Steps out of Logical Omniscience

MSc Thesis *(Afstudeerscriptie)*

written by

Anthia Solaki
(born December 11, 1992 in Thessaloniki, Greece)

under the supervision of **Prof. dr. Francesco Berto** and **Prof. dr. Sonja Smets**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:
June 28, 2017

Members of the Thesis Committee:
Prof. dr. Francesco Berto
Dr. Luca Incurvati
Dr. Katrin Schulz (chair)
Prof. dr. Sonja Smets
Dr. Jakub Szymanik

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

This thesis discusses the problem of logical omniscience, a defect of standard epistemic and doxastic logics which – unrealistically – predict that agents know/believe all consequences of their knowledge/beliefs. We first give a detailed account of the problem, argue for its importance and describe the kind of solution we are interested in. More specifically, we attach great value to the ability of real-life agents to engage in *bounded* reasoning. Then, once we provide the appropriate background notions from Dynamic Epistemic Logic, we continue with a comprehensive review of selected approaches to the problem. In doing so, certain criteria are flagged, in order to assess these attempts on a solid ground. Keeping these remarks in mind, we proceed with our own proposals against the problem, in hope of overcoming the challenges emphasized in the critical survey. These proposals prioritize the need to take reasoning steps in order to attain knowledge or belief. First, we improve step-wise solutions to the problem by providing two frameworks, RW, that captures reasoning steps as transitions between worlds, and IW, that employs impossible worlds. We present the main elements of RW, explain how it refines existing attempts and escapes omniscience, and provide a sound and complete logic with respect to a class of its models. We similarly analyze the contribution of IW, and extend it to a quantitative system, sensitive to the idea of resource consumption. Other extended settings, such as IWPA and IWp, facilitate a more elaborate study of reasoning and belief change. Finally, we devise a method to obtain complete axiomatizations for IW-like systems, that relies on a reduction of models with impossible worlds.

Acknowledgements

First, I'd like to thank my supervisors. I want to thank Franz, for his continuous support, insightful feedback, and contagious enthusiasm on the topic; Sonja, for her valuable comments, the fruitful meetings and for pointing me to the right material for my purposes. I'm also grateful to Fernando, for the very interesting discussions, his suggestions and, of course, for all the proofreading. On several occasions his remarks saved me from trouble. These two years at the ILLC wouldn't have been so nice without the MoL-people. I enjoyed the time we spent together in front of the whiteboard as well as the breaks for poor-quality pizza, joking and talking about politics. It's impossible not to mention my favourite, yet geographically scattered, group of people: Anastasia, Dina, Marianna and Vicky. Be it late-night skype calls or endless chatting, they could always stand by me – *Thanks*. Last, but certainly not least, I want to thank my family; my parents, for constantly supporting me and trusting in my way of doing things; my brother, who had his own way of reminding me that if I'm to live thousands of kilometres away, it'd better be for something worthwhile; my sister, for her word of encouragement and advice, and for being *there* when I most needed it.

Outline of the thesis

For the convenience of the reader, we hereafter explain the structure of the thesis:

- Chapter 1: The standard epistemic and doxastic logics are introduced in Section 1.1. We explain how these emerged as spin-offs of Modal Logic and discuss some of their properties. In Section 1.2, we present the various forms and manifestations of the *problem of logical omniscience*. The problem plagues the mainstream treatment of knowledge and belief because this ascribes unlimited inferential power to the agents. We then emphasize why it is important to resolve it, arguing against claims that have been put forward to justify this sort of idealization. This discussion also hints at elements seen as desirable for a proposed solution.

- Chapter 2: This chapter serves as the background for what follows. It introduces elements from the toolkit of *Dynamic Epistemic Logic*, that can help us draw a more realistic picture of reasoning. First, we present *public announcement logic* (Section 2.1). Second, we describe the contribution of plausibility models in the study of (static) belief change (Section 2.2). Third, we briefly discuss dynamic belief change triggered by various kinds of incoming information (Section 2.3).

- Chapter 3: In this chapter, we provide a detailed exposition and discussion on selected proposals against the problem, as found in the literature. They are classified according to the rationale and method they adopt. Apart from explaining their workings, we also assess them according to specific criteria. This survey allows us to spot useful tools and underlying ideas, but also reveals the open challenges that await.

- Chapter 4: This chapter constitutes our own attempt to resolve the problem, in a way that improves existing approaches and accounts for the real, dynamic nature of reasoning. More specifically, we design two settings, dubbed *Rule-based worlds* (RW) and *Impossible worlds* (IW), that break reasoning processes into *reasoning steps*. In Section 4.1.1, we present the main elements of RW, we compare it to a similar step-wise view and construct the complete logic Λ_{RW}. In Section 4.1.2, we similarly present IW. This approach additionally allows for a more detailed analysis of reasoning, further pursued in Section 4.2 and Section 4.3, mainly inspired by Chapter 2. Next, Section 4.4 provides a reduction of frameworks with *impossible worlds*, that, combined with material from Chapter 3, facilitates the construction of complete axiomatic systems.

- Chapter 5: Finally, we summarize our main points and suggest directions for further investigation on the topic.

Contents

Chapter 1

The problem of logical omniscience

1.1 Epistemic and doxastic logic

Since Hintikka's seminal work in Hintikka (1962), Logic has been instrumental in the formal study of propositional attitudes such as knowledge and belief. Standard epistemic and doxastic logics were developed as spin-offs of *Modal Logic* and made use of its main techniques, in particular of the *possible worlds semantics*. The core of this conception is that in knowing or believing something, one obtains a way of determining which the *actual* world is among a range of possibilities. *Possible worlds* articulate precisely this conception: they embody these logical possibilities. Although there is a lively debate on their metaphysical status[1], given our purposes, we content ourselves in considering possible worlds alternative scenarios, representations of the ways the world could be or could have been.

The standard approach accounts for knowledge by supplementing the language of propositional logic with a unary operator K such that $K\phi$ reads: "the agent knows that ϕ". Following the same fashion, we can add a unary operator B such that $B\phi$ reads "the agent believes that ϕ". Next, the semantic interpretations are given in terms of possible worlds: an agent knows(/believes) that ϕ if and only if in all possible worlds compatible with what the agent knows(/believes), it is the case that ϕ.

Of course, there can be more than one operator to accommodate settings with more than one agent. Then, by indexing the operators, we get $K_i\phi$, read as "agent i knows that ϕ", and likewise for belief. The content of this chapter can be accordingly generalized for multi-agent settings.

Departing from these initial remarks, we will give the concrete account of standard single-agent epistemic logic, starting off with the constructions of Modal Logic (Blackburn et al. (2001), Bezhanishvili and van der Hoek (2014), Fagin et al. (1995a)). We will also comment on how this can be adapted for doxastic and combined epistemic-doxastic frameworks.

Definition 1.1.1 (Syntax)**.** The language of single-agent epistemic logic is defined inductively as follows:

[1] One may consult Berto and Plebani (2015) for a review of the several schools of thought.

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi \wedge \phi \quad | \quad K\phi$$

with $p \in \Phi$ and Φ a set of propositional atoms.

The language of single-agent epistemic-doxastic logic is easily obtained by supplementing the previous definition with $B\phi$. The common boolean connectives are defined in terms of $\neg$ and $\wedge$ as usual. It is also useful to consider the *dual* operators $\hat{K}, \hat{B}$ where $\hat{K}\phi := \neg K \neg\phi$ and $\hat{B}\phi := \neg B \neg\phi$.

Next, we elucidate the technical details that show how the relational structures of Modal Logic are utilized in our context. More specifically, the compatibility of worlds with the agent's knowledge and belief is captured via primitive binary relations on possible worlds, reflecting epistemic and doxastic accessibility. We first present the standard modal account, followed by the discussion on the properties that furnish this fruitful adaptation.

Definition 1.1.2 (Kripke frames and models)**.**

1. A *Kripke frame* is a pair $\mathcal{F} = \langle W, R \rangle$, where:
   - W is a non-empty set of possible worlds.
   - R is a binary *accessibility* relation on W.
2. A *Kripke model* is a frame supplemented with a valuation $V : \Phi \to \mathcal{P}(W)$ assigning to each $p \in \Phi$ a subset $V(p)$ of W. Intuitively, $V(p)$ is the set of all worlds in the model where p is true. A pair (M, w) consisting of a model M and a designated world w of the model is called a *pointed model*.

As we will see, the accessibility relation can be used to denote epistemic or doxastic accessibility. Of course, frames and models might be endowed with more than one accessibility relation, thereby allowing for combined epistemic-doxastic settings.

We now proceed with the truth clauses and other key-definitions:

Definition 1.1.3 (Truth)**.** For a world w in a model $M = \langle W, R, V \rangle$, we inductively define that a formula ϕ is *true in M at world w* (notation: $M, w \vDash \phi$) as follows:

- $M, w \vDash p$ if and only if $w \in V(p)$, where $p \in \Phi$.
- $M, w \vDash \neg\phi$ if and only if $M, w \nvDash \phi$.
- $M, w \vDash \phi \wedge \psi$ if and only if $M, w \vDash \phi$ and $M, w \vDash \psi$
- $M, w \vDash K\phi$ if and only if for all worlds $u \in W$ such that wRu we have $M, u \vDash \phi$.

A set Σ of formulas is true at a world w of a model M (notation: $M, w \vDash \Sigma$) if all members of Σ are true at w.

Regarding belief, and denoting the doxastic relation with R_b, we simply define an extended frame $\mathcal{F} = \langle W, R, R_b \rangle$ and model $M = \langle W, R, R_b, V \rangle$ as suggested above. Then the following clause can be added:

- $M, w \vDash B\phi$ if and only if for all worlds $u \in W$ such that wR_bu we have $M, u \vDash \phi$.

Definition 1.1.4 (Truth in a model)**.** A formula is *(globally) true* (or *valid*) *in a model* if it is true at all possible worlds of the model. A set of formulas is true in a model if all of its members are true in the model.

Definition 1.1.5 (Validity)**.** A formula ϕ is *valid at a world* w *in a frame* $\mathcal{F}$ (notation: $\mathcal{F}, w \models \phi$) if it is true at w in every model $\langle \mathcal{F}, V \rangle$ based on $\mathcal{F}$. It is *valid in a frame* $\mathcal{F}$ (notation: $\mathcal{F} \models \phi$) if it is valid at every world w in $\mathcal{F}$. It is *valid on a class of frames* if it is valid in every frame of the class. It is *valid* if it is valid on the class of all frames. The set of all formulas that are valid on a class of frames F is called the *logic* of F.

Definition 1.1.6 (Logical Implication and Equivalence)**.**

1. A set of formulas Ψ *logically implies* ϕ with respect to a class of frames F, if for all $\mathcal{F} \in F$ and all worlds $w \in \mathcal{F}$: whenever $\mathcal{F}, w \models \psi$ for every $\psi \in \Psi$, then $\mathcal{F}, w \models \phi$. We will also say that ϕ is a *logical consequence* of Ψ.

2. Two formulas are logically equivalent if each logically implies the other. One is true precisely when the other is true.

Note that the definition can be restated with respect to a class of models. From the foregoing it follows that a formula is valid if it is a logical consequence of the empty set of formulas.

The definitions above constitute the basis to illustrate the direct contribution of Modal Logic to the construction of epistemic and doxastic frameworks. Apart from these basic elements though, the contribution extends further: the use of *characterization results* renders many properties of knowledge and belief amenable to formal study. In particular, the validity of certain formulas is associated with certain properties of the accessibility relation(s). The following definition sets the background for the investigation of these effects.

Definition 1.1.7 (Normal modal epistemic logic)**.** A *normal modal epistemic logic* Λ is a set of formulas that contains all instances of propositional tautologies, all instances of the *Kripke schema* (K): $K(\phi \to \psi) \to (K\phi \to K\psi)$ and is closed under *Modus Ponens* and the *Necessitation Rule* (N): from ϕ infer $K\phi$.

By suitable modifications of the operators, the (normal) doxastic counterpart is easily obtained.

As a result, certain logics, built on the addition of special schemes of formulas as axioms, induce certain algebraic properties on the accessibility relations. The classes of frames that are determined by those properties reflect useful properties of knowledge and belief, often revealing connections to epistemological corollaries.

To begin with, the class of all frames corresponds to the smallest normal modal logic, which is called **K**. Extensions of this logic are obtained via adding axiom schemes that seem plausible according to our intuitive understanding of knowledge/belief and the epistemological discussion that has long investigated how these attitudes can be discerned. We hereafter give an overview of properties that have been suggested for the adequate formal description of knowledge and belief as well as of the results of their inclusion at the logical level. In this sense, the standard modal constructions are transformed into epistemic and doxastic frameworks.

Veridicality

The axiom scheme that reflects the *veridicality of knowledge*, i.e. that if ϕ is known then ϕ is true, is called (T): $K\phi \to \phi$. Its addition results in the logic **T**. One can easily check that (T) corresponds to the class of those frames where for every world w, wRw, i.e. the class of all *reflexive* frames. Likewise, if one accepts veridicality of belief, the belief-version of the scheme should be added, turning R_b into a reflexive relation too. It is worth noticing that most formalizations do not assume veridicality for belief.

Consistency

The axiom scheme that reflects the *consistency of knowledge* is called (D): $K\phi \to \neg K\neg\phi$. It is equivalent to $\neg K(\phi \wedge \neg\phi)$. Its addition results in the logic **D**. The axiom is valid precisely on those frames where for any world w, there is some world u such that wRu, i.e. the class of all *serial* frames[2]. Accordingly, the belief-version of the axiom is $B\phi \to \neg B\neg\phi$ and corresponds to seriality of the doxastic accessibility relation.

Positive Introspection

The instances of the axiom (4) $K\phi \to KK\phi$ reflect the *positive introspection of knowledge*. The addition of this axiom scheme yields the logic **K4**. It characterizes the class of those frames where for any worlds w, u, v, if wRu and uRv then wRv, i.e. the class of all *transitive* frames. Positive introspection of belief works along the same lines.

Negative Introspection

The instances of the axiom (5) $\neg K\phi \to K\neg K\phi$ reflect the *negative introspection of knowledge* and result in the logic **K5**. This axiom scheme characterizes the class of those frames where for any worlds w, u, v, if wRu and wRv then uRv, i.e. the class of all *euclidean* frames. Negative introspection of belief again works along these lines.

While veridicality is often seen as an essential property for knowledge, this is not the case for belief; it is generally accepted that an agent might hold false beliefs. It has been argued that this is one of the properties that can be used to distinguish knowledge and belief. As a result, a belief-version of the axiom is not usually assumed and, in turn, doxastic accessibility need not be reflexive. On the other hand, the intuitive appeal of positive and negative introspection is considered debatable for both knowledge and belief, as is consistency of belief, and no absolute consensus has been reached regarding the inclusion of the respective axioms (Danto (1967), Hintikka (1962), Lemmon (1967), Stalnaker (2006)).

The aforementioned remarks are summarized in Table 1.1. Overall, combinations of these axioms result in logical systems of varying strength that are sound and complete with respect to those classes of frames complying with the analogous combinations of restrictions on the accessibility relation(s). Picking the *most* appropriate system depends on one's dispositions and goals.

We only notice that according to the received view (e.g. as in Fagin et al. (1995a)) (a) epistemic models are S5-models, that is models in which the (epistemic) accessibility relation is

[2] It is then easy to see that (D) can follow from other axioms, such as (T).

an equivalence relation (reflexive, transitive and symmetric or equivalently: reflexive and euclidean) and (b) doxastic models are KD45-models, that is models in which the (doxastic) accessibility relation is serial, transitive and euclidean. These modelling choices give rise to certain additional properties regarding the interaction between knowledge and belief operators under combined epistemic-doxastic frameworks. In other words, if we assign the proposed qualities to the epistemic and doxastic relations, the following validities easily follow:

- **Strong positive introspection of beliefs** $B\phi \rightarrow KB\phi$.
- **Strong negative introspection of beliefs** $\neg B\phi \rightarrow K\neg B\phi$.
- **Knowledge implies belief** $K\phi \rightarrow B\phi$.

