicate a finite set of computation resources. Define* $r ::= |Res|$ *as the number of resources. A dual process probabilistic model* M *is a structure* $\langle W^P, W^I, Pr, V, C, cp\rangle$ *where:*

- W^P *is a non-empty set of possible worlds and* W^I *a set of impossible worlds.*
- $Pr : W \rightharpoonup [0, 1]$ *is a probability function on a non-empty subset of* $W = W^P \cup W^I$*. Notice that this is a partial function and not all worlds may be included in the domain.*
- $V : W \to \mathcal{P}(\mathcal{L})$ *assigns a set of sentences to each* $w \in W$*. Specifically, it assigns the set of atomic formulas true at* w *to each* $w \in W^P$*. It assigns all atomic or composite true formulas to each* $w \in W^I$*. The truth values of the sentences in impossible worlds are not determined recursively, but assigned directly. This means that the set of formulas in the impossible worlds need not be deductively closed. However, some restrictions are still imposed on it to avoid trivial inconsistency since we want to model the doxatic status of logically-competent agents. These restrictions will be discussed below.*
- $C : R \to \mathbb{N}^r$ *is a function that assigns a cognitive cost to each inference rule.Note that when* $r > 1$*, there are more than one type of resources (time, space, etc.)*
- $cp \in \mathbb{N}^r$ *indicates the agent's cognitive capacity or computation resources.*

As introduced above, a pointed dual process probabilistic model (M, w) is one in which M is a dual process probabilistic model and w a designated world that represents

[6] The dual process plausibility model introduced by Solaki et al. (2021) is an extension of the plausibility model and is very similar to the dual process probabilistic model introduced below except it has an ordinal ranking function *ord* instead of a probability function Pr.

the actual world. Now we turn to the restrictions imposed on the set of sentences in any impossible world. In order to model the doxatic status of logically competent agent, we follow Solaki et al. (2021) and include a Minimal Consistency (MC) requirement. Specifically, it means that $\{\phi, \neg\phi\} \nsubseteq V(w)$ for all $w \in W^I$. This suggests that no trivial contradiction is included in any impossible world since an agent who cannot exclude an impossible world like this would be too logically incompetent.

3.2.2 Semantic Clauses

The semantics of the probabilistic logic can be derived by replacing the $\Box\phi$ and $B\phi$ formulas with $Cr_a(\phi)$.

Definition 4 (semantics)**.** *The semantics of formulas in possible worlds are defined inductively while the semantics of formulas in impossible worlds are not.*

For $w \in W^P$:

- $M, w \vDash p$ *iff* $p \in V(w)$
- $M, w \vDash \neg\phi$ *iff* $M, w \nvDash \phi$
- $M, w \vDash \phi \wedge \psi$ *iff* $M, w \vDash \phi$ *and* $M, w \vDash \psi$
- $M, w \vDash \Box\phi$ *iff* $M, w' \vDash \phi$ *for all* w' *such that* $Pr(w') \geqslant Pr(w)$
- $M, w \vDash Cr_a(\phi)$ *iff* $Pr([[\phi]]_M) = a$*, where* $[[\phi]]_M = \{s \in W : M, s \vDash \phi\}$
- $M, w \vDash [\Psi \Uparrow]\phi$ *iff* $M^{\Psi\Uparrow}, w \vDash \phi$
- $M, w \vDash \langle R_k \rangle\phi$ *iff* $M', w \vDash \phi$ *for some* (M', w) *which is* R_k*-accessible from* (M, w)*. A pointed model* (M', w) *is* R_k*-accessible from* (M, w) *if the pointed model* (M, w) *can be transformed into* (M', w) *after applying the particular rule of inference* R_k [7]*.*

For $w \in W^I$:

- $M, w \vDash \phi$ *iff* $\phi \in V(w)$

It can be seen that the credence over a particular proposition is the sum of the probabilities assigned to all the worlds in which the proposition is true. Compared to the qualitative logic of fast and slow thinking, the probabilistic extension is indeed more expressive. We can compare the cardinal values of the plausibility of worlds and talk about an agent's degrees of belief over propositions. Now we can introduce the model transformations of the probabilistic extension.

[7] The details about how these two model-changing operations work will be explained in more detail in the next section of model transformations

4 Model Transformations

4.1 The Fast Updater

The fast updater is a system-1 action of integrating new information. Given a piece of new information expressed as a sentence Ψ, the system incorporates Ψ by prioritizing the worlds satisfying it (Solaki et al., 2021). We include the impossible worlds to describe situations in which Ψ but not all its logically equivalent propositions are true. This setting is consistent with some findings in behavioural science. For example, the framing effect can happen when humans' decisions are affected by the way the options are presented. Tversky and Kahneman (1985) find that a majority of participants would choose the option presented with positive framing (saving 200 people) while much smaller percentage of participants chose the option presented with negative framing (400 people will die) even though these two options are logically equivalent. In our model, this feature is captured by the impossible worlds since a fast upgrade with Ψ does not ensure that the worlds in which the logical equivalence of Ψ is true are also prioritised. In addition, this action is assumed to induce no cognitive costs and the model after the fast update is not a result of cognitively costly stepwise deductive reasoning. Specifically, the definition of this action is shown below:

Definition 5 (Probabilistic model transformation by a System 1 upgrade)**.** *Given a probabilistic model: $M = \langle W^P, W^I, Pr, V, C, cp \rangle$, the new model after $\Psi \Uparrow$ transformation is $M^{\Psi\Uparrow} = \langle W^P, W^I, Pr^{\Psi\Uparrow}, V, C, cp \rangle$ where $Pr^{\Psi\Uparrow}$ is a member of the following set of functions $\{f : W \to [0,1] |$ for any $w, u \in W : f(w) \geqslant f(u)$ iff $(w \in [[\Psi]]$ and $u \notin [[\Psi]])$ or $(w \in [[\Psi]] \wedge u \in [[\Psi]]$ and $Pr(w) > Pr(u))$ or $(w \notin [[\Psi]]$ and $u \notin [[\Psi]]$ and $Pr(w) > Pr(u))\}$.*

This shows that, after the fast updater, all Ψ-worlds become more doxatically plausible to the agent than all $\neg\Psi$-worlds, and the relative ordering remains intact within the two zones. For example, if Ψ is true in both worlds w and u, then w and u are "in the same zone" and w is considered to be at least as plausible as u after the fast update if and only if w is considered to be at least as plausible as u before the update. This action of upgrade corresponds to the Lexicographic update in dynamic epistemic logic Van Benthem (2007). The agent implementing this upgrade strongly believes the source of the information because she thinks the source is highly reliable but still fallible.

4.1.1 Using KL-Divergence to restrict the probabilities in the new model

One challenge faced by the probabilistic approach is to specify what further restrictions need to be imposed on the new probability function. In the qualitative model, only the ordinal rankings of the plausibility matter. In contrast, our probabilistic settings carry more information since more than one probabilistic assignments can be reduced to the same ordinal numbers assignment. This requires us to propose more constraints on the probability functions. In the qualitative model of Solaki et al. (2021), the ordinal rankings change in a way such that the Ψ-worlds become more plausible than non-Ψ ones and the previous ordering remains intact within the two zones. This is because the fast updater requires no effort and the model should therefore be changed in a minimal way.

Similarly, we want the probability distribution in the new model to be the closest to the probability distribution in the original model given the constraints described above. A measure of probability divergence may serve the purpose. For example, one can use the concept of Kullback-Leibler Divergence to restrict the probability in the new model.

Definition 6 (The Kullback-Leibler Divergence Kullback and Leibler (1951))**.** *Let $S_1, S_2, ..., S_n$ be the possible values of a random variable S over which probability distributions P' and P are defined. The Kullback-Leibler divergence between P' and P is then given by $D_{KL}(P'||P) := \sum_{i=1}^{n} P'(S_i) log \frac{P'(S_i)}{P(S_i)}$.*

In our context, P is the probability distribution in the old model while P' is the probability distribution in the new model. The new probability P' should be one that satisfies two requirements:

1. It is a function from the set described in Definition 6.

2. $D_{KL}(P'||P)$ is minimized given the constraint in 1.

Requirement 2 ensures that the new probability distribution is the closest probability distribution given requirement 1.

Example 2. *For example, in Figure 2, (M, w_2) is the original model and $(M^{p \to r \Uparrow}, w_2)$ is the model derived after a fast upgrade of $p \to r$. In both models, w_2 is the possible world, while w_0 and w_1 are the impossible worlds. Since w_2 is logically closed, only the true atomic propositions are shown and all the logical consequence of the set $\{p, q, r\}$ are true in w_0. Only the listed formulas are true in the impossible world: $V(w_0) = \{p, p \to q, q\}$ and $V(w_1) = \{p, p \to q, q, p \to r\}$. Same as the model transformation in qualitative models, we require all the worlds in which $p \to r$ is true to be brought to the more plausible end of the ranking. This means that in the new model, the plausibility ranking would change from $w_2 \geqslant w_1 \geqslant w_0$ to $w_0 \geqslant w_2 \geqslant w_1$. Notice that the bigger the ordinal ranking is, the less plausible the world is. This shows that we would expect the probability assignment to be one such that $P'(w_1) \geqslant P'(w_2) \geqslant P'(w_0)$. Given this constraint on the probabilities of each world in the new model, we try to minimise the KL-divergence between the new probability distribution P' and the original probability distribution P where $P(w_0) = 1/2, P(w_1) = 1/4, P(w_2) = 1/4$. With some calculations, we get the probability distribution P' such that $P'(w_0) = 1/3, P'(w_1) = 1/3, P'(w_2) = 1/3$ would minimise the divergence. Therefore, we conclude that this P' is the probability distribution in the new worlds. As a result the agent's credence in the formula $p \to r$ increases from $1/2$ to $2/3$.*

There are some potential problems faced by this KL-divergence approach. Firstly, the new probability distribution is not sensitive to the priors in some cases. For example, consider a model that just have two worlds: $W = \{w_0, w_1\}$. Let the probability distribution before the fast updater be one such that $P(w_0) > P(w_1)$. Assume that w_0 is a non-Ψ world while w_1 is a Ψ world. Then after the $\Psi \Uparrow$ transformation, the new probability distribution should be one such that $P'(w_0) \leqslant P'(w_1)$. Then it is easy to show that the probability function that satisfies this restriction and minimizes $D_{KL}(P'||P)$ is $P'(w_0) = P'(w_1) = 1/2$. Similar result holds for all situations where the

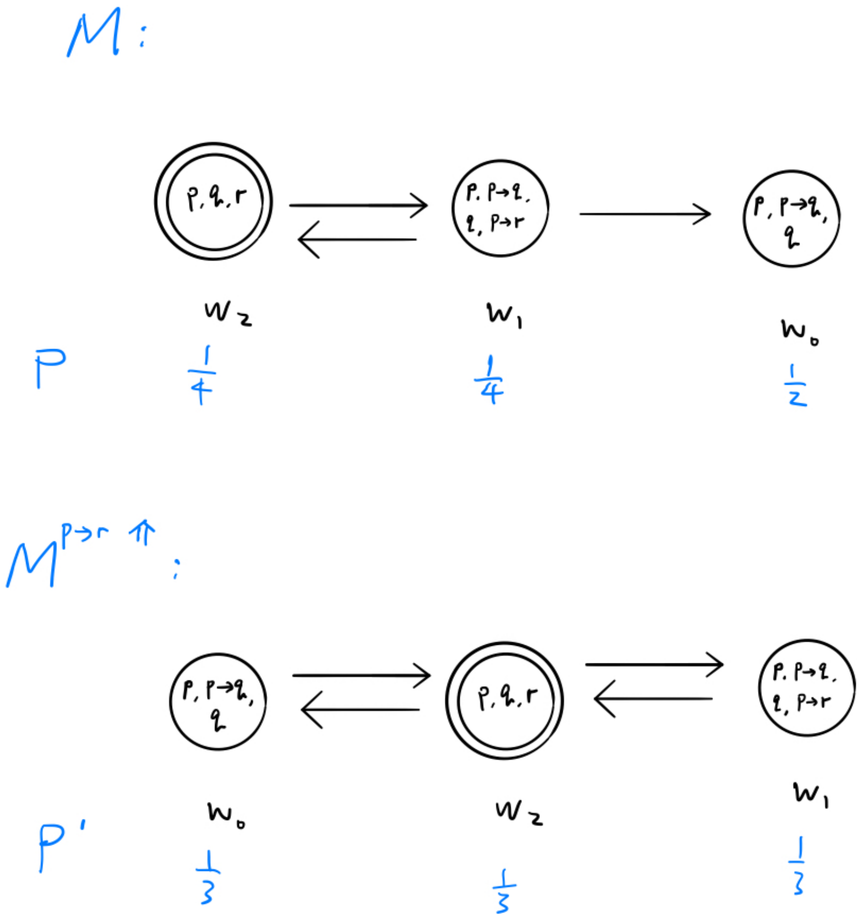


Figure 2: An Example of the Fast Updater.

order of the worlds are switched.