Table 1.1: Common logics

Logic	Axioms	Class of frames
K	(K)	All
T	(K), (T)	Reflexive
D	(K), (D)	Serial
K4	(K), (4)	Transitive
K5	(K), (5)	Euclidean
KD45	(K), (D), (4), (5)	Serial, Transitive, Euclidean
S4	(K), (T), (4)	Reflexive, Transitive
S5	(K), (T), (5)	Reflexive, Transitive, Symmetric

This is the mainstream logical landscape drawn by the hintikkian approach, heavily influenced by the machinery of normal modal logics. As we will see, the seemingly smooth integration of knowledge and belief in this picture faces serious objections.

1.2 The problem

We have seen the main properties of epistemic and doxastic logics. Despite the benefits reaped by exploiting Modal Logic in the formal study of knowledge and belief, there is a certain cost. The *problem of logical omniscience* (identified in Halpern and Pucella (2011), Hintikka (1975), Moses (1988), Parikh (2008), among others), is an inherent defect of this treatment. It manifests itself as follows:

> Suppose that an agent at a world w knows all formulas in a set Ψ and that Ψ logically implies ϕ. Because of the former assumption, all formulas of Ψ hold at every world epistemically accessible from w. Due to the latter assumption, ϕ holds at these worlds as well. Therefore, the agent knows ϕ at w.

This closure property constitutes the *full* problem of logical omniscience. Notice that the problem can be easily restated for belief. Given the aim of providing a theory for actual reasoners, it is not difficult to spot the malignancy of the problem. The predictions of this approach are not accurate; the brightest mathematician might know all axioms of set theory without thereby knowing all their consequences. Or, although a winning strategy for a game might follow mathematically from a given state of the game and players are aware of the latter, it is not the case that they always play according to the winning strategy; otherwise, many games would be pointless and uninteresting. The performance of real-life agents is inhibited by their limited memory, computational capacity, biases, faulty reasoning etc. That is to say that real-life agents are fallible and resource-bounded, therefore not well accommodated within these settings.

Equally alarming considerations arise from special cases of the full form. In addition, even if certain modifications alter the kind of structures and the notion of truth in a manner that avoids the full problem, the divergence from reality is retained through weaker problematic closure principles. More specifically, all these special and weaker forms are given below, following the work of van Ditmarsch et al. (2007) and Fagin et al. (1995a).

1. If ϕ is valid, then the agent knows ϕ. (*Knowledge of valid formulas*)
2. If the agent knows ϕ and ϕ logically implies ψ, then the agent knows ψ. (*Closure under Logical Implication*)
3. If the agent knows ϕ and ϕ is logically equivalent to ψ, then the agent knows ψ. (*Closure under Logical Equivalence*)
4. If the agent knows ϕ and also knows $\phi \to \psi$, then the agent knows ψ. (*Closure under Material Implication*)
5. If the agent knows ϕ and $\phi \to \psi$ is valid, then the agent knows ψ. (*Closure under Valid Implication*)
6. If the agent knows ϕ and also knows ψ, then the agent knows $\phi \wedge \psi$. (*Closure under Conjunction*)
7. If the agent knows ϕ, then the agent knows $\phi \vee \psi$. (*Closure under Disjunction*)

Again, the foregoing can be restated for belief.

Knowledge of all valid formulas is a special case of the full form as validity boils down to logical consequence from the empty set. The discrepancy between the standard treatment and real agents is once again apparent. For example, it is not realistic to expect that agents believe or know every propositional tautology irrespective of its complexity. In the same line of reasoning, consider Goldbach's Conjecture[3]; if it is true, then it is true at all possible worlds and if it is false, then it is false at all possible worlds. As a result, in any case, the correct response to the Conjecture is known by any agent, according to the possible worlds account. In contrast to these predictions, though, Goldbach's Conjecture remains an unsolved mathematical problem. Yet another illustration of the problem arises from the Closure under Logical Implication, that follows immediately from the full form, as well as from the Closure under Logical Equivalence, predicting that an agent knows any formula that is equivalent to a formula she knows. Closure under Material Implication is a weaker principle not necessarily following from the full form. However, it coincides with Closure under Logical Implication in standard modal logic. Closure under Valid Implication is equivalent to Closure under Logical Implication because in

[3] Every even integer greater than 2 can be expressed as the sum of two primes.

standard modal logic $\phi \to \psi$ is valid precisely when ϕ logically implies ψ. Closure under Conjunction and Disjunction are special cases of full logical omniscience, if the set $\{\phi, \psi\}$ logically implies $\phi \wedge \psi$ and if ϕ logically implies $\phi \vee \psi$ – respectively.

Another counterintuitive quality that is nevertheless attributed to agents, at least under systems that contain the axiom (D), is that of Consistency: agents never know/believe both ϕ and $\neg\phi$. However, in the real-world, cognitively limited reasoners often maintain inconsistent beliefs, whether they realize it or not. More specifically, it has been claimed that agents might even believe a contradiction *explicitly* and consider themselves justified in doing so, as *dialetheists* believe that a particular sentence, the *liar sentence*, is simultaneously true and false (Priest (2006)).

At this point, it is worth noticing that this kind of idealization, as indicated by the aforementioned list, is also observed in mainstream attempts of logical modelling on belief change. As also observed in the next chapter, the building block of the predominant *AGM* approach (Alchourrón et al. (1985)), the *belief sets*, also suffer from closure principles that lead to omniscient agents. According to this approach, the beliefs of an agent are represented by a set of sentences in a formal language. This set is taken to be closed under logical consequence, i.e. if p is in a belief set and q logically follows from p, then q is already in the set. But of course, this too entails that agents are expected to believe all consequences of their beliefs, thus leading to the undesired properties of the full form of logical omniscience. In addition, if the belief set contains both p and $\neg p$, i.e. the agent holds some inconsistent belief, then her belief state necessarily collapses to the trivial one, as she is expected to believe *everything*. Moreover, following the AGM postulates: if two sentences p and q are logically equivalent, then believing the one amounts to believing the other. However, we often revise our beliefs influenced by the mode of presentation and the frame under which the revision takes place. In these cases, we might end up believing the one without believing the other.

Despite the discrepancy between logical predictions and reality, there have been attempts to defend the standard paradigm and view its properties as inevitable or even desired tools. For instance, Stalnaker (1991) and Yap (2014) examine reasons that could justify the extent of the idealization. First, this is sometimes defended as the means to reach the mechanisms underlying the complex theory of knowledge and belief. Motivated by certain examples from other disciplines, e.g. the use of frictionless planes in physics, it has been argued that the isolated study of individual components of larger theories increases our understanding of them, even if we miss out on their interconnections. For example, external forces may be ignored and the realistic picture may be only partially drawn because the internal dynamics tend to move the system in question towards an equilibrium. In this line of reasoning, idealization can be justified by viewing the fallibility of agents as a kind of "cognitive friction" that interferes with the reasoning process yet the latter eventually reaches an equilibrium where perfect rationality is attained. A second reason backing idealization lies in the need for simplification: the cost of distortion is assessed as unimportant when compared to the benefits of simplifying. Thirdly, another source of justification is presented by virtue of normativity: although the standard logics draw an ideal picture, far from the actual inner workings of knowledge and belief, they are still considered acceptable and valuable as they set the standard that rational agents *ought to* comply with.

However, these arguments cannot completely alleviate the worries on logical omniscience. To begin with, the distortion induced by the closure principles is not negligible. Resorting to the idealized models of other disciplines seems more like a convenient analogy[4]. In particular,

[4] For all fairness, there is a large discussion on idealization and abstraction as tools for scientific investigation, in

there is no reason, theoretical or empirical, to assume that the reasoning process, constantly influenced by external information, ever reaches an equilibrium of spotless rationality. As a result, one cannot do away with logical omniscience by merely suggesting that it poses no threat in the long run. Moreover, considering the descriptive use of epistemic and doxastic logics, the argument for simplicity is ineffective because the extent of the chasm between idealized and real agents is substantial enough to obscure many of the benefits. Apart from the fact that the closure principles are not aligned with ordinary intuitions, there is also concrete evidence that sheds light on actual cognitive states and highlights the extent of the defect. Cognitive science and psychology of reasoning import experimental evidence suggesting that subjects' performance in reasoning tasks, e.g. in the *Wason selection task* or the *suppression task*, is not always consonant with logical predictions (Stenning and van Lambalgen (2008)). Furthermore, Parikh (2008), prompted by Daniel Kahneman's work on behavioural economics, argues that human belief states are neither consistent nor usually closed under logical inference. In general, the shift from classical to behavioural economics (Kahneman (2003), Simon (1955)) endorses the revision of idealized models of perfect rationality so that limited resources and the framing of decision-making are taken into account. In addition, the experiments discussed in Alxatib and Pelletier (2011) and Ripley (2011) show that in certain cases, agents hold – at least *prima facie* – inconsistent beliefs. This does not mean that they are "absurd" nor willing to believe *everything*, as the standard account predicts. Next, appealing to normativity to secure standard epistemic and doxastic logics from objections also faces counterarguments: there seem to be good reasons, for example, to account for the fact that agents do not know *all* consequences of their knowledge even while aiming for a normative model of how we *ought to* reason. That is, acknowledging our own fallibility is often seen as a prerequisite to rationality. Stalnaker (1991) specifically reports on the view that rational agents should believe that some of their own beliefs are false. Forcing one to commit to models that are *either* non-normative *or* representing omniscient agents might as well be a false dilemma. A normative model can still focus on a *moderately* rational agent, who is able to conduct finite chains of inferences avoiding blatant inconsistencies, despite being non-omniscient. Finally, Hintikka's own understanding of the problem did not presuppose any kind of defense of his standard systems due to normativity:

> Logical truths are not truths which logic forces on us; they are not necessary truths in the sense of being unavoidable. They are not truths we must know, but truths which we can know without making use of any factual information. [...] The fact that the so-called laws of logic are not "laws of thought" in the sense of natural laws seems to be generally admitted nowadays. Yet the laws of logic are not laws of thought in the sense of commands, either, except perhaps laws of the sharpest possible thought. Given a number of premises, logic does not tell us what conclusions we ought to draw from them; it merely tells us what conclusions we may draw from them – if we wish and we are clever enough.[5]

Consequently, there is not enough support to defend the modelling of agents with infinite inferential powers as a means to say how they ought to perform.

We have thus far presented the problem of logical omniscience and emphasized its importance. However, the intuitive considerations and the experimental evidence that dictate the attack against logical omniscience also urge us to demarcate another feature of real agents'

general (see for example Stokhof and van Lambalgen (2011), for similar considerations). The wider study, that touches upon philosophy of science, is beyond our scope.

[5] Hintikka (1962), p.37.

epistemic/doxastic states. Although real agents are fallible and non-omniscient, they still are *logically competent*; their rationality might be bounded but it is not absent. In particular, we often fail in making complex inferences as we lack the necessary time, memory or computational power. Even if these are sufficient, incomplete reasoning or biases interfere with our judgment. Yet, we do engage in bounded reasoning: noticing that it is (once again) raining in Amsterdam, we would normally take our raincoats before leaving home. This is because from our beliefs that (a) it is raining and that (b) whenever it is raining, we need a raincoat, we infer that we should wear the coat and act accordingly. Furthermore, people seemingly holding inconsistent beliefs, are still considered (moderately) rational. We might hold false beliefs without this preventing us from reasoning and operating in the world without *much* trouble. The interdisciplinary empirical data mentioned earlier also contributes to the case for logical competence. For instance, subjects' performance in the Wason selection task was remarkably improved when it was stated as imitating a familiar social norm (Griggs and Cox (1982)). In van Benthem et al. (2016), we also encounter a defense of logical competence on the grounds of these task-dependent fluctuations of performance.

> Clearly, people are not irrational, and if they ignored logic all the time, extracting the wrong information from the data at their disposal, it is hard to see how our species could survive. What seems to be the case is rather an issue of representation of reasoning tasks, and additional principles that play a role there.[6]

In addition, the subjects of Alxatib and Pelletier (2011) were able to provide good reasons for claiming that a certain suspect is both tall and not tall. Their responses triggered the re-evaluation of classical logic and the extended study of phenomena of vagueness rather than the re-evaluation of the subjects' mental capacities. Along the same lines, the research on decision theory and economics stresses the importance of the availability of resources, the pursuit of a satisfactory but not always optimal solution, the influence of fallacies in decision-making etc., without suggesting that agents' activity collapses to irrationality. Therefore, a successful attempt to model actual epistemic/doxastic states presupposes that agents are logically competent and more specifically they do not miss out on trivial consequences of what they know or believe.

As a result, logical closure principles, either illustrated in the possible worlds semantics or the purely syntactic belief sets, give rise to the problem of logical omniscience. It is therefore essential to revise the current outlook on logical modelling of propositional attitudes, if we are to capture their realistic effect.

[6] van Benthem et al. (2016), p. 2-35.

Chapter 2

Dynamic Epistemic Logic

Chapter 1 discussed the paradigmatic accounts on epistemic and doxastic logic as well as the major problem of logical omniscience. These accounts, however, are purely static; they model knowledge and belief, as held at a particular moment. As a result, they cast aside the constant changes of attitudes triggered by both our "internal" mental processes (e.g. performing inferences) and our "external" interactions (e.g. the information exchange that takes place during a discussion). It is therefore clear that merely focusing on a glimpse of an agent's epistemic/doxastic state yields a rather limited modelling, that omits real-life actions interfering with our reasoning. Such deficiencies can be treated by using tools from *Dynamic Epistemic Logic* (DEL), that puts *model change* under scrutiny. The vast variety of systems designed within this field allows for modelling of a plethora of attitudes and of multiple phenomena, especially concerning multi-agent settings. Given our purposes though, we only review a selection of these[1] – more specifically those that provide the background for the content of the next chapters – and restrict our attention to the changing states of a single agent.

The general idea is to enrich the standard language by modal operators that correspond to the actions capable of altering an agent's epistemic or doxastic state. Their effect is then captured via *model transformations*. If a formula is of the form $[\,]\phi$ with $[\,]$ such an operator, then it is evaluated at a particular world in a model by examining what the truth value of ϕ is at the transformed model. That is, formulas involving action operators are evaluated by utilizing transitions from the original model, activated by the action of the corresponding operator.

2.1 Public Announcement Logic

In what follows, we summarize *Public Announcement Logic* (PAL) (Plaza (2007)), because it offers a clear illustration of the above and provides the foundations to better understand more complex actions as well as the details for some of the proposals of Chapter 3 and Chapter 4. To begin with, its language is the extension of the standard language with modal operators $[\psi!]$ such that $[\psi!]\phi$ reads "after the public announcement of ψ, ϕ is true". The announcement is

[1] For detailed surveys, one can consult Baltag and Renne (2016), van Ditmarsch et al. (2007).

thought of as truthful and absolutely reliable; this is the motivation behind the definition of the transformed model $M^{\psi!}$ as a model in which all not-ψ worlds are eliminated. Formally:

Definition 2.1.1 (Model transformation by public announcement)**.** Given a Kripke model $M = \langle W, R, V \rangle$, its transformation by $!\psi$ is a model $M^{\psi!} = \langle W^{\psi!}, R^{\psi!}, V^{\psi!} \rangle$ where:

- $W^{\psi!} = \{w \in W \mid M, w \vDash \psi\}$
- $R^{\psi!} = R \cap (W^{\psi!} \times W^{\psi!})$
- $V^{\psi!}(p) = V(p) \cap W^{\psi!}$

The truth clauses are then supplemented with the extra clause: $M, w \vDash [\psi!]\phi$ if and only if $M, w \nvDash \psi$ or $M^{\psi!}, w \vDash \phi$. The first part of the clause is such to obey the restriction to *truthful* announcements: if the announced sentence is false, then $[\psi!]\phi$ is vacuously true.