Secondly, some may wonder whether this is a plausible rule of transformation given the assumption that the fast updater requires no effort. It may take a lot of effort to calculate the new probability distribution that satisfies the requirement about Ψ and minimizes the KL-divergence at the same time. How can we expect an agent to derive the new probability model without consuming any computation resources? The same problem applies to other distance and divergence measures as well, e.g., the Hellinger distance, the total variation distance, and the earth mover's distance.

In terms of the first problem, we could modify the divergence measure to make the new probabilities more distinct from each other. One could also gain insight from experiments in behavioural sciences and modify the measure so that it is in accordance with the experiment results. In terms of the second problem, we respond by saying that such fast update rules are supposed to be an approximation of the real update. It may be costly to calculate the exact numbers that correspond to the measure but it may be possible to get an approximate result with a very small cost. Also, we choose to minimise the KL divergence because we think that it reflects the intuition that we want to make the minimum effort required to incorporate the new information.

4.2 The Slow Controller

This part introduces the action of slow controller that requires cognitive cost to conduct inference rules, which reflects the features of the system 2 way of thinking. We start by introducing the notion of R_k-expansion and R_k-accessibility in the probabilistic context, which is adapted from the R_k-accessibility relation introduced by Solaki et al. (2021).

Let $V^*(u)$ [8] denote the set of true formulas in u. A world u is an R_k-expansion of v if $V^*(u) = V^*(v) \cup \{con(R_k)\}$ and the minimal consistency requirement is not violated in u, where $con(R_k)$ is a particular conclusion of R_k derived from one particular set of premises $pr(v) \subseteq V^*(v)$. If $pr(v) \nsubseteq V^*(v)$ for any R_k, we take the only R_k expansion to be itself. For a world w, if the minimal consistency requirement is violated in all the resulting worlds after an application of R_k, then w does not have any R_k-expansion. We illustrate the idea will an example.

Example 3. *In Figure 3, there are three impossible worlds w_1, w_2, w_3 and one possible world w_4. The black arrows indicate the plausibility relation while the blue dotted arrows indicate the MP-expansion (modus ponens expansion) relation. Here we explain some of the MP-expansion relations and the rest can be left as an exercise. w_2 is a MP-expansion of w_1 because q is a conclusion of modus ponens with the the set of premises being $\{p, p \rightarrow q\} \subseteq V^*(w_1)$ and $V^*(w_2) = V^*(w_1) \cup \{q\}$. w_3 is also a MP-expansion*

[8] Note that V^* is different from the function V introduced in Definition 3: When $w \in W^P$, $V(w)$ includes only the atomic formulas true at w but $V^*(w)$ includes all the true formulas, atomic or composite. When $w \in W^I$, they both include all the true formulas, atomic or composite.

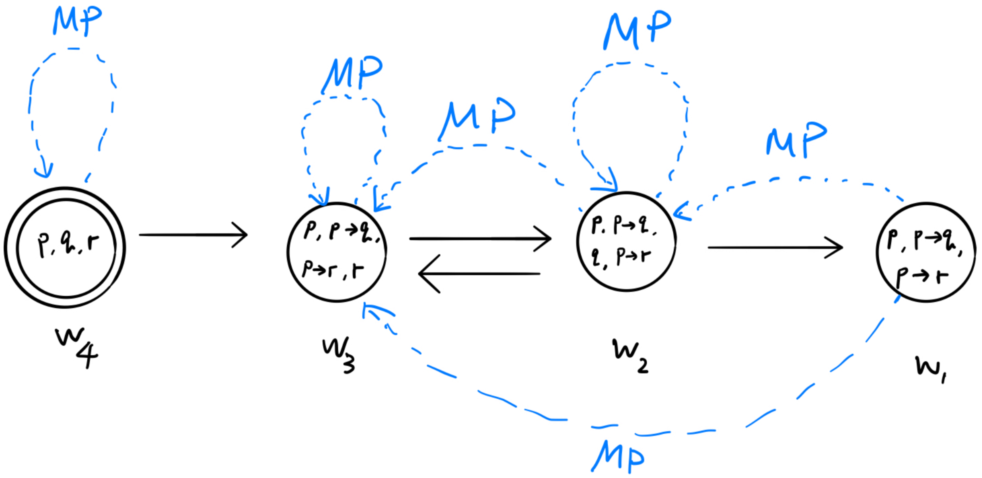


Figure 3: An Example of the Fast Updater.

of w_1 because r is a conclusion of modus ponens with the the set of premises being $\{p, p \to r\} \subseteq V^*(w_1)$ and $V^*(w_3) = V^*(w_1) \cup \{r\}$. The only MP-expansion of w_4 is itself because a possible world is deductively closed and all the conclusions of a particular rule of inference are true in this possible world. Therefore, for any rule of inference R_k, the only R_k-expansion of a possible world w is w itself. We use v^{R_k} to denote the set of R_k-expansions of the world v. In this example, $w_1^{MP} = \{w_2, w_3\}$ and $w_4^{MP} = \{w_4\}$.

Let $P_{\geqslant}(w) := \{u \in W | Pr(u) \geqslant Pr(w)\}$ be the set of worlds assigned a probability at least as big as that of w in a pointed model. A pointed probabilistic model (M', w) is R_k-accessible from a pointed probabilistic model (M, w) if it can be generated by replacing each world $w' \in P_{\geqslant}(w)$ with a R_k-expansion of w, while the probabilities assigned to each world are adapted accordingly. The detail of the transformations is illustrated in the following steps:

1. Given a set of non-empty sets A, we call a set constructed by selecting exactly one element from each of the element in A a choice. $\mathcal{C}$ is a choice function whose input is a set of sets A and returns the set of all choices constructed from A. For example, $\mathcal{C}(\{\{w_0\}, \{w_1, w_2\}\}) = \{\{w_0, w_1\}, \{w_0, w_2\}\}$ and each element ($\{w_0, w_1\}$ or $\{w_0, w_2\}$) in $\mathcal{C}(\{\{w_0\}, \{w_1, w_2\}\})$ is called a choice. Let (M, w) be a pointed model. Let $P^{R_k}(w) := c$ where c is some choice in $\mathcal{C}(\{v^{R_k} | v \in P_{\geqslant}(w)\})$. Each element in the set $P^{R_k}(w)$ is an R_k-expansion of the worlds which were assigned with a probability at least as high as that of w in the original pointed model.

2. The set of worlds in the R_k-accessible pointed model is $W^{R_k} = W \backslash \{u \in P_{\geqslant}(w) | u \notin P^{R_k}(w)\}$. This shows that the worlds that were assigned with a probability higher

than that of w in the original pointed model and are not in the choice of R_k-expansion of worlds are excluded from the new model.

3. This part describes the new probability assignment function Pr^{R_k}:
   - If $u \notin P_{\geqslant}(w) \cup P^{R_k}(w)$, then $Pr^{R_k}(u) = Pr(u)$. This shows the worlds that were assigned with a probability lower than that of w and were not in the choice of R_k-expansion of worlds are assigned with the same probability as they were in the original model.
   - If $u \in P_{\geqslant}(w)$ and $u \notin P^{R_k}(w)$, this case has already been discussed in step 2: These worlds are excluded from the model and do not have a probability assigned to them.
   - If $u \in P^{R_k}(w)$ and $u \in P_{\geqslant}(w)$, then $P^{R_k}(u) = \Sigma Pr(v_i)$ where v_is are the worlds in the original model that R_k-expanded to u. Note that this assumes a given choice of the R_k-expansions of the worlds.
   - If $u \in P^{R_k}(w)$ and $u \notin P_{\geqslant}(w)$, then $P^{R_k}(u) = \Sigma Pr(v_i)$ where v_is are the worlds in the original model that R_k-expanded to u and the worlds (if there are any) in the original model that are identical to u but were assigned with a probability smaller than that of w.

4. After the application of the rule R_k, $cp^{R_k} := cp - C(R_k)$. The application of the inference rule consumes the corresponding cognitive cost.

One major difference between this slow controller transformation and that proposed by Solaki et al. (2021) is the way the probability/plausibility is decided in the new model after transformation. In the qualitative model, when a world $u \in P^{R_k}(w)$ and there are some world $v \in P_{\geqslant}(w)$ such that $u \in v^{R_k}$ for the particular choice c, the ranking of u in the new model is simply the ranking of the most plausible world from which it originated in the old model. This strategy is due to the ordinal ranking of the plausibility of worlds. By replacing the ordinal plausibility ranking with cardinal probability, we compute the probability of world u in the new pointed model as the sum of the probability of those worlds from which it originated.

Example 4. *Assume that there are three impossible worlds and one possible world that are considered by an agent. w_0, w_1, w_3 are impossible worlds while w_2 is a possible world. In the possible world, $w_2 = \{p, q, r\}$. In the impossible worlds, all the true sentences are listed. This means that $V(w_0) = \{p, p \rightarrow q, q\}$, $V(w_1) = \{p, p \rightarrow q, q, q \rightarrow r\}$, $V(w_3) = \{p, p \rightarrow q, q, q \rightarrow r, r\}$.*

As shown in Figure 4, the black arrows indicate the plausibility relation. For example, a black arrow from w_1 to w_0 indicates that w_0 is at least as plausible as w_1 to the agent. The double arrows between w_1 and w_2 indicates that these two worlds are equally plausible to the agent. Actually not all the plausibility relations are shown since the rest of them can be deduced from what are shown in Figure 4. To illustrate this point, the full model is shown in Figure 5. Since the plausibility relation is reflexive and transitive, it can be seen that the full model in Figure 5 can indeed be derived from that in Figure 4.

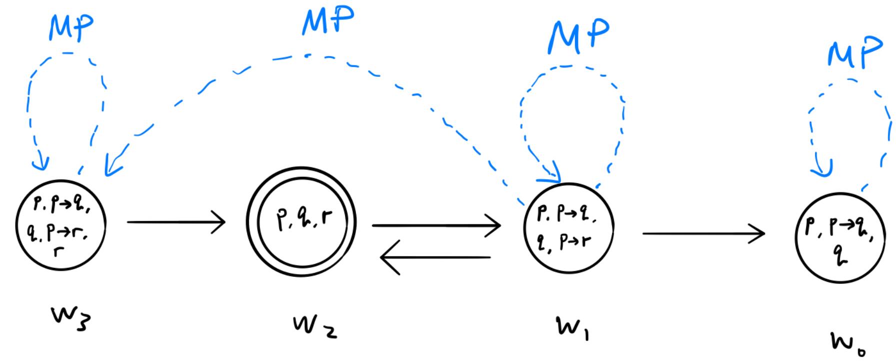


Figure 4: The Probabilistic Model for Example 4.