It can be shown that the addition of the following axioms and rule to the axioms and rules of **S5** [2] results in a sound and complete axiomatic system. These axioms are often called *reduction axioms*, for they reduce the complexity of formulas with announcements. Indeed, we can gradually end up with formulas that do not involve announcements at all, i.e. formulas of our basic language. Subsequently, the completeness of PAL follows immediately from the completeness of **S5**.

- $[\psi!]p \leftrightarrow (\psi \rightarrow p)$
- $[\psi!]\neg\phi \leftrightarrow (\psi \rightarrow \neg[\psi!]\phi)$
- $[\psi!](\phi \wedge \chi) \leftrightarrow ([\psi!]\phi \wedge [\psi!]\chi)$
- $[\psi!]K\phi \leftrightarrow (\psi \rightarrow K([\psi!]\phi))$
- From ϕ infer $[\psi!]\phi$

The above can be easily adapted for frameworks involving belief.

A subtle point that is worth a remark is that announced sentences do not always preserve their truth value *after* the announcement. The most prominent case in point is *Moore formulas*, such as $p \wedge \neg Bp$: it is not hard to see why the very announcement of this defeats its truth. Moore formulas then indicate that PAL is not closed under substitution.

With this illustration in mind, we only emphasize that it is possible to generalize the intuition behind action-induced change and thus study more sophisticated real-life scenarios. This has been achieved due to the construction of *action models* and *product updates*, introduced in Baltag et al. (1998).

2.2 Belief change and plausibility models

Dynamic Epistemic Logic also incorporates ideas from Belief Revision. According to the *AGM theory* (Alchourrón et al. (1985)), an agent's beliefs are given by a logically closed set of sentences, her belief set. This belief set might be expanded, contracted or revised, in the face of

[2] The same holds if we substitute **S5** with other appropriate systems – appropriate, in the sense of being sound and complete with respect to a class of models closed under the announcements.

new information, represented by a sentence ϕ. The corresponding operations, *expansion*, *contraction*, and *revision* are ruled by the *AGM postulates*. However, their status is controversial, as they too suffer from concerns on their adequacy in capturing realistic belief change.

Until now we have elaborated on the effect of public announcements. It is natural to think of them as a kind of expansion, given that the elimination of worlds results in an enrichment of the agent's factual knowledge, an idea investigated in van Ditmarsch et al. (2004). Yet one may come up with examples of incoming information such that the basic view of expansion does not suffice; in particular, if the announced sentence contradicts existing beliefs, then the agent ends up believing *everything*. In order to deal with *changing* beliefs, DEL primarily relies on *plausibility models* (Baltag and Smets (2008)). These allow us to express various grades of knowledge and belief. Importantly, *conditional beliefs* express what is believed, depended on certain incoming pieces of information, and in this way, they manage to model *static belief change*, that is, belief change in a non-changing situation.

Definition 2.2.1 (Plausibility model)**.** A *plausibility model* M is a structure $\langle W, \geq, V\rangle$ where:

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

Abbreviations such as $>$, $\leq$, $<$ are defined as usual; for example, we will use $w < u$ to say that $u \geq w$ and $w \ngeq u$ with the slash denoting a negated relation. In order to make precise the notion of a *locally well-preordered relation*, first consider the binary relation $\sim$ on W such that for $w, u \in W$: $w \sim u$ if and only if $w(\geq \cup \leq)^* u$. Then *local connectedness*[3] amounts to: if $w \sim u$ then $w \geq u$ or $u \geq w$. *Converse well-foundedness* amounts to: for each non-empty set $P \subseteq W$, the set of its minimal elements (i.e. $min(P) := \{w \in P \mid \forall u \in P : u \nless w\}$) is non-empty. Bringing together reflexivity, transitivity, local connectedness and converse well-foundedness, we obtain the definition of a locally well-preordered relation. The intuitive appeal of reflexivity and transitivity is obvious, given the reading of $\geq$. Local connectedness is invoked to say that the agent should be able to assign a relative plausibility between any two worlds considered possible. Converse well-foundedness is imposed to avoid infinite chains of more and more plausible worlds; being able to retrieve the set of "the most plausible worlds" is instrumental for the definitions that follow.

In order to describe other attitudes, we supplement the standard epistemic language with a modal operator $\Box$ such that $\Box\phi$ stands for "ϕ is defeasibly known by the agent". *Defeasible knowledge* (or *safe belief*) is a weaker notion distinguished from the ordinary K reading, discussed in Lehrer and Paxson (1969), Lehrer (2000) and formalized in Stalnaker (2006). While K denotes an infallible and irrevocable kind of knowledge, that persists even in the face of false incoming information, defeasible knowledge only persists in the face of *true* incoming information[4]. The semantics, in terms of plausibility models, is given by:

[3] In simplified settings, we could simply impose *connectedness* on $\geq$, that would amount to $w \geq u$ or $u \geq w$, for every $w, u \in W$, and thus obtain a definition of $\sim$ in terms of these two cases.

[4] Lehrer's *justification game*, as in Lehrer (2000) and Fiutek (2013) is illustrative for the study of defeasible knowledge. Roughly, and to connect it with our description, suppose that an agent x, the Claimant, holds a justified true belief and an agent y, the Critic, who is truthful and omniscient, challenges x with several objections. For the Claimant's belief to count as defeasible knowledge, she should be able to overcome the Critic's objections and pass the justification game successfully.

Definition 2.2.2 (Semantics-plausibility models)**.**

- $M, w \vDash p$ if and only if $w \in V(p)$.
- $M, w \vDash \neg\phi$ if and only if $M, w \nvDash \phi$.
- $M, w \vDash \phi \wedge \psi$ if and only if $M, w \vDash \phi$ and $M, w \vDash \psi$.
- $M, w \vDash K\phi$ if and only if $M, u \vDash \phi$ for all $u \in W$ with $w \sim u$.
- $M, w \vDash \Box\phi$ if and only if $M, u \vDash \phi$ for all $u \in W$ with $u \leq w$.

Furthermore, belief can be also accommodated within this setting. As promised earlier, we can talk about the agent's conditional beliefs, denoted by $B^{\psi}\phi$ and interpreted as "the agent believes ϕ, conditional on ψ". Conditional belief can be given as an expression involving the dual of K ($\hat{K}\phi := \neg K \neg\phi$) and $\Box$ as follows: $\hat{K}\psi \rightarrow \hat{K}(\psi \wedge \Box(\psi \rightarrow \phi))$. Its corresponding truth clause can be obtained in a simple way by:

$$M, w \vDash B^{\psi}\phi \text{ if and only if } M, u \vDash \phi \text{ for all } u \in min\{u \in W \mid w \sim u \wedge u \in [[\psi]]\}$$

with $[[\psi]]$ denoting the set of worlds where ψ is true.

It is then easy to view plain belief as a special case of conditional belief, and more specifically as $B^{\top}\phi$; since $\top$ is always true, $B\phi$ amounts to *unconditional* belief of ϕ. Then naturally:

$$M, w \vDash B\phi \text{ if and only if } M, u \vDash \phi \text{ for all } u \in min\{u \in W \mid w \sim u\}$$

Given that conditional, and thus plain, belief is expressible in terms of K and $\Box$, it has been shown (Baltag and Smets (2008)) that a sound and complete axiomatization (with respect to the class of pointed plausibility models) for this variety of notions is obtained by:

- The $S5$ axiom schemes and rules for K.
- The $S4$ axiom schemes and rules for $\Box$.
- $K\phi \rightarrow \Box\phi$.
- $K(\Box\phi \rightarrow \psi) \vee K(\Box\psi \rightarrow \phi)$.

2.3 Dynamic belief change due to hard and soft information

Turning to the dynamics of plausibility models, the account of public announcements, as sources of hard information, can be adapted for plausibility models too.

Definition 2.3.1 (Plausibility model transformation by public announcement)**.** Given a plausibility model $M = \langle W, \geq, V\rangle$, its transformation by $\psi!$ is a model $M^{\psi!} = \langle W^{\psi!}, \geq^{\psi!}, V^{\psi!}\rangle$ where:

- $W^{\psi!} = \{w \in W \mid M, w \vDash \psi\}$
- $\geq^{\psi!} = \geq \cap (W^{\psi!} \times W^{\psi!})$
- $V^{\psi!}(p) = V(p) \cap W^{\psi!}$

The truth clause for sentences of the form $[\psi!]\phi$ is then given in the same spirit as above. A sound and complete axiomatization can be obtained by supplementing the axiom schemes and rules of any static logic corresponding to the model class we are interested in, and the PAL reduction axioms mentioned above, with a reduction axiom for defeasible knowledge[5].

$$[\psi!]\Box\phi \leftrightarrow (\psi \to \Box(\psi \to [\psi!]\phi))$$

However, real-life interaction does not only involve truthful and absolutely reliable information. For example, cases in which the source is partially trusted are suggestive of actions bringing along "softer" information, that only changes our beliefs but not our knowledge. This is why van Benthem (2007) suggested other policies of belief change. Plausibility models, which offer a more detailed outlook to the states of the agent, enable us to study the effect of such actions. The main idea is that soft information cannot really eliminate a world. Rather, it changes the plausibility ordering so that the incoming information is somehow "prioritized", without altogether discarding the other possibilities. For our purposes, we will focus on the revision operation of *radically upgrading* with ψ ($\psi\Uparrow$), that rearranges worlds in a way that renders all ψ-worlds more plausible than all $\neg\psi$-worlds, and leaves intact the ordering within these two zones[6]. Given that our language is suitably extended with operators $[\psi\Uparrow]$, the following definition leads to the truth clause for $[\psi\Uparrow]\phi$:

Definition 2.3.2 (Model transformation by radical upgrade)**.** Given a plausibility model $M = \langle W, \geq, V\rangle$, its transformation by $\psi\Uparrow$ is a model $M^{\psi\Uparrow} = \langle W^{\psi\Uparrow}, \geq^{\psi\Uparrow}, V^{\psi\Uparrow}\rangle$ where:

- $W^{\psi\Uparrow} = W$
- $\geq^{\psi\Uparrow} = (\geq \cap (W \times [[\psi]])) \cup (\geq \cap ([[\neg\psi]] \times W)) \cup (\sim \cap ([[\neg\psi]] \times [[\psi]])$
- $V^{\psi!}(p) = V(p)$

Then, $M, w \vDash [\psi\Uparrow]\phi$ if and only if $M^{[\psi\Uparrow]}, w \vDash \phi$.

We can obtain a complete axiomatization (van Benthem (2007), van Ditmarsch et al. (2015)) for the dynamic logic of radical upgrade by augmenting any complete axiomatization on the static models by the following reduction axioms and rule:

- $[\psi\Uparrow]p \leftrightarrow p$
- $[\psi\Uparrow]\neg\phi \leftrightarrow \neg[\psi\Uparrow]\phi$
- $[\psi\Uparrow](\phi \wedge \chi) \leftrightarrow [\psi\Uparrow]\phi \wedge [\psi\Uparrow]\chi$
- $[\psi\Uparrow]K\phi \leftrightarrow K[\psi\Uparrow]\phi$
- $[\psi\Uparrow]B^{\chi}\phi \leftrightarrow (\hat{K}(\psi \wedge [\psi\Uparrow]\chi) \wedge B^{\psi\wedge[\psi\Uparrow]\chi}[\psi\Uparrow]\phi) \vee (\neg\hat{K}(\psi \wedge [\psi\Uparrow]\chi) \wedge B^{[\psi\Uparrow]\chi}[\psi\Uparrow]\phi)$
- From ϕ, infer $[\psi\Uparrow]\phi$

[5] Reduction axioms for conditional beliefs can be analogously obtained. We confined ourselves to $\Box$, given the way conditional beliefs were defined in terms of K and $\Box$. Consult van Ditmarsch et al. (2015), Chapter 7, for logics built on conditional beliefs.

[6] Radical (or lexicographic) upgrade is widely discussed in van Benthem (2007), and it is also representable with the machinery found in Baltag and Smets (2008).

This synopsis of elements from the influential DEL literature smooths the path towards the discussion on the treatment of the problem of logical omniscience. This is not to come as a surprise. The problem itself is indicative of the gap between the standard, static systems and reality. Dynamic epistemic logic brings us closer to real-life scenarios; the foregoing hint at some of the meaningful ways it has done so. It is therefore clear that once dynamics join forces, the idealized, breeding ground for omniscience, gets a substantial strike. This is exactly why material from this chapter contributes to proposals in the literature (surveyed in Chapter 3) as well as to our own suggestions (made in Chapter 4).

Chapter 3

Dealing with the problem: a critical survey

In this chapter, we examine prominent attempts to cope with the problem of logical omniscience. The examination is structured according to the following classification:

- Syntactic approaches.
- Approaches that propose a distinction between *implicit* and *explicit* attitudes, invalidating the problematic closure principles with respect to the latter. These comprise *awareness structures*, *algorithmic structures*, *justification logics* and *logics of justified knowledge and belief*.
- Impossible-worlds frameworks, that extend the usual set of worlds with impossibilities. Such approaches are divided into: elementary ones, involving worlds that are either not closed by any notion of logical consequence or (only) closed under some non-classical notion of logical consequence; Jago's approach as described in Jago (2014), imposing a suitable structure on the epistemic space; the attempt of Rasmussen and Bjerring (2015) who aim at a dynamic framework that traces the evolution of an agent's reasoning process.

Apart from explaining each proposal's contribution[1] towards the solution of the problem, we also comment on their adequacy according to both general criteria and proposal-specific objections. The major general criterion of our evaluation is testing whether the avoidance of logical omniscience is accompanied by an overall attractive modelling of agents' bounded, but not absent, rationality. That is, merely escaping the forms of the problem does not suffice; an attractive approach should also reflect that agents are (moderately) logically competent. In fact, what we want to avoid is non-omniscience collapsing into total irrationality and ignorance. We are hesitant to accept that one might fail in knowing even the most trivial consequences of what she knows. Of course, strictly determining what can count as "trivial" consequence is not an easy task, especially given that *any* inference can be unfolded as a chain

[1] In doing so, we will remain faithful to the descriptions as found in the literature. However, slight modifications should be tolerated, in the interest of readability and notational consistency.

consisting of “easy”steps. Despite the vague nature of the notion of moderate logical competence, we still value its integration into a proposed framework, at least from a normative point of view: it represents how agents *ought to* perform. Another criterion emerges if we further assess the explanatory power in capturing these subtle differences. In other words, we want to see whether the approaches are intuitively plausible and aligned with our understanding of what actually goes on whenever real agents reason. For example, a technically sufficient solution that nonetheless relies on ad-hoc and not independently motivated assumptions and modifications should not be considered entirely successful. Besides, resolving the problem per se would not have required extreme effort if we had just tweaked the semantics of the standard systems in accordance with the very goal of destroying the unwelcome closure principles. However, in that case, in the absence of any (other) concrete incentive to motivate the modification, it is doubtful whether the result would pertain to the propositional attitudes we examine, at least in a meaningful manner. Furthermore – and unsurprisingly – finding a way out of the problem often requires the introduction of additional machinery. The danger then lurks in obtaining new or weaker forms of logical omniscience with respect to these newly introduced elements. It is therefore worth checking whether logical omniscience is avoided without simultaneously generating further problems.

3.1 Syntactic approaches

3.1.1 Syntactic structures

The main idea behind this syntactic approach, described in Eberle (1974), Fagin et al. (1995a), Halpern and Pucella (2011), is to identify the agent’s epistemic state with the set of formulas that she knows, at each possible world[2]. Indeed, we *explicitly* list these formulas at a primitive level, without relying on the usual recursive definition and thus on the epistemic accessibility relation. More concretely:

Definition 3.1.1 (Syntactic structure)**.** A *syntactic structure* $\langle W, C\rangle$ is a pair consisting of a set of worlds W and a valuation function C that assigns truth values to *all* formulas at all worlds.