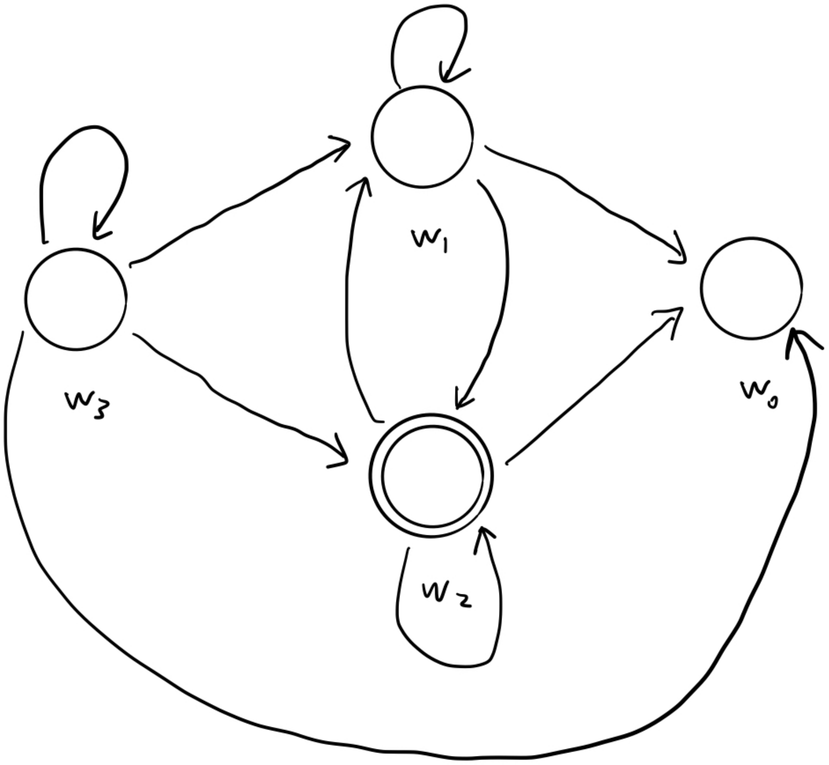


Figure 5: The Full Probabilistic Model for Example 4.

The blue arrows indicate the MP-expansion relations. For example, the blue arrow from w_1 to w_3 indicates that w_3 is an MP-expansion of w_1. This is because the agent can apply Modus Ponens to p and $p \to r$ and get r. The MP-expansion relation for the possible world w_2 is not drawn since w_2 is a possible world and the only MP-expansion of w_2 is itself.

There is also a probability assignment to both the possible and impossible worlds. Let $Pr(w_0) = 30\%$, $Pr(w_1) = Pr(w_2) = 25\%$, $Pr(w_3) = 20\%$. Note that the sum of the probability equals 1 and the ranking of the numbers assigned to each world is in line with the plausibility relation between them: the black arrow is pointing from each world to all the worlds (including itself) at least as plausible as itself. The credence assigned to each of the propositions can also be derived from the probability distribution defined over worlds. For example, the agent's credence over the proposition $p \to r$ is 0.7 ($Cr_{0.7}(p \to r)$) because the proposition $p \to r$ is true only in worlds w_1, w_2, w_3. The sum of these worlds' probabilities is $Pr([[p \to r]]_M) = Pr(w_1) + Pr(w_2) + Pr(w_3) = 0.25 + 0.25 + 0.2 = 0.7$.

Now I will show how does this model change after the action of slow controller. Specifically, we assume that the agent only applies Modus Ponens in this example. The overall cognitive resources the agent has is $cp_0 = (10, 8)$ and the cognitive cost of Modus Ponens is $C(MP) = (2, 3)$. After each application of Modus Ponens the cognitive resources will be reduced. For example, after the first application of Modus Ponens, the cognitive resources will change to $cp_1 = cp_0 - C(MP) = (10, 8) - (2, 3) = (8, 5)$.

Given the pointed model (M, w_2), we start by computing the set of the MP-expansions of the worlds considered at least as plausible as w_2. $\{v^{MP} | v \in P_{\geqslant}(w_2)\} = \{\{w_2\}, \{w_1, w_3\}, \{w_0\}\}$. This shows that there are two choices of the MP-expansions of worlds because worlds w_0 and w_2 have only one MP-expansion and w_1 has two MP-expansions and $1 \times 1 \times 2 = 2$ choices are available. In the first case, $P^{MP}(w_2) = \{w_2, w_1, w_0\}$. In the second case, $P^{MP}(w_2) = \{w_2, w_3, w_0\}$.

1. *In the first case, $w_3 \notin P_{\geqslant}(w) \cup P^{R_k}(w)$, then $Pr^{R_k}(w_3) = Pr(w_3) = 20\%$. $u \in P^{R_k}(w)$ and $u \in P_{\geqslant}(w)$ holds for all $u \in \{w_0, w_1, w_2\}$, then $P^{R_k}(u) = \Sigma Pr(v_i) = Pr(u)$ in this example since the MP-expansions of each world are themselves. The new pointed model is identical with the original model (M, w_2) and the probabilities of each world remain the same.*

2. *In the second case, $u \in P^{R_k}(w)$ and $u \in P_{\geqslant}(w)$ holds for all $u \in \{w_0, w_2\}$, then $P^{R_k}(u) = \Sigma Pr(v_i) = Pr(u)$. $P^{R_k}(w_0) = Pr(w_0) = 30\%$, $P^{R_k}(w_2) = Pr(w_2) = 25\%$. $w_1 \in P_{\geqslant}(w)$ and $w_1 \notin P^{R_k}(w)$, w_1 is excluded from the model and does not have a probability assigned to it [9]. $w_3 \in P^{R_k}(w)$ and $w_3 \notin P_{\geqslant}(w)$, then $P^{R_k}(w_3) = \Sigma Pr(v_i) = Pr(w_1) + Pr(w_3) = 45\%$ where v_i are the worlds in the original model*

[9] Whether the world does not have a probability assigned to it or has a probability of 0 needs further discussions. Intuitively, a probability of 0 means that the agent is aware of the proposition and considers it as impossible while no probability assignment means that the agent is simply not aware of it. A reason in support of the choice of removing the world from the domain instead of assigning a probability 0 to it may be that the agent is subject to space (memory) constraint and she tends to

that R_k-expanded to u and the worlds (if there are any) in the original model that are identical to u but were assigned with a probability smaller than that of w.

5 Philosophical Implications

In this section, we discuss the philosophical implications of the proposed framework. We will do so by reviewing the Dutch Book argument from the perspective provided by the framework sketched above. By doing this, we aim to shed some light on the discussions of reasoning with bounded resources in a probabilistic context.

5.1 The Dutch Book Argument

The Dutch Book argument has been used to show that a rational agent's degrees of belief should behave like probabilities (satisfy Kolmogorov axioms) because otherwise they would be subject to a sure loss. It seems that the type of rational agents assumed in this argument is logically omniscient in the sense that they have unlimited computation resources and do not make any mistakes in the process of deductive reasoning. For example, the Normalization condition of Kolmogorov axioms can be interpreted as saying that a rational agent should always assign a credence of 1 to a tautology T ($Cr(T) = 1$). This is true in our framework when there are only possible worlds involved. Since any tautology T is true in all possible worlds and the sum of the probabilities assigned to all the worlds in the model is 1, a rational agent should always be certain about a tautology T. Similarly, it can also be derived from the axioms that for any two propositions A and B, if $A \rightarrow B$, then $Cr(A) \leqslant Cr(B)$ since in every possible world in which A is true, B is also true while the reverse is not necessarily true.

It is straightforward to see that there is indeed a correspondence between the concepts in a probability measure space and those from our framework. The intuition is that, for a probability measure space (X, S, p) in which X is the set of elementary events, S the set of general events and p the probability measure function, each elementary event corresponds to a possible world and each general event corresponds to a set of possible worlds in our framework. For example, consider tossing two coins once. Let H_i indicate that coin i lands head and T_i indicate that coin i lands tail. Then the set of elementary events $X = \{H_1 \wedge H_2, H_1 \wedge T_2, T_1 \wedge H_2, T_1 \wedge T_2\}$, while the set of general events $S = \mathcal{P}(X)$. Accordingly, we can construct a possible world structure in which there are four possible worlds and exactly one elementary event is true in each of the worlds. Any general event is simply a set of some possible worlds and the probability function measure corresponds to the credence function. Hence, we could identify the correspondence between the concepts in these two structures[10].

remove these implausible worlds from the set of worlds considered by her. For simplicity, we assume that the the world is removed in this example.

[10] In this paper, we assume that there is a finite number of elementary events and therefore the number of possible worlds $|W|$ and the subsets of the set of all possible worlds $|\mathcal{P}(\mathcal{W})|$ are also finite. The case when $|W|$ is infinite needs further discussions.

The dutch book argument in this case requires a rational agent to assign credence to possible worlds (elementary events) in a way such that for any possible world w_i, $0 \leqslant Cr(w_i) \leqslant 1$ and $\sum\limits_i Cr(w_i) = 1$ and the credence of the general events can be computed recursively according to the Kolmogorov axioms. Actually, different rational agents may assign different credence to the same event, but since their credence both satisfy the axioms, their credence would all be bound by the rules about tautologies and implications introduced above.

However, this assumption about rationality is too demanding for rational agents with bounded resources. For example, if an agent is faced with a tautology, the Dutch Book argument says that she should assign a credence of 1 to it but the problem of deciding whether a propositional formula is a tautology $TAUT = \{\varphi : \varphi \text{ is a formula satisfied by every assignment}\}$, is **coNP**-complete [11]. It is usually assumed that the complexity class **P** includes those problems that can be solved feasibly and a decision problem not in **P** is reckoned to be infeasible to solve. Together with some other assumptions about the relationship between different complexity classes [12], the fact that TAUT is **coNP**-complete suggests that it can be infeasible for a rational agent who has bounded computational resources to decide whether a given formula φ in propositional logic is a tautology [13]. Specifically, as the length of the input formula increases, the required computation resources increase too fast and makes the task infeasible for an agent who has bounded resources. Therefore, a boundedly rational agent may not be certain of a complex tautology. As a result, the agent's degrees of belief are not coherent and the agent can be dutch-booked. However, we should not accuse her for being irrational just because of her lack of resources to compute to get the correct answer, otherwise there exists no rational agents in this world and the idea of rationality is too ideal to have any practical value.

Instead of asking what can be achieved if we have infinite resources, we should ask what can be achieved given the limited resources we have. The probabilistic extension in this paper provides a framework for modelling the situations. The non-logically closed impossible worlds are introduced, which makes it possible for an agent to consider A being true in a world but not B being true even though A entails B. Since impossible worlds may be incomplete, some tautologies may not be true in an impossible world. This enables the agent to assign a credence not equal to 1 to a tautology. One may wonder whether the Dutch Book argument becomes vacuous completely. We would claim that this is not the case. The Dutch Book argument can still be applied in cases when the given resources are sufficient for a rational agent to deduce the correct conclusions. For example, we may make a claim like "given the degrees of complexity of the propositions and the resources a particular rational agent has, it is very likely that she will assign a set of coherent credence to these propositions". In this case, we

[11] Intuitively, **coNP**-complete problems are the "hardest" problems in **coNP** because any other problem in the class can be reduced to it.

[12] Some examples of widely believed assumptions are $\mathbf{P} \neq \mathbf{NP}, \mathbf{NP} \neq \mathbf{coNP}$.

[13] The VALIDITY problem for other logics can be even harder. For example, it is **RE**-compete in first-order logic and **PSPACE**-complete in Intuitionistic Logic and Modal Logic (K,T,S4)(Dean, 2021).

expect the rational agent to not to be dutch-booked. In the case when the propositions are complex or the resources are very limited, we may expect the agent's degrees of belief to be incoherent. In order to talk about such distinctions, we define the concept of $Res-$accessibility using the concept of R_k-accessibility.

Definition 7 ($Res-$accessibility)**.** *For two dual process probabilistic models M_1 and M_k and a set of inference rules R, M_k is Res-accessible from M_1 ($M_1 \rightsquigarrow_{Res} M_k$) iff there exists a sequence of rules of inference from R: $R_1, R_2, ..., R_{k-1}$ such that there exists a sequence of intermediate models $M_2, M_3, ..., M_{k-1}$ with each M_{i+1} (including M_k) being R_i-accessible from M_i[14] and $\sum_i C(R_i) \leqslant Res$.*

The $Res-$accessibility relation enables us to discuss which models are accessible from a particular model given the resources. To illustrate this idea, Figure 6 gives the sketch of an example of the R_k-accessibility relation between different models. Each of the circle in the figure indicates a model. Therefore, the content of each of the circle is something similar to a probabilistic model depicted in Figure 4. The black arrow indicates the corresponding R_k-accessibility relation. For example, M_2 is R_b-accessible from M_1 and M_4 is R_d-accessible from M_2 and R_e-accessible from itself. Figure 7 translates the R_k-accessibility relation in Figure 6 to Res_k-accessibility relation. For example, M_3 is Res_3-accessible from M_1 where Res_3 corresponds to the sum of the cost of the corresponding rules of inference needed in the deduction. Note that more than one Res_k-accessibility relation could exist between two models because there can be different sequences of rules of inference that lead to the same result and some lines of reasoning are more efficient than others. Since the minimum Res required to move from M to M' indicates the minimum effort it takes to switch to the new model, we could use it to indicate how difficult (and also likely) it is for an agent to reach M' from M. In Figure 8, we assume that there are two types of resources (x, y), each vector Res_k indicates the minimum cost required to move from M_1 to M_k while the modulus of the vector can be used to measure how plausible it is for the agent to reach M_k. Note that this has to be done under the strict constraint of the total resources held by the agent. For example, even though $|Res_l| < |Res_k|$, it is possible for the agent to reach M_k but not M_l because Res_k exceeds the limit of the total resources on the dimension of x.