The crucial difference from standard Kripke models is that the truth values of compound formulas are determined directly from the syntactic valuation C instead of being computed recursively, based on the valuation of atomic formulas. As a result, syntactic structures can be considered generalizations of standard Kripke models as accessibility relations are no longer relevant in obtaining the truth value of formulas such as $K\phi$. In this case, we can view each Kripke model $\langle W, R, V\rangle$ as a syntactic structure $\langle W, C\rangle$ such that $C(w)(\phi) = 1$ whenever $M, w \vDash \phi$.

It is not hard to see how the syntactic approach deals with all forms of the problem. First and regarding the full form, the value of $K\phi$ is not affected by the truth values of the formulas in Ψ nor by the Logical Implication from Ψ to ϕ. More specifically, *Knowledge of valid formulas* fails because the construction of the valuation function could be such that the truth values of ϕ and $K\phi$ diverge. *Closure under Logical Implication*, *Closure under Material Implication*, *Closure under Valid Implication* likewise fail because the value of $K\psi$ can be suitably tailored.

[2] A doxastic counterpart of this approach can be easily obtained. Belief may just replace knowledge in what follows.

The truth values of $K\phi$ and $K\psi$ do not have to agree, just because ϕ and ψ are logically equivalent, therefore *Closure under Logical Equivalence* might fail. The values of $K\phi$ and $K\psi$ do not put any constraints on the values of $K(\phi \wedge \psi)$ and $K(\phi \vee \psi)$, thus avoiding *Closure under Conjunction* and *Closure under Disjunction*. Finally, we can also invalidate the consistency of knowledge, i.e. $\neg K(\phi \wedge \neg\phi)$, by considering the independence of values between $\phi \wedge \neg\phi$ and $\neg K(\phi \wedge \neg\phi)$.

Given that our prime interest lies in invalidating the closure principles, we might want to preserve the standard account as far as propositional connectives are concerned. It is therefore reasonable to impose constraints such as (a) $C(w)(\neg\phi) = 1$ if and only if $C(w)(\phi) = 0$, and (b) $C(w)(\phi \wedge \psi) = 1$ if and only if $C(w)(\phi) = 1$ and $C(w)(\psi) = 1$.

However, the syntactic response to the problem is not entirely satisfactory, despite avoiding all forms of it. Even by imposing the constraints and given the connection to the standard epistemic models, it cannot describe any interesting property of knowledge and belief. The only formulas that are valid in such structures are propositional tautologies. By assigning arbitrary truth values we indeed avoid the problem but our understanding of propositional attitudes is not facilitated because no interesting philosophical benefits can be reaped from this kind of formalization. More importantly, since knowledge or belief assertions are assigned truth values arbitrarily, there is nothing to ensure that agents know or believe at least *some* consequences of what they know or believe, thus the desideratum on capturing logical competence is not fulfilled. Had we attempted to preserve the epistemologically interesting properties or add desired elements of a realistic portrayal via suitable modifications of the syntactic valuation, we would have ended up with an ad-hoc, artificial and unnatural embedding of the standard modelling device in the syntactic structures. This obviously lacks independent motivation and presupposes an acknowledgement of the superiority of the standard epistemic-doxastic systems, which goes against the very project of proposing a fuller and more attractive alternative.

3.1.2 Rasmussen

Another syntactic attempt, now constructing a dynamic logic whose axiomatization is such to escape the problem, is described in Rasmussen (2015). The author emphasizes that the source of the problem lies in the difficulty to jointly satisfy the following two requirements:

(R1) The knowledge of resource-bounded agents is not closed under any non-trivial logical law (*Non-Closure*).

(R2) If a resource-bounded agent knows the premises of a valid inference and knows the relevant inference rule, then, given sufficient resources, the agent can infer the conclusion (*Non-ignorance*).

According to this diagnosis, any approach that solely designs a static framework is destined to be inferior in terms of realistic modelling. Static systems cannot effectively approximate real-life situations because they neglect the reasoning process that resulted in a particular epistemic or doxastic state. This is why Rasmussen builds on Duc's *dynamic epistemic logic* (Duc (1997)), who augments the standard epistemic language by dynamic operators $\langle F_i \rangle$, such that $\langle F_i \rangle\phi$ reads "ϕ is true after some reasoning process performed by agent i". The main idea is that while the necessitation rule "from ϕ infer $K\phi$" is not derivable in this logic, thereby avoiding a form of omniscience, the rule "from ϕ infer $\langle F_i \rangle K_i \phi$" is derivable, thereby showing that agents can indeed *come to know* validities but *only* if they think "hard enough". Rasmussen expands this idea, aiming at a full description of an agent's reasoning process. This task boils down to accounting for (i) the specific applications of inferences rules involved in a reasoning process, (ii) the chronology of these applications of inference rules, (iii) the cognitive cost of each application of an inference rule. With this observation in mind, the logical language $\mathcal{L}_D(\Phi)$, and the axiomatization of the proposed logic L_D are defined as follows:

Definition 3.1.2 (Language $\mathcal{L}_D(\Phi)$, Rasmussen (2015))**.** The language $\mathcal{L}_D(\Phi)$ is defined inductively from a set of atomic sentences Φ, a knowledge operator K, and a set of dynamic operators $\langle R_i \rangle^{\lambda_i}$ for $1 \leq i \leq n$ as follows:

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi \to \phi \quad | \quad K\phi \quad | \quad \langle R_i \rangle^{\lambda_i}$$

with $p \in \Phi$.

The dual modality $[R_i]^{\lambda_i}\phi$ is defined as $\neg\langle R_i \rangle^{\lambda_i}\neg\phi$. Then, $\langle R_i \rangle^{\lambda_i}\phi$ intuitively reads "ϕ is the case after *some* application of R_i at cognitive cost λ_i", with "any" replacing "some" for the dual case. Cognitive costs can be thought of as natural numbers.

In order to present the axiomatization, the following abbreviations are used to denote arbitrary sequences of dynamic operators[3]:

$$\langle \ddagger \rangle^i := \langle R_i \rangle^{\lambda_i} \ldots \langle R_j \rangle^{\lambda_j}$$

$$[\ddagger]^i := [R_i]^{\lambda_i} \ldots [R_j]^{\lambda_j}$$

where $R_i, \ldots, R_j$ are arbitrary inference rules and $i = \lambda_i + \ldots + \lambda_j$. The first abbreviation intuitively says that "after some application of R_i at cognitive cost λ_i followed by … followed by some application of R_j at cognitive cost λ_j, ϕ is the case. For the intuitive reading of the second abbreviation, again replace "some" by "any".

Definition 3.1.3 (Axiomatization of L_D, Rasmussen (2015))**.** Let $\phi, \psi \in \mathcal{L}_D(\Phi)$, $\Gamma \subseteq \mathcal{L}_D(\Phi)$, and $\langle \ddagger \rangle^i, \langle \dagger \rangle^j$ (also, $[\ddagger]^i, [\dagger]^j$) denote arbitrary sequences of dynamic operators. The logic L_D has the following axiom schemes:

- (PC) All substitution instances of propositional tautologies.
- (A1) $\langle \ddagger \rangle^i K\phi \to \phi$ (*Veridicality*)
- (A2) $\langle \ddagger \rangle^i K\phi \to \langle \ddagger \rangle^i [\dagger]^j K\phi$ (*Persistence*)
- (A3) $\langle \ddagger \rangle^i \phi \wedge \langle \dagger \rangle^j \psi \to \langle \ddagger \rangle^i \langle \dagger \rangle^j (\phi \wedge \psi)$ (*Succession*)
- (A4) $\langle \ddagger \rangle^i (\phi \wedge \psi) \to \langle \ddagger \rangle^i \phi$ (*Elimination*)

L_D has the following inference rule:

- (MP) If $\Gamma \vdash \phi$ and $\Gamma \vdash \phi \to \psi$ then $\Gamma \vdash \psi$.

[3] Note that a sequence can be empty.

It is evident that for a doxastic framework, axiom (A1) might be dropped at it only imitates the veridicality axiom (T) usually adopted for standard epistemic systems. (A2) says that known sentences remain known as reasoning progresses. This still presupposes two idealizing assumptions (a) on agent's infallible memory, and (b) on sentences preserving a truth value throughout the whole reasoning process. (A3) is imposed to express that a reasoning process can succeed another, and finally yield the conjunction of their outcomes. (A4) simply states that ϕ is the case after a reasoning process if both ϕ and ψ are the case after it.

An extension of L_D, denoted by L_D^Λ, comprises appropriate axioms for specific inference rules from a set Λ, with which the agent is equipped. For illustrative purposes, consider Modus Ponens (MP), Conjunction Introduction (CI) and Double Negation Elimination (DNE), as the inference rules in Λ. The main idea behind the axiomatization is that the agent can come to know certain formulas with the additional cognitive cost of applying an inference rule. For example, if the agent knows ϕ and $\phi \to \psi$ then the agent can derive ψ by applying Modus Ponens at a particular cognitive cost. Using the abbreviation $\Delta := \phi \wedge \ldots \wedge \psi$ to denote arbitrary conjunctions in the language, L_D^Λ is axiomatized as follows:

Definition 3.1.4 (Axiomatization of L_D^Λ, Rasmussen (2015))**.** Let $\Lambda = \{MP, CI, DNE\}$ and let Δ be an arbitrary conjunction of sentences in $\mathcal{L}_D(\Phi)$. Furthermore, let μ, κ, ν denote the cognitive costs of MP, CI, DNE respectively. L_D^Λ extends L_D with the following axiom schemes:

- (MP_D) $\langle \ddagger \rangle^i (\Delta \wedge K\phi \wedge K(\phi \to \psi)) \to \langle \ddagger \rangle^i \langle MP \rangle^\mu (\Delta \wedge K\phi \wedge K(\phi \to \psi) \wedge K\psi)$ (MP*-success*)
- (CI_D) $\langle \ddagger \rangle^i (\Delta \wedge K\phi \wedge K\psi) \to \langle \ddagger \rangle^i \langle CI \rangle^\kappa (\Delta \wedge K\psi \wedge K\phi \wedge K(\phi \wedge \psi))$ (CI*-success*)
- (DNE_D) $\langle \ddagger \rangle^i (\Delta \wedge K\neg\neg\phi) \to \langle \ddagger \rangle^i \langle DNE \rangle^\nu (\Delta \wedge K\neg\neg\phi \wedge K\phi)$ (DNE*-success*)

Of course, the same pattern can be generalized for any inference rule R with premises $\phi_1, \ldots, \phi_n$ and conclusion ψ, and a cognitive cost λ:

$$(R_D) \quad \langle \ddagger \rangle^i (\Delta \wedge K\phi_1 \ldots \wedge K\phi_n) \to \langle \ddagger \rangle^i \langle R \rangle^\lambda (\Delta \wedge K\phi_1 \ldots \wedge K\phi_n \wedge K\psi)$$

The following theorem of the extended logic essentially distinguishes Duc's logic from L_D^Λ and manifests the accuracy of this framework in describing reasoning processes:

Theorem 3.1.1 (Application, Rasmussen (2015))**.**

$$K\neg\neg\phi \wedge K(\phi \to \psi) \to \langle DNE \rangle^\nu \langle MP \rangle^\mu \langle CI \rangle^\kappa K(\phi \wedge \psi)$$

is a theorem of L_D^Λ.

The theorem illustrates the dynamic nature of reasoning by keeping track of the applications of the inference rules, their chronology and their cognitive costs. Additionally, it does so by taking into account the complexity of an agent's deduction. That is, unlike Duc's system, we can account for the fact that not all deductions are equally hard and the evolution of reasoning is thus described in a more elaborate way.

We are now ready to see how this approach deals with the problem, given the two desiderata raised above. First, consider that $\Lambda = \varnothing$: *Non-closure* is obviously satisfied because agents are, by definition, incapable of applying any inference rules, hence knowledge cannot be closed under any logical law. Also, *Non-ignorance* is trivially satisfied as there are no requirements on the inferential abilities of the agent if she does not have any inference rules available. Next, consider that $\Lambda \neq \emptyset$; taking an arbitrary inference rule R with premises $\phi_1, \ldots, \phi_n$ and conclusion ψ, the R-closure $K\phi_1 \wedge \ldots K\phi_n \rightarrow K\psi$ is not a theorem of our logic because if $R \in \Lambda$, then its corresponding axiom is of the form:

$$\langle \ddagger \rangle^i (\Delta \wedge K\phi_1 \wedge \ldots K\phi_n) \rightarrow \langle \ddagger \rangle^i \langle R \rangle^\lambda (\Delta \wedge K\phi_1 \ldots K\phi_n \wedge K\psi)$$

which says that the agent needs to *reason* to attain knowledge of ψ, unlike R-closure which dictates that whenever the agent knows $\phi_1, \ldots, \phi_n$, she automatically knows ψ, too. If $R \notin \Lambda$ then the result trivially holds. On the other hand, *Non-ignorance* is satisfied because the axiom does predict how, given sufficient resources, the agent can derive the relevant conclusion.

Interestingly, this approach captures both non-omniscience and non-ignorance. Of course, syntactic manipulations, and in particular the introduction of cognitive costs, allow us to modify the system as we please. Yet the problem is usually semantically retained because it is precisely the commitment to a notion of truth that poses the challenge. By merely modifying the axioms, without providing semantics, it seems that Rasmussen bites the bullet. In particular, it is easy to see that no trivial possible-worlds framework could work. That is, how a model would change as the outcome of an application of sequences of inference rules is not a trivial matter. What can make it even more challenging is capturing what the effect of different – quantitative – cognitive costs would be on such a model. Obtaining the validity of the proposed axioms and theorems then remains an open issue. This renders the choice of axioms somewhat controversial: why are *these* formulas the most appropriate to capture the desired features of reasoning? What is more, and again in the absence of the useful properties of possible-worlds semantics, we miss out on interesting properties about knowledge and belief. Moreover, the way the desideratum on agents' rationality is stated, i.e. *Non-ignorance*, raises suspicion about its adequacy. Although it is tailored in a way that matches the subsequent axiomatization[4], there are reasons to assume that even when given sufficient resources, e.g. infinite time, agents will still be fallible. For example, they might just be biased or reluctant to reason according to the logical rules. In this sense, it is worth emphasizing that Rasmussen sheds light merely on what an agent *can* do – what is in principle affordable when certain resources are available – and what the agent cannot do – when running out of those. The axiomatization, though, says little on what an agent *ought to do*, in general. Despite the normative nature of the notion of competence as suggested in the beginning of this chapter, Rasmussen's characterization and solution remain largely descriptive. All in all, his approach only partially overcomes the problem.

3.2 Implicit versus explicit attitudes

It has been argued that the problem of logical omniscience is in fact an indication for a distinction between *implicit* and *explicit* propositional attitudes. For example, Levesque (1984) suggests that closure principles do not refer to what we *actually* know or believe but rather to another kind of concept: what is implicit in what we know or believe, even without us realizing it.

> [...] if an agent imagines the world to be one where α is true and if α logically implies β, then (whether or not he realizes it) he imagines the world to be one

[4] This might even exacerbate our worries on the explanatory power and motivation underlying the axioms; an objection of ad-hocness is unavoidable.

where β also happens to be true.[5]

In addition, the distinction between implicit and explicit attitudes is endorsed by results in psychology. For example, it has been documented (Schacter (1987), Schacter and Tulving (1994)) that *implicit memory* affects our performance as we resort to our accumulated experience without making a conscious recall, unlike *explicit memory* that usually involves a deliberate and conscious act of recall. One could therefore argue that there are different shades of knowledge, corresponding to different ways it can be acquired.

In any case, we need to tell apart what is explicitly, directly known/believed by an agent and what the world would be like if what she knows/believes was true, even if its consequences are not consciously accessible to her. The approaches below promise to model this difference and show that logical omniscience is overruled with respect to the former while remaining an unproblematic, and even desired, feature with respect to the latter.