In the context of the dutch book argument, we could now say that it is possible for the agent who has resources Res and whose current credence is depicted in model M to have coherent degrees of belief if there exists a model M' in which the credence assigned to the propositions listed in the betting scenario is coherent in M' and $M \rightsquigarrow_{Res} M'$. Similarly, we could also discuss whether it is highly plausible or not for a particular agent to get coherent degrees of belief. Notice that we may not be able to claim that the agent will definitely get coherent credence because the agent may choose a very inefficient sequence of rules of inference and can therefore not reach the model which would have been easily accessible had she chosen a more efficient sequence of reasoning. An extreme example would be that the agent keeps applying the rule of inference

[14] We omit the designated world which is the same across models in this definition.

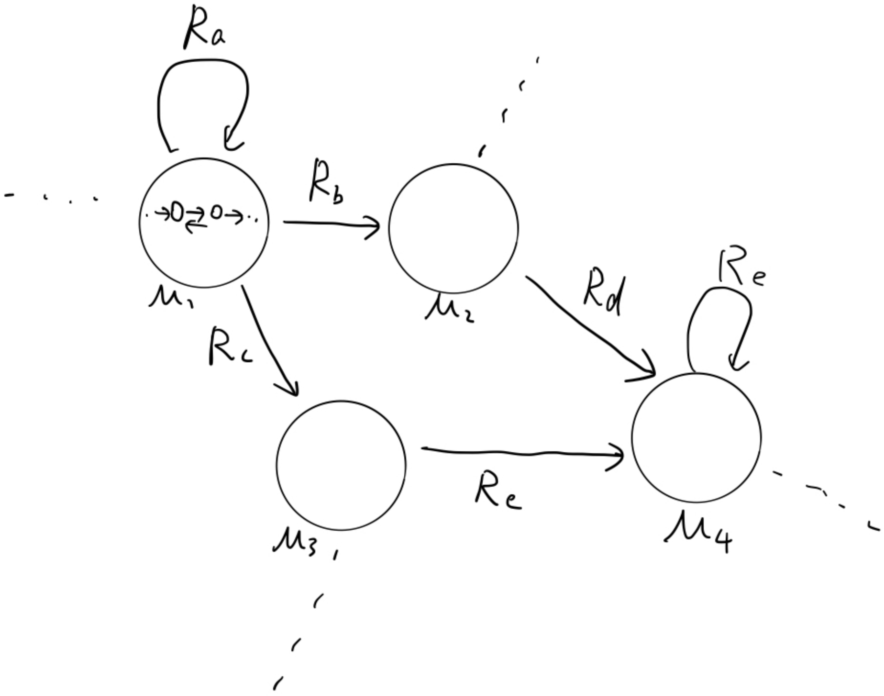


Figure 6: The R_k-accessibility relations.

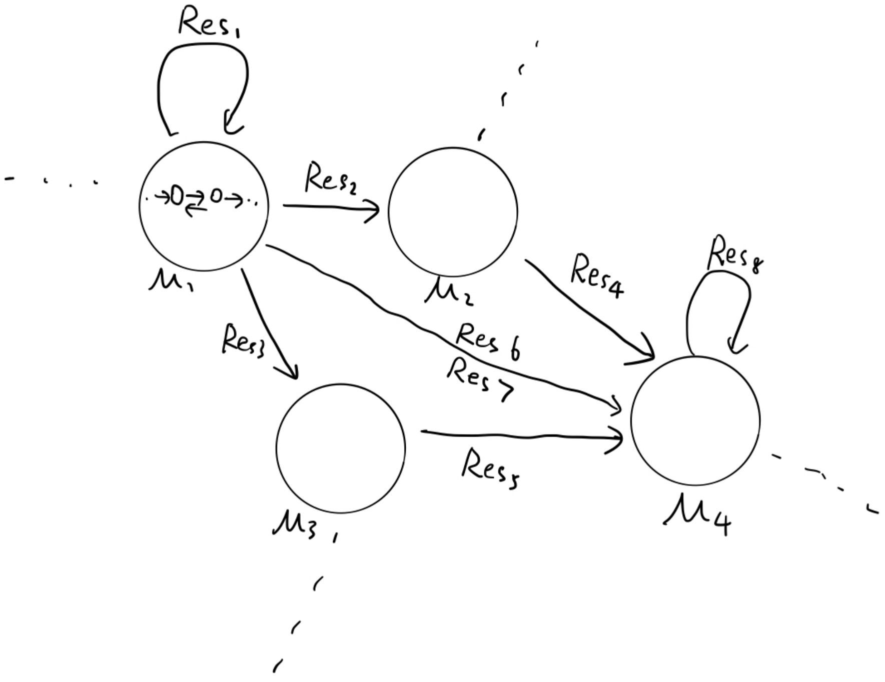


Figure 7: The Res_k-accessibility relations.

that brings her back to the same model, as the reflexive relation of R_a shown in Figure 6.

One may wonder if there is anything more that we could say about the choice of rules of inference at a particular model, which is meaningful if we want to use the model to make predictions about the agent's epistemic status. One naive way to model the model transformations process of an agent is to use random walks. This means that we assume the agent randomly chooses a rule to apply among the rules available to her at a particular node (the model that the agent is currently at). More sophisticated models could be derived if we enrich the model with more interesting features such as assuming a random walk with weighted probabilities or assuming that the agent would avoid a path already taken by her before. As a result, we may be able to come up with a particular probability for each model which is used to claim that after a certain time period, the agent has a probability of p to end up being in model M'.