3.2.1 Awareness

One of the ways that have been proposed to formally capture this distinction is based on the notion of *awareness*. According to this view, as presented in Fagin and Halpern (1987) and Fagin et al. (1995a), agents are not logically omniscient because they cannot believe/know things they are not aware of. In order to account for the concepts that the agent is aware of, the standard single-agent language is supplemented with an operator A such that $A\phi$ reads "the agent is aware of ϕ" and, also, with an operator K^e such that $K^e\phi$ reads "the agent explicitly knows ϕ". Next, our familiar models are enriched by an additional component:

Definition 3.2.1 (Awareness structure, Fagin et al. (1995a))**.** An *awareness structure* is a Kripke model M augmented by an awareness function $\mathcal{A}$, i.e. a tuple $\langle W, R, V, \mathcal{A}\rangle$, such that $\mathcal{A}$ associates a set of formulas from our extended language with each world $w \in W$. Intuitively, this set is the set of all formulas that the agent is aware of at w.

The awareness function behaves arbitrarily; for instance, an agent might be aware of both ϕ and $\neg\phi$, or aware of $\phi \wedge \psi$ but unaware of ϕ. Once we obtain these new structures, we can turn to the truth clauses of $A\phi$ and $K^e\phi$.

Definition 3.2.2 (Semantics for awareness, Fagin et al. (1995a))**.** The definition and notation follow that of Definition 1.1.3, only now supplemented with:

- $M, w \vDash A\phi$ if and only if $\phi \in \mathcal{A}(w)$.
- $M, w \vDash K^e\phi$ if and only if $M, w \vDash K\phi$ and $M, w \vDash A\phi$.

Unsurprisingly, for the case of belief $B^e\phi$ (once the operators B^e and B are added to the language to denote explicit and implicit belief) we substitute $K\phi$ with $B\phi$ in the last clause, i.e. $B^e\phi$ boils down to $B\phi$ and $A\phi$.

Consequently, while implicit knowledge remains subject of the closure principles that yield logical omniscience, the problem can be avoided in its explicit form. For example, if an agent explicitly knows ϕ, and ϕ logically implies ψ, then it need not be that she explicitly knows ψ; although $K\psi$ still holds, it might be that $\psi \notin \mathcal{A}(w)$ for some world of the model. Likewise, not all valid formulas are explicitly known/believed as an agent might not be aware of them.

[5] Levesque (1984), p. 198.

Through manipulations of the awareness function, the remaining closure principles can be similarly destroyed.

In the face of criticisms stemming from the arbitrariness of the awareness function, and resembling the considerations discussed in Section 3.1.1, it seems reasonable to impose certain restrictions on what an agent is aware of:

1. $\phi \wedge \psi \in \mathcal{A}(w)$ if and only if $\psi \wedge \phi \in \mathcal{A}(w)$, to ensure that the awareness function is coarse-grained enough to disregard the order of conjuncts.
2. $\phi \in \mathcal{A}(w)$ if and only if $\neg\phi \in \mathcal{A}(w)$, to ensure that an agent cannot be aware of ϕ without being aware of its negation and vice-versa.
3. If $\phi \in \mathcal{A}(w)$ and ψ is a subformula of ϕ then $\psi \in \mathcal{A}(w)$, to ensure that awareness of a formula presupposes awareness of its constituent parts.
4. If $\phi \in \mathcal{A}(w)$, then $A\phi \in \mathcal{A}(w)$, to ensure that an agent self-reflects and is therefore aware of what she is aware of.
5. If wRu then $\mathcal{A}(w) = \mathcal{A}(u)$, to ensure that an agent knows the formulas she is aware of[6].
6. $\mathcal{A}(w)$ contains exactly those formulas that the agent can decide on whether they follow from the information at w, in some given period of time.

Next, it is important to note that, unlike syntactic structures, awareness logics preserve useful results of the standard systems and uncover some interesting properties of knowledge and belief.

Theorem 3.2.1 (Axiomatization for awareness, Fagin et al. (1995a))**.** By adding the axiom $K^e\phi \leftrightarrow A\phi \wedge K\phi$ to the axioms and rules of **K**, we can obtain a sound and complete axiomatization for the logic of awareness [7].

Specifically, we can obtain properties for explicit knowledge corresponding to axiom (K) and inference rule (N), once we relativize to awareness: i.e. consider $K^e\phi \wedge K^e(\phi \rightarrow \psi) \wedge A\psi \rightarrow K^e\psi$ and "from ϕ infer $A\phi \rightarrow K^e\phi$". Analogous instances can be obtained for the Positive Introspection axiom and the Negative Introspection axiom, but only if we further demand that the aforementioned restriction 5 holds.

To sum up, this approach manages to invalidate the problematic cluster of principles via attaching the intuitively plausible requirement of awareness to (explicit) knowledge. It also naturally preserves elements of the standard system. However, logical competence of agents is not preserved; it can easily be the case that an agent fails to know a trivial consequence of her knowledge merely due to the construction of the awareness function. Even if we resort to restrictions listed above, e.g. the reasonable assumption of awareness closure under subformulas, the problem is somehow retained, e.g. *Closure under Material Implication* persists. Indeed, in this case, awareness of ϕ and awareness of $\phi \rightarrow \psi$ yields awareness of ψ too.

[6] Demanding that $\mathcal{A}(w) \subseteq \mathcal{A}(u)$ whenever wRu can suffice. Given that standard epistemic models are symmetric and since this account can be preserved for (implicit) knowledge, we can immediately adopt the other inclusion as well.

[7] Clarifying that this is the case for the plain awareness system, i.e. without any of the restrictions mentioned above. Furthermore, in Fagin and Halpern (1987) we can find a similar axiomatization (for belief) building on **KD45**, provided of course that the accessibility relation of the awareness structures satisfies the corresponding properties.

Combining this with the implicit *Closure under Material Implication*, that ensures that $K\psi$, we get that $K^e\psi$. Avoiding such plausible restrictions – that can be seen as attempts to preserve logical competence – in the pain of logical omniscience, unveils the difficulty to generate a balanced account via awareness. What is more, despite the close connection to the standard systems, *awareness* still forfeits desirable properties of logical modelling such as the correspondence between properties of accessibility relations and propositional attitudes, as pointed out in Konolige (1988). Suppose for example that $B^e\phi$ is true and consider the introspective explicit belief $B^eB^e\phi$. Its truth depends on the status of both $A\phi \to BA\phi$ and $B^e\phi \to AB^e\phi$, none of which reflecting properties of doxastic accessibility. It is therefore clear that several aspects of this approach are exposed to criticism.

3.2.2 Algorithmic Knowledge

According to the *algorithmic knowledge* approach (hereafter adapted from Halpern and Pucella (2011), Halpern et al. (1994)), knowledge is attained via a *knowledge algorithm*. Given a language formed as in the previous section, only without the awareness operator, the algorithm takes a formula as input and outputs "yes", if the agent computes that the formula is true and "no", if the agent computes that the formula is false (and "?" otherwise). Formally:

Definition 3.2.3 (Algorithmic knowledge structure)**.** An *algorithmic knowledge structure* is a tuple $M = \langle W, R, V, A\rangle$ where $\langle W, R, V\rangle$ is a Kripke structure and A is a *knowledge algorithm* that returns "yes", "no" or "?", given a formula ϕ.

Definition 3.2.4 (Semantics for algorithmic knowledge)**.**

- The clauses of Definition 1.1.3.
- $M, w \vDash K^e\phi$ if and only if $A(\phi)$="yes".

As a result, the agent attains explicit knowledge only of those facts that she can explicitly compute. Since knowledge algorithms are not subject to any restriction, an agent can explicitly know both ϕ and $\phi \to \psi$ without explicitly knowing ψ; or an agent can explicitly know ϕ, which is logically equivalent to ψ, without explicitly knowing ψ. That is, the algorithmic knowledge approach also escapes the problem in a syntactic manner, using algorithms to capture the notion of explicit knowledge and utilizing their arbitrariness to differentiate it from its implicit counterpart[8].

An important class of knowledge algorithms consists of *sound knowledge algorithms*. When a sound algorithm outputs "yes" given an input ϕ, then the agent knows ϕ, in the standard sense. When a sound algorithm outputs "no" given an input ϕ, then the agent does not know ϕ, in the standard sense.

Definition 3.2.5 (Sound Algorithm)**.** An algorithm is *sound in a structure M* if for any world w in M and any formula ϕ: $A(\phi)$= "Yes" implies $M, w \vDash K\phi$ and $A(\phi)$="No" implies $M, w \vDash \neg K\phi$.

We have good reasons to restrict our attention to sound algorithms since in this case $K^e\phi \to K\phi$ and thus $K^e\phi \to \phi$ are valid, both reflecting desirable properties of explicit knowledge. But then it can be easily seen that algorithmic knowledge structures can be reduced to instances of awareness and thereby inherit the criticisms of the previous section.

[8] This strategy can be trivially extended to belief.

3.2.3 Justification Logic

Justification Logic is another framework that focuses on explicit attitudes, developed to reason about epistemic justification. The starting point for its endeavor is to supplement the language of standard propositional logic with justification assertions $t : F$, interpreted as "t is a justification for F". In this sense, it introduces an epistemologically anticipated component of justification to the conventional epistemic logic accounts, that solely deal with *what* is known, neglecting *why* it is known. It is precisely in this notion of justification that the "explicitness" lies.

Before continuing with Justification Logic, based on the paradigmatic expositions in Artemov and Fitting (2016), Artemov (2011) and Artemov (2008), we first need definitions of *justification terms* and *justification formulas*. We gather these in:

Definition 3.2.6 (Language of Justification Logic)**.** For countable sets of justification constants, justification variables and propositional atoms, justification terms are built as follows:

$$t \quad ::= \quad c_i \quad | \quad x_i \quad | \quad t \cdot t \quad | \quad t + t$$

We use c_i to denote the justification constants and x_i to denote the justification variables.

Formulas are built as follows:

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi \to \phi \quad | \quad t : \phi$$

where p denotes a propositional atom.

In order to grasp this definition, one should think of justification constants as unanalyzable justifications and of justification variables as unspecified justifications. Applying a justification term to a formula, e.g. as in $t : \phi$, can take the intuitive interpretation of t being a justification for ϕ. In our case, the standard knowledge assertion "ϕ is known" is replaced by "t is a justification of ϕ". In this sense, justification terms may be viewed as the usual modal operator $\Box$, expanded with labels. As we will see, the treatment of these terms is no different to the treatment of the modal operators. The conceptual difference is that now the notion of knowledge becomes more expressive, as it is reinforced with an evidence-based foundation.

Regardless of the intuitive reading, we can always define operations between justification terms. The operation symbol $\cdot$ indicates an one-step deduction according to Modus Ponens, that is, given justifications s and t, if s is a justification of $\phi \to \psi$ and t a justification of ϕ, then $s \cdot t$ is produced, as a justification of ψ. Formally this is written as $s : (\phi \to \psi) \to (t : \phi \to (s \cdot t) : \psi)$. For example, Artemov (2011) remarks that if justifications are taken as Hilbert-style proofs, then $s \cdot t$ can be seen as a new proof obtained by s and t by performing Modus Ponens to all possible premises $\phi \to \psi$ from s and ϕ from t. The operation symbol $+$ says that, given justifications s and t, $s + t$ is a justification produced for everything justified by s or t. Formally we write $s : \phi \to (s + t) : \phi$ and $s : \phi \to (t + s) : \phi$. For example, $s + t$ can be thought of as a body of evidence that "gathers" evidence without performing logical inferences. Specifically, $+$ aggregates all the evidence provided by s and t so that the resulted $s + t$ supports everything supported by s or t.

With this material in hand, we can continue with specific Justification Logic systems.

Definition 3.2.7 (Basic Logic, Artemov (2011))**.** The Basic Logic J_0 is axiomatized by:

- (CP) All instances of classical propositional tautologies.

- (Application Axiom) $s : (\phi \to \psi) \to (t : \phi \to (s \cdot t) : \psi)$.
- (Sum Axioms) $s : \phi \to (s + t) : \phi$ and $s : \phi \to (t + s) : \phi$.
- (MP) The inference rule Modus Ponens.

In addition, if we want to postulate that an axiom A has a justification for the agent, we postulate $e_1 : A$ for some justification constant e_1. Continuing in the same manner, by $e_n : e_{n-1} : \ldots : e_1 : A$ we postulate that e_n is a justification for $e_{n-1} : \ldots : e_1 : A$.

Definition 3.2.8 (Constant Specification, Artemov (2011))**.** A *Constant Specification* (CS) for a given justification logic L is a set of formulas $e_n : e_{n-1} : \ldots : e_1 : A$ $(n \geq 1)$, where A is an axiom of L and $e_1, \ldots, e_n$ are justification constants with indices $1, \ldots, n$. We also assume that if $e_n : e_{n-1} : \ldots : e_1 : A$ is in CS then $e_{n-1} : \ldots : e_1 : A$ is in CS, that is, intermediate specifications are included in a given CS.

Note that the justification terms of the basic system are not necessarily factive. In order to incorporate *Factivity*, one has to add the relevant axiom $t : \phi \to \phi$, thereby getting an extension of the basic system. Of course, there can be further operations such as the *Positive Introspection operation '!'*, with $!t$ interpreted in such a way that if an agent accepts t as sufficient justification for ϕ, then $!t$ serves as sufficient justification for $t : \phi$. Subsequently, other axioms can be added (in our example, $t : \phi \to !t : (t : \phi)$), resulting in further extensions. A case in point is LP, Gödel's *Logic of Proofs*, which can be seen as the extension of the basic logic with the axioms: $t : \phi \to \phi$ and $t : \phi \to !t : (t : \phi)$. Such extensions already hint at a particular correspondence with the standard modal logics such as **T**, **K4**, **S4**, **KD45**, **S5** etc. For example, the extension of the basic system axiomatized by the positive introspection axiom has **K4** as a modal-logic counterpart[9].

We now proceed with the semantical account of Justification Logic, based on Artemov and Fitting (2016):

Definition 3.2.9 (Fitting model)**.** A *Fitting* model is a model $M = \langle W, R, E, V \rangle$ where W is a non-empty set of worlds, R is an accessibility relation on W and V a valuation function that maps a propositional atom to a set of worlds (containing those worlds in which the atom is true). E is an *evidence function*, which takes as arguments a justification term t and a formula ϕ and maps them to a set of worlds; these are the worlds where t is *admissible* (or relevant) evidence for the formula ϕ. This function additionally satisfies the following:

- (Application) $E(s, \phi \to \psi) \cap E(t, \phi) \subseteq E(s \cdot t, \psi)$
- (Sum) $E(s, \phi) \cup E(t, \phi) \subseteq E(s + t, \phi)$

The first condition ensures the validity of the Application Axiom, whereas the second ensures the validity of the Sum Axioms. For justification logics with a constant specification, we also require:

- (Constant Specification) If $c : X \in CS$ then $E(c, X) = W$

By virtue of the definition of *Constant Specification*, it is clearly understood that axioms are justified *ex officio* and what is justified in this way cannot be analyzed any further; this is why

[9] For a detailed investigation of correspondence results, consult Fitting (2014).

constants (or arrays of constants) are employed. This additional clause precisely manifests what constants were introduced for: they denote the reasons for "atomic" assumptions that are universally accepted and therefore available throughout the whole model.