6 Problems remained for future work

This paper has sketched a framework for the probabilistic extensions of the logic of fast and slow thinking by constructing the probabilistic counterparts to the plausibility model, the syntax and the semantics of the logic. It has also made a response to some of the challenges faced by this approach of using impossible worlds to model non-

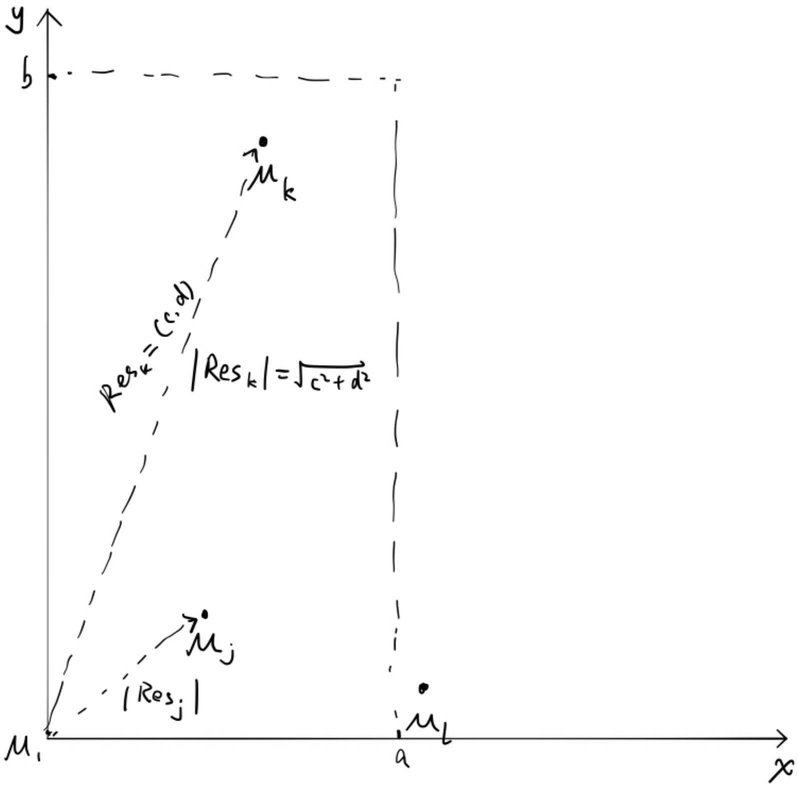


Figure 8: The plausibility of reaching a model.

ideally rational agents. In addition, it has discussed the philosophical implications it can provide for some philosophy of science problems that involve deductive reasoning and probability. Specifically, it reassesses the Dutch Book argument from the perspective of bounded rationality using the probabilistic framework provided in this paper.

There are several questions that can be discussed in future work. Firstly, one can discuss whether it is too demanding for the agents to switch from the qualitative ranking of plausibility of worlds to a cardinal probability assignment. One could borrow ideas from the ordinal and the cardinal utility theory to assess this question. Secondly, one could compare our approach with other approaches of probabilistic update: see Baltag and Smets (2008), Van Benthem (2003), Kooi (2003) for using conditional probabilities to update the degrees of belief. One could also explore other ways of updating degrees of belief such as Jeffrey's rule of conditioning (Jeffrey, 1990) and Dempster's rule of combination (Dempster, 2008). This could be combined with empirical research that tests each of these theories against data collected from experiments and real life. Other than comparing and proposing different approaches of update, one could also enrich the logic. For example, one could follow Fagin and Halpern (1994); Halpern and Tuttle (1993) and add linear inequalities to the language. One could also study different sources of probability or probabilistic aspects of incoming information: see Van Benthem et al. (2009). More interesting further research questions can be found in these papers mentioned.

References

Artemov, S. and Kuznets, R. (2006). Logical omniscience via proof complexity. In *International Workshop on Computer Science Logic*, pages 135–149. Springer.

Artemov, S. and Kuznets, R. (2009). Logical omniscience as a computational complexity problem. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 14–23.

Artemov, S. and Kuznets, R. (2014). Logical omniscience as infeasibility. *Annals of pure and applied logic*, 165(1):6–25.

Baltag, A. and Renne, B. (2016). Dynamic Epistemic Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2016 edition.

Baltag, A. and Smets, S. (2008). Probabilistic dynamic belief revision. *Synthese*, 165(2):179–202.

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

Cresswell, M. J. (2016). *Logics and languages*. Routledge.

Dean, W. (2021). Computational Complexity Theory. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2021 edition.

Dempster, A. P. (2008). Upper and lower probabilities induced by a multivalued mapping. In *Classic works of the Dempster-Shafer theory of belief functions*, pages 57–72. Springer.

Duc, H. N. (2001). *Resource Bounded Reasoning about Knowledge*. PhD thesis, Leipzig, Univ., Diss., 2001.

Fagin, R. and Halpern, J. Y. (1987). Belief, awareness, and limited reasoning. *Artificial intelligence*, 34(1):39–76.

Fagin, R. and Halpern, J. Y. (1994). Reasoning about knowledge and probability. *Journal of the ACM (JACM)*, 41(2):340–367.

Halpern, J. Y. and Tuttle, M. R. (1993). Knowledge, probability, and adversaries. *Journal of the ACM (JACM)*, 40(4):917–960.

Hintikka, J. (1979). Impossible possible worlds vindicated. In *Game-theoretical semantics*, pages 367–379. Springer.

Jeffrey, R. C. (1990). *The logic of decision*. University of Chicago press.

Kooi, B. P. (2003). Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information*, 12(4):381–408.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

Moreno, A. (1997). *How to avoid knowing it all*. Citeseer.

Rantala, V. (1982). Impossible worlds semantics and logical omniscience. *Acta Philosophica Fennica*, 35:106–115.

Rendsvig, R. and Symons, J. (2021). Epistemic Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition.

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

Solaki, A., Berto, F., and Smets, S. (2021). The logic of fast and slow thinking. *Erkenntnis*, 86(3):733–762.

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

Tversky, A. and Kahneman, D. (1985). The framing of decisions and the psychology of choice. In *Behavioral decision making*, pages 25–41. Springer.

Van Benthem, J. (2003). Conditional probability meets update logic. *Journal of Logic, Language and Information*, 12(4):409–421.

Van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of applied non-classical logics*, 17(2):129–155.

Van Benthem, J., Gerbrandy, J., and Kooi, B. (2009). Dynamic update with probabilities. *Studia Logica*, 93(1):67–96.