Definition 3.2.10 (Semantics for justification logics)**.** The definition of a formula ϕ being *true at world w in a (Fitting) model M*, denoted by $M, w \models \phi$, follows the usual inductive procedure for the propositional part. Furthermore, an extra clause is added to account for the justification terms:

$$M, w \models t:\phi \quad \text{if and only if for all } v \in W \text{ with } wRv\ M, v \models \phi \text{ and } w \in E(t, \phi)$$

This means that $t:\phi$ is true at world w if ϕ is true at all accessible worlds *and* t is admissible evidence for ϕ at the world w. The first part imitates the usual hintikkian characterization. The second part imposes another constraint: evidence t should be admissible for ϕ at w. It can be therefore stated that this characterization is "explicit" in the sense that the traditional (implicit) condition is augmented by another condition regarding the features of the available evidence. Consequently, one might fail to know something for a particular reason at a world, either because this is not "knowable" already in the standard sense, or because the reason that backs it, is irrelevant.

Once we have presented the main elements of Justification Logic, we can continue with its contribution towards the problem of logical omniscience. To do so in a most comprehensive manner, we start with the critique facing other epistemic systems, as fired from the justification camp; it is in these terms that its proponents then build their own solution. Specifically, in Artemov and Kuznets (2009) the inherent defect of the standard logics is located in that they represent knowledge without taking its origin into account. As a result, the fact that acquiring knowledge is a subject of certain restrictions in the agents resources (e.g. time, memory etc.) is underestimated. In Artemov and Kuznets (2013), the authors also locate the insufficiency of other strategies against omniscience, exemplified in sacrificing agents' rationality, in their qualitativeness: allowing or prohibiting knowledge assertions is determined by lists of the known formulas, the formulas the agent is aware of etc. This is why the justification camp argues for a quantitative method, that uses justifications to navigate in the internal inferential process of the agents and suggests approaching the problem via proof and time complexity. In this way, one can capture that rational agents can successfully make simple, small inferences yet complex chains of inferences, despite consisting of smaller ones, might still be inaccessible, thereby making agents non-omniscient. To sum up, the authors set the bar of their solution in obtaining a quantitative criterion, that (a) targets the insufficiency of having knowledge assertions without accompanying feasible justifications, and (b) discerns omniscient and non-omniscient systems based on *how much* information about the background reasoning is required to avoid logical omniscience.

Let's now delve into their own proposal. As mentioned above, it is essential to ensure that agents are fundamentally rational. According to the current proposal, what usually prevents them from acquiring knowledge is the boundedness of resources. In order to express boundedness, one has to gain an insight into the resources and quantify over them. In other words, for succeeding in knowing ϕ in an epistemic system, the "cost" needed to achieve this should be somehow measured. The size of the internal proof that allows stipulating knowledge of ϕ is an adequate such measure [10]. Logical omniscience then arises whenever: "for some 'short'

[10] Of course, there might be several interpretations on what is meant by "proof-size". For example, assuming

valid knowledge assertions F *is known*, it is impossible to feasibly find proofs of F in $\mathcal{E}$."[11]

This observation and the search of a precise characterization of what "feasibly find proofs of F" means, fuel the design of two tests, called LOT (Logical Omniscience Test) and SLOT (Strong Logical Omniscience Test) aiming in detecting whether a system gives rise to the problem. As the authors put it in Artemov and Kuznets (2009), p.14.:

> "An epistemic system $\mathcal{E}$ is *not logically omniscient* if for any valid-in-$\mathcal{E}$ knowledge assertion A of type F *is known*, there is a proof of F in $\mathcal{E}$, the complexity of which is bounded by some polynomial in the size of A."
>
> "We suggest a more general Strong Logical Omniscience Test (SLOT) based on time complexity: an epistemic system $\mathcal{E}$ is *not logically omniscient* if for any valid-in-$\mathcal{E}$ knowledge assertion A of type F *is known*, a proof of F in $\mathcal{E}$ can be found in polynomial time in the size of A."

We can already make some initial remarks. Both tests demonstrate that there is a link between the size of the knowledge assertion and the ability of the system to feasibly produce sufficient evidence for the object of knowledge. The difference is that LOT uses the complexity of the proof of the object of knowledge while SLOT uses the time required to find the proof. Given that a proof found in polynomial time in the size of the assertion is of polynomial-size, the strong test indeed entails the weak. Furthermore, both tests depend on the proof system and the measure of the size of formulas. Yet, LOT additionally depends on the measure of the size of proofs.

In order to clarify how the tests work, we proceed with a formal presentation. Some preliminary definitions make the notions involved in the tests precise. First, let L a logic of some language $\mathcal{L}$.

Definition 3.2.11 (Preliminaries, Artemov and Kuznets (2009))**.**

- *A proof system for L* is a polynomial-time computable function $\mathcal{E} : \Sigma^* \to L$ from the set of words in some alphabet, called proofs, onto the set of L-valid formulas.
- *A measure of size for proofs* is a function $l : \Sigma^* \to \mathbb{N}$.
- *A measure of size for individual formulas* $|\cdot| : \mathcal{L} \to \mathbb{N}$ (e.g. number of logical symbols in the formula).
- L is called an *epistemic system* if some subset $r\mathcal{L} \subseteq \mathcal{L}$ is taken as a set of *knowledge assertions*. Each knowledge assertion $A \in r\mathcal{L}$ has an intended meaning "formula F is known" for a unique formula F. The function $OK : r\mathcal{L} \to \mathcal{L}$ that extracts the object of knowledge from each knowledge assertion is required to be computable in polynomial time in $|A|$ and validity-preserving: for any $A \in r\mathcal{L}$, $L \vdash A \Rightarrow L \vdash OK(A)$.
- *The reflected fragment rL* is the set of all valid knowledge assertions: $rL = L \cap r\mathcal{L}$.

Definition 3.2.12 (LOT and SLOT, Artemov and Kuznets (2009))**.** For $\mathcal{E}$ a proof system for an epistemic system L, or simply an *epistemic proof system*, we define:

Hilbert-style proof systems, the size measures can be associated with the number of proof-steps or the logical symbols in a derivation.

[11] For $\mathcal{E}$ an epistemic system; Artemov and Kuznets (2009), p. 14.

- (LOT): $\mathcal{E}$ is not logically omniscient or passes LOT, if there exists a polynomial P such that for any valid knowledge assertion $A \in rL$, there is a proof of $OK(A)$ in $\mathcal{E}$, with the size bounded by $P(|A|)$.
- (SLOT): $\mathcal{E}$ is strongly not logically omniscient or passes SLOT, if there is a deterministic algorithm, polynomial in $|A|$, that, for any valid knowledge assertion $A \in rL$ is capable of restoring a proof of $OK(A)$ in $\mathcal{E}$.

It now becomes clearer that "feasibly find proofs in $\mathcal{E}$", according to LOT, points at using the proof size of the assertion. This is so because LOT wants to make explicit the difference between "I know that Goldbach's Conjecture holds" and "Such-and-such make Goldbach's Conjecture hold". In the first case, the knowledge assertion is short and as a result there is a deficit of "raw material" to construct a proof. In the second case, the knowledge assertion contains some array of justifications, that provide information to verify this response to the conjecture. Accordingly, SLOT strengthens the requirements. More information on the background reasoning should be encoded, if we are to avoid omniscience. An increased size of a knowledge assertion suggests that finding a proof is also fast because we already have footprints for the construction of the proof: the justification terms. For example, knowing a winning strategy of a game would require that the assertion encodes enough information to retrieve a proof that narrates each of the steps necessary to reach victory. In a game like tic-tac-toe, storing this information can be easy, assuming that the symmetries on the table are exploited to prove that the strategy is indeed winning(or non-losing). In a game like chess, this is –as far as we currently know– unattainable because the complexity of proving that a strategy is winning (or non-losing) remains infeasible. A non-omniscient system would be able to discern these cases, unlike an omniscient one.

Given these tests, we are able to make some exegetical points on why justification logics avoid omniscience. In particular, the goal is to show that agents are logically omniscient under standard epistemic systems but not under justification ones. In the former, the difficulty and the internal effort to actually reach knowledge is not reflected in the language; no matter how hard it was for knowledge to be gained, these logics just do away with it, by using the K operator. On the other hand, the latter provide information on why F is known: this information, if included in the assertion, is sufficient to recover feasible proofs for F. In case a proof cannot be constructed, then knowledge acquisition fails. Overall, by employing LOT or SLOT and the knowledge assertions with incorporated justification terms, we can maneuver through a non-idealized agent's proof search and then explain her non-omniscience in terms of the resource-boundedness that affects this process.

Indeed, using these tests of (non-)omniscience, agents of modal epistemic logics, such as **S4**, are proved to be omniscient with respect to both LOT and SLOT. On the other hand, agents of justification logics are proved to be *not* logically omniscient. For instance, consider LP, the Hilbert-style proof system and the size of a proof being the number of formulas in it. According to Artemov and Kuznets (2006), for each valid knowledge assertion $t:F$, it is shown that there is a Hilbert-style derivation, making a linear number of steps ($3|t|+2$ steps being enough with $|t|$ taken as the number of symbols in t). Even when bringing the two traditions together under a combined framework, as in Artemov and Kuznets (2009) where a logic S4LP which combines **S4** and LP is devised, the distinction between explicit and implicit knowledge regarding logical omniscience is still sharp. More specifically, S4LP allows for a more realistic portrayal of knowledge, as both implicit and explicit versions of it are present. Its language includes both the usual epistemic assertions $K\phi$ and justification assertions $t:\phi$. The former denotes the standard implicit kind of knowledge while the latter, as we have seen, additionally requires feasible witnesses for the knowledge claim and it is precisely because of this that we are able to explain and control logical omniscience. The axiomatization of this new logic is given by the combination of **S4** and LP, supplemented with:

$$t:\phi \rightarrow K\phi \qquad \textit{(connection principle)}$$

It is shown that, indeed, implicit knowledge in S4LP gives rise to logical omniscience, whereas explicit knowledge does not.

Overall, Justification Logic's core idea to offer reasons on *why* something is known, provides natural ways to get closer to real agents' mental processes and avoid logical omniscience. In addition, this attempt succeeds in bridging rationality and non-omniscience. By focusing on the boundedness of resources, as exemplified in the quantitative use of justification terms, it ensures that the solution does not in any sense imply that agents are incapable of making simple inferences. This is also why Justification Logic is aligned with our intuitions on how reasoning actually progresses. The mere introduction of justifications resettles the notion of knowledge and assists the distinction between implicit and explicit knowledge. The tests LOT and SLOT complete the picture and suggest a way out while avoiding the construction of trivialized logics. However, the way the tests are devised hints at a flavour of knowability in that it presupposes that agents follow the *right* reasoning trajectory. As a result, it does not deliver a full account for those situations whereby an agent simply follows the wrong track, maintains inconsistent beliefs etc. This can be seen as lying beyond the scope of this attempt if we view it as a resource-sensitive proposal that provides us with an intuitive way to express (quantitatively) the resource-boundedness without delving into determining a strict bound, i.e. a bound more precise that merely polynomial in the size of the assertion. In addition, the response to the problem is parameterized by the measure of the size of the formulas/proofs and the proof system itself. For example, if we assume a Hilbert-style proof system, the size of the proofs can be the number of steps used in the derivation or the number of logical symbols appearing in the derivation. Likewise, the size of the formula can be thought of as the number of logical symbols in the formulas or its word length. The margin of arbitrariness involved in this selection deems the status of the tests, and hence of the evaluation of epistemic systems, debatable. The authors themselves acknowledge that one should be careful in this choice as omniscience-free systems can be artificially engineered to merely pass the tests and simply throw out standard knowledge assertions (Artemov and Kuznets (2006)). However, it seems more credible, from the explanatory point of view, to choose the proof system based on some empirical indication on what the agents we are interested in usually adopt. For example, experimental evidence from psychology of reasoning (Dieussaert et al. (2000), Stenning and van Lambalgen (2008)) suggests that certain rules, e.g. Modus Ponens, are preferred from others, e.g. Modus Tollens, by real-life human agents. Prioritizing such remarks might allow us to present more plausible epistemic systems, instead of following a recipe that produces systems that might be, on one hand, omniscience-free but on the other, loosely motivated.

3.2.4 Logics of Justified Belief and Knowledge

Logics of justified belief and knowledge and their dynamic extensions build on Justification Logic, Dynamic Epistemic Logic and Belief Revision in order to capture the difference between implicit and explicit attitudes, while incorporating the effect of dynamical changes. As a result, the idea of introducing justification terms to better approximate real-world knowledge and belief is extended to allow for explicit analogues of many notions from the literature on Dynamic Epistemic Logic (such as defeasible knowledge) and for justification-sensitive dynamics. The following definitions set the background of this attempt:

Definition 3.2.13 (Language $\mathcal{L}_{JB}$, Baltag et al. (2012)). The language $\mathcal{L}_{JB}$ comprises both a set of propositional formulas $\mathcal{F}$ and a set of evidence terms $\mathcal{T}$. Given a set Φ of propositional atoms:

- $\phi \quad ::= \quad \bot \quad | \quad p \quad | \quad \neg\phi \quad | \quad \phi\wedge\phi \quad | \quad Et \quad | \quad t \gg \phi \quad | \quad \Box\phi \quad | \quad K\phi \quad | \quad Y\phi$, with $p \in \Phi$
- $t \quad ::= \quad c_\phi \quad | \quad t \cdot t \quad | \quad t + t$

Spelling out the meaning of these constructions:

- Et means that the evidence t is available to the agent but it is not necessarily accepted by her.
- $t \gg \phi$ means that t is admissible evidence for ϕ: if accepted, this evidence supports ϕ. To avoid confusion, we should clarify that, apart from a symbol of the language, $\gg$ will also be used to denote the smallest binary relation on $\mathcal{T} \times \mathcal{F}$ such that (a) $c_\phi \gg \phi$, (b) if $t \gg (\psi \Rightarrow \phi)$ and $s \gg \psi$, then $(t \cdot s) \gg \phi$, and (c) if $t \gg \phi$ or $s \gg \phi$, then $(t + s) \gg \phi$.
- $\Box$ means that the agent implicitly *defeasibly* knows ϕ. As explained in Chapter 2, defasible knowledge stands for a justified true belief stable under the introduction of any *true* information.
- $K\phi$ means that the agent implicitly *infallibly* knows ϕ. In this case and in line with the the previous chapter, the incoming information is not necessarily true. It might include false evidence, misleading testimony etc. Yet infallible knowledge survives even in the face of this misinformation.
- $Y\phi$ is a temporal operator, meaning that "yesterday" (before the last epistemic action) ϕ was true.
- c_ϕ is an evidential certificate, a "canonical" piece of evidence in support of a sentence ϕ.
- $s \cdot t$ is a compound piece evidence, where the operation $\cdot$ combines two pieces of evidence s and t using Modus Ponens.
- $t + s$ is a body of evidence, where the operation $+$ combines two pieces of evidence by aggregating all the evidence provided by t and s without performing logical inferences.

In addition, consider $sub(\phi)$, the set of subformulas of a formula ϕ: for the boolean cases, the set is obtained in the standard way of defining subformulas; then, $sub(Et) := \{Et\}$, $sub(t \gg \theta) := \{t \gg \theta\}$, $sub(\Box\theta) := \{\Box\theta\} \cup sub(\theta)$, $sub(K\theta) := \{K\theta\} \cup sub(\theta)$, and $sub(Y\theta) := \{Y\theta\} \cup sub(\theta)$. The abbreviations below will also be useful:

Definition 3.2.14 (Abbreviations, Baltag et al. (2012)).

- $con_t := \bigwedge\{\theta \mid t \gg \theta\}$, i.e. the conjunction of all formulas for which t is admissible evidence.

- $B\phi := \Diamond\Box\phi$, which says that the agent implicitly believes ϕ.
- $A(t) := \bigwedge_{c_\phi \in sub(t)} B\phi$, which says that the agent implicitly accepts evidence t.
- $G(t) := \bigwedge_{c_\phi \in sub(t)} \Box\phi$, which says that t is good implicit evidence.
- $I(t) := \bigwedge_{c_\phi \in sub(t)} K\phi$, which says that t is infallible implicit evidence.
- $t : \phi := A(t) \wedge t \gg \phi$, which says that t is implicit evidence for belief of ϕ.
- $B^e\phi := B\phi \wedge Ec_\phi$, which says that the agent explicitly believes ϕ.
- $\Box^e\phi := \Box\phi \wedge Ec_\phi$, which says that the agent explicitly defeasibly knows ϕ.
- $K^e\phi := K\phi \wedge Ec_\phi$, which says that the agent explicitly infallibly knows ϕ.
- $t :^e \phi := t : \phi \wedge Et$, which says that t is explicit evidence for belief in ϕ.

Next, we present the models of this approach; notice that they bring together plausibility models and the notion of justification.

Definition 3.2.15 (Model for justified attitudes, Baltag et al. (2012)). A *model for justified attitudes* is a structure $M = \langle W, [[\cdot]], \sim, \geq, \rightsquigarrow, E\rangle$, where:

- W is a non-empty set of possible worlds.
- $[[\cdot]]$ is a valuation map from Φ to $\mathcal{P}(W)$.
- $\geq, \sim, \rightsquigarrow$ are binary relations on W, representing relative plausibility, epistemic indistinguishability, and immediate temporal precedence respectively. Certain conditions are imposed on these relations.
  - $\sim$ is taken to be an equivalence relation and $\geq$ a preorder.
  - Indefeasibility: $w \geq u \rightarrow w \sim u$.
  - Local Connectedness: $w \sim u \rightarrow (w \geq u \vee u \geq w)$
  - Propositional Perfect Recall: $(w \rightsquigarrow u \sim u') \rightarrow \exists w'(w \sim w' \rightsquigarrow u')$.
  - Uniqueness of Past: $(w' \rightsquigarrow w \wedge w'' \rightsquigarrow w) \rightarrow w' = w''$.
  - Persistence of Facts: $w \rightsquigarrow w' \rightarrow (w \in [[p]] \leftrightarrow w' \in [[p]])$.
- E is an evidence map from W to $\mathcal{P}(\mathcal{T})$. Further conditions are imposed regarding E.
  - Evidential Perfect Recall: $w \rightsquigarrow w' \rightarrow \{t^Y \mid t \in E(w)\} \subseteq E(w')$.
  - (Implicit) Evidential Introspection: $w \sim u \rightarrow E(w) = E(u)$.
  - Subterm Closure: if $t \cdot t' \in E(w)$ or $t + t' \in E(w)$ then $t \in E(w)$ and $t' \in E(w)$, that is the constituent evidential pieces should be available too.
  - Certification of Evidence: if $t \in E(w)$ and $t \gg \phi$ then $c_\phi \in E(w)$, that is all explicit knowledge can be certified.

A pointed model is a pair (M, w) consisting of a model M and a designated world w in M called the "actual world".

If we require well-foundedness of $\geq$ and $\rightsquigarrow$ (i.e. there are no infinite chains of plausibility and temporal precedence), we get the *standard models*, so that truth of belief ($B\phi$) can be captured in terms of "truth in the most plausible worlds".

Definition 3.2.16 (Semantics for justified attitudes, Baltag et al. (2012))**.** Truth at a world w in a model M is inductively defined for formulas $\phi \in \mathcal{F}$, based on the following clauses:

- $M, w \nvDash \bot$
- $M, w \vDash p$ if and only if $w \in [[p]]$
- $M, w \vDash \neg\phi$ if and only if $M, w \nvDash \phi$
- $M, w \vDash \phi \wedge \psi$ if and only if $M, w \vDash \phi$ and $M, w \vDash \psi$
- $M, w \vDash Et$ if and only if $t \in E(w)$
- $M, w \vDash t \gg \phi$ if and only if $t \gg \phi$
- $M, w \vDash \Box\phi$ if and only if $M, u \vDash \phi$ for every $u \leq w$
- $M, w \vDash K\phi$ if and only if $M, u \vDash \phi$ for every $u \sim w$
- $M, w \vDash Y\phi$ if and only if $M, u \vDash \phi$ for every $u \rightsquigarrow w$

We can now proceed with the proof system and the main results on soundness and completeness:

Definition 3.2.17 (Axiomatization of JB, Baltag et al. (2012))**.** The axiomatization of JB is defined in Table 3.1.

Table 3.1: The theory JB

AXIOM SCHEMES
Axioms of Classical Propositional Logic
$Et \rightarrow KEt$ (Knowledge of Available Evidence)
$E(t \cdot s) \rightarrow Et \wedge Es$ $E(t + s) \rightarrow Et \wedge Es$ (Subterm Closure)
$t \gg \phi \wedge Et \rightarrow Ec_\phi$ (Certification of Available Evidence)
$t \gg \phi$ if $t \gg \phi$ $\neg(t \gg \phi)$ if $t \not\gg \phi$ (Admissibility)
S5 axioms for K
S4 axioms for $\Box$
$K\phi \rightarrow \Box\phi$ (Indefeasibility)
$K(\phi \vee \Box\psi) \wedge K(\psi \vee \Box\phi) \rightarrow (K\phi \vee K\psi)$ (Local Connectedness)
$Y(\phi \rightarrow \psi) \rightarrow (Y\phi \rightarrow Y\psi)$ (Normality of Y)
$YK\phi \rightarrow KY\phi$ (Propositional Perfect Recall)
$YEt \wedge \neg Y\bot \rightarrow Et^Y$ (Evidential Perfect Recall)
$\neg Y\phi \rightarrow Y\neg\phi$ (Uniqueness of Past)
$Yp \leftrightarrow (\neg Y\bot \rightarrow p)$ (Persistence of Facts)
RULES
From ϕ and $\phi \rightarrow \psi$ infer ψ (Modus Ponens)
From ϕ infer $\Box\phi$ ($\Box$ N)
From ϕ infer $K\phi$ (KN)
From ϕ infer $Y\phi$ (Y N)

Theorem 3.2.2 (Completeness of JB, Baltag et al. (2012))**.**

- JB is sound and strongly complete with respect to the class of all models.
- JB is sound and weakly complete with respect to the class of the standard models.

Till now, the framework is static. However, it is important to consider how explicit belief and knowledge are *practically* attained. During discourse, the interlocutors share information with one another and prompt actions in the manipulation of evidence. Depending on whether the source is considered absolutely reliable or strongly trusted, the receiver updates or upgrades with the given piece of evidence. During reflection, an individual might form or become aware of some piece of evidence e.g. of the instance of an axiom or of her own implicit beliefs and non-beliefs. Furthermore, she can perform Modus Ponens to compose previously collected pieces of evidence and form a new one. On the whole, acquiring evidence terms and thus explicit belief and knowledge is a dynamic process and this is why machinery from DEL is integrated in the JB framework.

In what follows, we describe four types of epistemic actions and present the extension of the basic language to *a language with updates* capable of modelling these dynamical scenarios.

- **Availability of evidence,** $t+$: the evidence term t becomes available, either because the agent becomes aware of the possibility of such evidence (e.g. awareness of an axiom instance or introspection on her own beliefs) or because the agent forms this term. It is useful to underline that availability does not necessarily entail acceptance of evidence. The modal operator corresponding to $t+$ is denoted by $[t+]$ and the formula $[t+]\phi$ is informally read as "after t becomes available, ϕ is true". The precondition for $t+$ is $pre_{t+} := \top$, i.e. the action can always happen and no further conditions need to be imposed. Its *evidence set*, that is the set of all the evidence terms that become available due to the action, is $\mathcal{T}_{t+} := sub(t) \cup \{c_\theta \mid s \gg \theta \text{ for some } s \in sub(t)\}$.
- **Combination of evidence,** $t \otimes s$: given two available evidence terms t and s the action $t \otimes s$ forms a new term $t \cdot s$, corresponding to combining the terms by performing Modus Ponens. The modal operator corresponding to $t \otimes s$ is denoted by $[t \otimes s]$ and the formula $[t \otimes s]\phi$ is informally read as "after t and s are combined through Modus Ponens, ϕ is true". The precondition for $t \otimes s$ is $pre_{t\otimes s} := Et \wedge Es$. i.e. t and s should be already available. Its *evidence set* is $\mathcal{T}_{t\otimes s} := \{t \cdot s\} \cup \{c_\theta \mid t \cdot s \gg \theta\}$.
- **Update with hard evidence,** $t!$: this is the action of updating with some "hard" evidence term t, that is coming from an absolutely infallible source. This notion is analogous to the one of the standard update under DEL, only now the input is an evidence term and not a proposition. We also note that the introduced evidence becomes available and accepted by the agent, but in its past form as it unveils properties of the world as it was *before* the update. The modal operator corresponding to $t!$ is denoted by $[t!]$ and the formula $[t!]\phi$ is informally read as "after updating with hard evidence t, ϕ is true". The precondition for $t!$ is $pre_{t!} := con_t = \bigwedge\{\theta \mid t \gg \theta\}$, i.e. t is indeed "hard" evidence (it supports a true proposition). Its *evidence set* is $\mathcal{T}_{t!} := sub(t) \cup \{c_\theta \mid s \gg \theta \text{ for some } s \in sub(t)\}$.
- **Upgrade with soft evidence,** $t \Uparrow$: this is the action of upgrading with some "soft" evidence t, that is coming from a reliable but not infallible source. This notion is analogous

to the action of radical upgrade in DEL. The new evidence is also strongly accepted but not infallibly known. The modal operator corresponding to $t\Uparrow$ is denoted by $[t\Uparrow]$ and the formula $[t\Uparrow]\phi$ is informally read as "after upgrading with soft evidence t, ϕ is true". The precondition for $t\Uparrow$ is $pre_{t\Uparrow} := \top$. Its *evidence set* is $\mathcal{T}_{t\Uparrow} := sub(t) \cup \{c_\theta \mid s \gg \theta$ for some $s \in sub(t)\}$.

The dynamic account is then captured by the following:

Definition 3.2.18 (Language with updates, Baltag et al. (2012))**.** The *basic language extended with updates* $\mathcal{L}^{act} := (\mathcal{T}^{act}, \mathcal{F}^{act})$ contains the new modal formulas obtained by the operators $[\alpha]$ where $\alpha = \{t+, t\otimes s, t!, t\Uparrow\}$, for every $t, s \in \mathcal{T}$[12].

Definition 3.2.19 (Semantics for updates, Baltag et al. (2012))**.** The satisfaction relation $(M, w) \vDash \phi$ is extended to accommodate the dynamic modalities $[\alpha]\phi$ where $\alpha \in \{t+, t\otimes s, t!, t\Uparrow\}$:

$$x \vDash_M [\alpha]\phi \text{ if and only if } x^\alpha \vDash_{M[\alpha]} \phi$$

with x^α representing the "updated" world and $M[\alpha] := (W^\alpha, [[\cdot]]^\alpha, \sim^\alpha, \geq^\alpha, \rightsquigarrow^\alpha, E^\alpha)$ where

$W^\alpha := W \cup \{w^\alpha \mid w \in [[pre_\alpha]]\}$

$E^\alpha(w) := E(w)$ for $w \in W$

$E^\alpha(w^\alpha) := \{u^Y \mid u \in \mathcal{T}(\alpha) \cup E(w)\}$

$[[p]]^\alpha := [[p]] \cup \{w^\alpha \in W^\alpha \mid w \in [[p]]\}$

$\sim^\alpha := \sim \cup \{(w^\alpha, u^\alpha) \mid w \sim u\}$

$\rightsquigarrow^\alpha := \rightsquigarrow \cup \{(w, w^\alpha) \mid w \in [[pre_\alpha]]\}$

$\geq^\alpha := \geq \cup \{(w^\alpha, u^\alpha) \mid w \geq u\}$ for $\alpha \in \{t+, t\otimes s, t!\}$

$\geq^{t\Uparrow} = \geq \cup \{(w^{t\Uparrow}, u^{t\Uparrow}) \mid (w \notin [[con_t]] \wedge u \in [[con_t]]) \vee (w \notin [[con_t]] \wedge w \geq u) \vee (v \in [[con_t]] \wedge w \geq u)\}$

$\geq^{t\Uparrow} = \geq \cup \{(w^{t\Uparrow}, u^{t\Uparrow}) \mid w \geq u\}$ for $t \notin \mathcal{T}^e$

Finally, there is a dynamic axiomatization as well:

Theorem 3.2.3 (Theory with dynamics, Baltag et al. (2012))**.** DJB, the theory of dynamic justified belief, is composed by Table 3.1 and Table 3.2.

Theorem 3.2.4 (Soundness and completeness of DJB, Baltag et al. (2012))**.** $\vdash \phi$ if and only if $\vDash \phi$ for each $\phi \in \mathcal{F}^{act}$.

A variant of this approach is described in Baltag et al. (2014). The newly introduced notion is that of *actual availability of "conclusive" ("good") evidence*. The main idea is that conclusive evidence is fully reliable, as is the case with its constituent parts. More specifically, the availability of good evidence t, denoted by Et in the language of this system, expresses that (a) the agent has actually observed all basic pieces of evidence used as the building blocks for t, and (b) the agent actually constructed the argument t. The semantics, axiomatization and completeness results are then adapted to account for this new notion. It is worth noticing that the implicit-versus-explicit distinction within this framework is extended to generate implicit and explicit versions of the axioms of the several justification logic(s). For example,

[12] Note that their introduction extends both the set of formulas *and* the set of terms; think of c_ϕ.

Table 3.2: The theory DJB

	AXIOM SCHEMES
Persistence of facts:	$[\alpha]p \leftrightarrow (pre_\alpha \to p)$
Functionality:	$[\alpha]\neg\phi \leftrightarrow (pre_a \to \neg[\alpha]\phi)$
Distributivity of conjunction	$[\alpha](\phi \wedge \psi) \leftrightarrow [\alpha]\phi \wedge [\alpha]\psi$
Evidence dynamics:	$[\alpha]Et^Y$ for $t \in \mathcal{T}(\alpha)$ $[\alpha]Et^Y \leftrightarrow (pre_\alpha \to Et)$ for $t \notin \mathcal{T}(\alpha)$ $[\alpha]Es \leftrightarrow \neg pre_\alpha$ for $s \notin \{t^Y \mid t \in \mathcal{T}\}$
Admissibility dynamics:	$[\alpha](t \gg \phi) \leftrightarrow (pre_\alpha \to t \gg \phi)$
Knowledge dynamics:	$[\alpha]K\phi \leftrightarrow (pre_\alpha \to K[\alpha]\phi)$ $[\alpha]\Box\phi \leftrightarrow (pre_\alpha \to \Box[\alpha]\phi)$ for $\alpha \in \{t+, t \otimes s, t!\}$ $[t\Uparrow]\Box\phi \leftrightarrow \Box(\neg con_t \to [t\Uparrow]\phi) \wedge (con_t \to \Box[t\Uparrow]\phi \wedge K(\neg con_t \to [t\Uparrow]\phi))$
Temporal dynamics;	$[\alpha]Y\phi \leftrightarrow (pre_\alpha \to \phi)$

the Application Axiom $t:(\phi \to \psi) \to (s:\phi \to (t \cdot s):\psi)$ is only valid in its implicit form, since for an explicit counterpart we would additionally require that the combined term $t \cdot s$ is actually available to the agent, i.e. $t:^e (\phi \to \psi) \to (s:^e \phi \wedge E(t \cdot s) \to (t \cdot s):^e \psi)$ is valid whereas $t:^e (\phi \to \psi) \to (s:^e \phi \to (t \cdot s):^e \psi)$ is not.

The contribution of this kind of strategies against the problem of logical omniscience consists in that explicit knowledge(/belief) avoids the closure principles as availability of the relevant pieces of evidence is not always guaranteed, but it rather depends on the value of $E(w)$ at each world. Similar to Justification Logic, the problem persists for implicit knowledge(/belief), which is viewed, according to this framework, as mere *potential* knowledge(/belief) whereas explicit knowledge requires that agents actually go through the trouble of collecting evidence and correctly validating the needed certificates. In addition, the dynamic extension provides the tools to spell out how actions affect the construction of our evidential stack. As a result, it sheds light on our *actual* reasoning processes and their interactive features. It is worth noticing that it does so while still alluding to the results of standard (dynamic) epistemic systems and their explanatory power. Besides, the merge of justification and plausibility facilitated the introduction of other evidence-based notions of knowledge and belief, thereby allowing us to capture a wider range of phenomena.

However, the treatment of the problem of logical omniscience can be reduced to the one pursued via awareness functions; the function E can be thought of as an awareness function, as this determines which formulas are explicitly known/believed. In that sense, the agent is aware of only those formulas that are evidenced by a term in $E(w)$. The difference is that the value of $E(w)$ actually unpacks the explicit justifications for those formulas that the agent knows/believes and their structure is suitable to reflect these properties. Yet, it still inherits the drawbacks of the awareness structures. Logical competence cannot be guaranteed for the same reasons. In addition, Baltag et al. (2014) discusses another kind of "omniscience" emerging within such frameworks, that can be seen as equally alarming as the ones hitherto discussed. Essentially, it extends the effect of closure principles, generating analogous problematic instances but now with regard to the justification terms. Consider a property P that the set of evidence available to the agent $E(w)$ satisfies, such as "$E(w)$ is closed under any correct application of the Modus Ponens operation $s \cdot t$" or "$E(w)$ is not finite". Then, on the same line of reasoning followed for standard omniscience forms, one can doubt whether resource-bounded agents can actually acquire *all* evidence terms predicted by the property P. It is only via suitably modifying the function E that we can prevent such occasions[13], again though

[13]At this point and despite the similarities highlighted above, we should note that under awareness frameworks,

stumbling across the challenge of striking a balance between competence and fallibility, now regarding the management of evidence. Finally, the otherwise fruitful integration of actions in a justification-involving framework does not *directly* influence the escape from logical omniscience. That is, it is not itself responsible for destroying the closure principles, e.g. due to capturing the complexity of reasoning, but it rather suggests ways the range of the function E can change, which can in turn (indirectly) lead to the desired effect.

3.3 Impossible worlds

3.3.1 Elementary approaches

Another popular approach to address the problem suggests expanding the set of worlds with worlds that "look possible and hence must be admissible as epistemic alternatives but which none the less are not logically possible"[14]. Indeed, a world where Fermat's Last Theorem fails to hold is impossible but it might not be identified as such by a limited agent, that lacks, for example, the appropriate mathematical background. As a result, this world still is a doxastic possibility for her. By extending the characterization of belief and knowledge to *all* doxastically/epistemically accessible worlds, both possible and impossible, we obtain a way out of the problem. Indeed, if an agent knows that ϕ, then ϕ is the case at all worlds epistemically accessible to her, possible and impossible. If, in addition, ϕ logically entails ψ, ψ must be the case at the possible worlds yet there can be an impossible world that does not validate it. Due to the new characterization, we conclude that the agent does not know ψ. We can destroy the other closure properties in a similar fashion. In what follows, we examine concrete proposals that exemplify this idea.

Definition 3.3.1 (Rantala interpretation, Berto (2013))**.** A *Rantala interpretation* is a structure $\langle W, N, R, V \rangle$, where

- W is the non-empty set of all worlds.
- N is the subset of (normal) possible worlds. Therefore, $W - N$ is the set of impossible worlds.
- R is the accessibility relation.
- The valuation function V assigns truth values in the standard, recursive way at possible worlds. However, at the impossible worlds of $W - N$, *all* formulas are assigned a truth value by V directly.

As a result of the arbitrariness of V when it comes to worlds in $W - N$, these become completely anarchic and not closed under any non-trivial consequence relation. By allowing such worlds to be epistemically/doxastically accessible, we can cope with any problematic closure principle. For example, an agent might not know all propositional tautologies because a world where a tautology fails is epistemically accessible for her.

The obvious shortcoming of this initial approach to utilize impossible worlds in epistemic-doxastic settings is that logical competence is not preserved. On the contrary, the setting is trivialized and agents appear as subjects of full ignorance. That is, they fail even in conducting

this form of omniscience could not have been expressed, in the absence of appropriate machinery.

[14] Hintikka (1975).

trivial inferences as the ill-behaved assignment allows for worlds verifying ϕ but not ψ, even when ψ is a trivial consequence of ϕ. In addition, there is no independent motivation behind tailoring the valuation function in this way. It only seems relevant as a means to tackle logical omniscience but it does not reveal any interesting property nor an intuitive explanation on failures of reasoning.

Of course, one might suggest that partially closing the worlds, via imposing certain restrictions on the ill behaviour of the assignment could alleviate these worries. Suppose, for example, that we require that each world that verifies ϕ also verifies at least some "easy" logical consequences of ϕ. Then it can be shown (Bjerring (2013), Jago (2014), Rasmussen (2015)) that this closure under "easy" logical consequence *collapses* into closure under full logical consequence, since any inference can be spelled out as a chain of trivial inferences.

Another elementary approach to the problem consists in interpreting the "impossibility" of impossible worlds in terms of non-classicality. Impossible worlds are then closed under logical consequence in some (weaker) *non-classical* logic L. While this approach manages to avoid logical omniscience with respect to classical logic – as the weaker logic L will obviously prevent the agents from drawing all classical consequences – it fails with respect to the non-classical logic. Taking, for instance, a paraconsistent logic, agents still end up knowing all paraconsistent consequences of what they know. However, there is no reason to assume that this conclusion is less problematic. The same arguments making the treatment of (classical) logical omniscience a worthwhile task, can be replicated to show why a new cluster of non-classical closure results is equally problematic. Thus, this approach is also inadequate.

It is therefore clear that the mere introduction of impossible worlds cannot provide a full alternative to realistic logical modelling of knowledge and belief, mainly because the problem is retained in another form and logical competence is completely overlooked.

3.3.2 Jago

In Jago (2014) we find another system in which worlds not closed under logical consequence can still be accessible. However, the set of worlds is structured in a way that distinguishes subtle and obvious impossibilities. The underlying motivation is that there are *blatantly* inconsistent worlds (containing obvious contradictions that agents easily unveil with limited reasoning) and *subtly* inconsistent worlds (containing hidden contradictions, that agents with limited resources cannot spot). This latter kind cannot be easily ruled-out by a priori reasoning. So such worlds might be epistemically accessible despite their impossibility and agents' rationality.

To begin with, the worlds of this framework are not necessarily complete and consistent. For instance, an incomplete world can represent neither that A nor that $\neg A$ while an inconsistent world can represent both. That is, A and $\neg A$ are not the only options there are, nor are they mutually exclusive, contrary to the mainstream view of logically possible worlds. This is why it is useful to consider *double worlds*, pairs of sets of sentences, i.e. $w = \langle w^+, w^- \rangle$ with w^+, w^- representing what holds and what fails in w respectively.

Then, the domain of worlds, called *epistemic space*, is structured in such a way that epistemic possibility is captured in terms of the normative relations *among worlds* and not in terms of intra-world normative relations. By making the epistemic space, and not worlds themselves, admitting the normative principles, respect to the logical deductive links is reflected on the way worlds are linked with one another.

We hereafter discuss step-by-step the construction and the notions involved in the epistemic space.

- **The proof-rules**
  The set of worlds is structured via proof-rules corresponding to the logical connectives. The steps taken in the construction of a proof are translated into connections between worlds. Suppose, for example, that a world w represents A and B but not $A \wedge B$ while another world w' does represent $A \wedge B$. Since $A \wedge B$ is obtainable through Conjunction Introduction, an arrow is drawn from w to w'. In this way, the normative relations are indeed reflected on the structure of the space. Of course, a strict method presupposes that there is a *specific* rule system of use (in our case, a version of Gentzen's intuitionistic calculus). According to this, there are sequents of the form $\Gamma \vDash \Delta$, meaning that the conjunction of all sentences in Γ entails the disjunction of all sentences in Δ. The proof-rules are the ones involving the logical connectives and the structural rule of *Identity*, as presented in Buss (1998). Proofs are practically constructed in tree-form, beginning with the sequent to be proved, as a root. Then we proceed backwards applying the rules, aiming for an instance of Identity. Once this point is reached, the sequent is deemed valid due to the soundness of the rules.

  In addition, and to anticipate the discussion that follows, structure in terms of proof-rules is suggestive of a criterion to scale worlds in terms of their inconsistency. The more steps needed to unveil a contradiction, the subtlier it is; therefore the corresponding world becomes an epistemic possibility. However, what constitutes enough steps and thus what draws the line between epistemic (im)possibilities is vague.

- **From proofs to world-graphs**
  Continuing our description, if $\Gamma \vDash \Delta$ is valid, a world in which all sentences in Γ are true and all sentences in Δ are false is an inconsistent world. Double-worlds, which can precisely represent what is the case and what is not the case, can be structured immediately due to the proof-rules and this observation. In other words, by considering a double-world $\langle \Gamma, \Delta \rangle$, the corresponding sequent is valid if and only if the world is inconsistent. This correspondence allows for sequent-calculus trees to be transformed into world-graphs, such that the inconsistent world w is, as its sequent, placed at the bottom of a"world-proof" and then leaves are similarly built, based on their sequent counterparts. World-inconsistency is therefore captured in terms of the structure induced by proof-rules.

- **The rank**
  The next goal is to determine the degree of inconsistency. Intuitively, this is illustrated in the size of the world-proof: small world-proofs indicate that the world's contradiction can be easily retrieved, so the world is blatantly inconsistent. On the other hand, large world-proofs indicate the difficulty to uncover the contradiction, so the world is subtly inconsistent. More concretely, we define the size of a proof as the number of its non-leaf nodes. Then, we collect the proofs corresponding to a world w in a set G_w. If there are such proofs, i.e. $G_w \neq \varnothing$, then the rank is given by $r(w) = min\{|G| \mid G \in G_w\}$, otherwise $r(w) = \omega$. Next, we order worlds according to $r(w)$: $w \leq w'$ if and only if $r(w) \leq r(w')$. In this way, we obtain a full account for all worlds: the most blatantly inconsistent worlds are the ones with small proof size. Worlds that never appear in the root of a world-proof are the possible worlds. The rest are inconsistent in some way or another, but not of the

same degree since they might have different ranks, therefore different levels in the ease one uncovers their contradiction. According to the ordering, if $w \leq w'$ the contradictions of w' are at least as difficult to uncover as w's ones. So if one accepts w as epistemically possible, one should also accept w'. A converse statement holds for $w' \leq w$ when w is not epistemically possible. In this way, the set of epistemically possible worlds only contains those impossible worlds for which no obvious a priori impossibilities can be discovered. Of course, where the line is drawn in what is considered small/large rank is not sharply determined.

- **The epistemic space**
  We can now strictly define the *epistemic space* E:

  Definition 3.3.2 (Epistemic space)**.** Fixing an object language L as a set of logically primitive sentences P closed under $\neg, \wedge, \vee$ and $\rightarrow$, an *epistemic space* E is a tuple $\langle W, V^+, V^-, r\rangle$ where W is the set of all worlds, the functions V^+, V^- are labelling functions of type $W \rightarrow \mathcal{P}(L)$, assigning a set of sentences to each world (practically determining what is true and false at each world, respectively) and r is a ranking function of type $W \rightarrow N \cup \{\omega\}$, as described above.

  We also define the rank of E as $min\{r(w) \mid w \in W\}$. This is an indicator of how trivial the most trivially impossible world is. It is useful as it also conveys a measure of epistemic possibility for the worlds, i.e. a relatively large space-rank indicates that no world is trivially impossible thus all worlds can be considered epistemically possible. In these cases, we get a *genuine epistemic space*. In general, for different integers n with $r(w) > n$ – n encoding the different intuitions on what constitutes a "large" rank – we say that we obtain different *sharpenings* of the epistemic space.

  The epistemic space, as constructed above, and the fine-grained worlds that comprise it, allow for an account of epistemic content [15] capable of overcoming the shortcomings of the standard approach. In particular, the epistemic content of 'A' is identified with the pair $\langle |A|^+, |A|^- \rangle$ where the first component includes the epistemically possible worlds according to which 'A' is true and the second those according to which 'A' is false. Of course, since epistemic possibility is, as we have seen, indeterminate, the indeterminacy is also inherited by content membership.

  In order to formally incorporate this view of content under his framework, Jago defines the *pointed space* and the *epistemic n-entailments.*

  Definition 3.3.3 (Pointed space, Jago (2014))**.** A *pointed space* is a pair $\langle E, w\rangle$ where $w \in W$ in E. Its rank is the rank of E. We will say that $E, w \vDash A$ if and only if $A \in V^+ w$ in E and $E, w \dashv A$ if and only if $A \in V^- w$ in E. For a pointed space E, $E \vDash \Gamma$ if and only if $E \vDash A$ for each 'A' $\in \Gamma$ and $E \dashv \Gamma$ if and only if $E \dashv A$ for at least one 'A' $\in \Gamma$.

  Definition 3.3.4 (Epistemic n-entailment, Jago (2014))**.** For any integer $n \in N \cup \{\omega\}$, a set of premises Γ *epistemically n-entails* 'A' , $\Gamma \vDash^e_n A$ if and only if for all pointed spaces E of rank $r > n$, $E \vDash \Gamma$ only if $E \not\dashv A$.

  For sufficiently large n, the epistemic n-entailments include all the inferences we would like to count as trivial. Epistemic n-entailments, with n ranging in some admissible values, provide sharpenings of trivial inferences: an inference from Γ to A is determinately

[15] That is, notions of content related to knowledge, belief, information etc.

trivial if and only if $\Gamma \vDash_n^e A$ for all n in the range, determinately non-trivial if and only if $\Gamma \nvDash_n^e A$ and it is indeterminate if it is trivial, otherwise.

- **Agents' epistemic states**
  Next we proceed with the modelling of an agent's epistemic state, using an epistemic accessibility relation R. Yet since it is indeterminate which worlds are epistemically possible, it might be indeterminate whether two worlds w and v are connected via R_i and accordingly whether the agent i knows a statement becomes in turn indeterminate. In other words, if we consider the set of worlds R_i-accessible from a fixed world w, i.e. $f_i w = \{u / wR_i u\}$ (called *projection function* of agent i), then membership in this set is indeterminate. We will say that the agent i knows that A, according to w, just in case $f_i w \subseteq |A^+|$.

  Epistemic Models
  Now, the goal is to build epistemic models, augmenting the spaces with epistemic accessibility relations. Then, roughly speaking, truth in a model will be truth in a particular sharpening of the epistemic space and *determinate* truth will require truth in *all* sharpenings. Before we put this account forward, we need some preliminary tools:

  - The language is extended with a *determinacy operator* $\triangle$ such that $\triangle A$ abbreviates "it is determinate that A is the case" and accordingly an *indeterminacy operator* $\triangledown$ with $\triangledown A$ standing for $\neg \triangle A \wedge \neg \triangle \neg A$.
  - Instead of the standard epistemic accessibility relations R_i, $f_i w$ will be used, denoting the set of worlds epistemically accessible from world w for agent i.

Definition 3.3.5 (Epistemic model, Jago (2014))**.** An *epistemic model* is a tuple $M = \langle W^P, W^I, V^+, V^-, r, f_1, \dots, f_k \rangle$, where W^P and W^I are sets of worlds (possible/impossible), V^+, V^-, r are as in Definition 3.3.2 and f_i the projection function for agent i. Additionally: $W^\cup := W^P \cup W^I$ and the rank of M is $min\{r(w) \mid w \in W^\cup\}$.

Definition 3.3.6 ('A'-variant, Jago (2014))**.** Suppose f is a projection function in an epistemic model M and 'A' $\in L$ a sentence. Then the 'A'-variant of f, f^A is defined as :

$$f_i^A w = \begin{cases} (f_i w \cap \{w / \text{`}A\text{'} \in V^+ w\}) \cup (f_i w \cap W^P) & \text{if } f_i w \subseteq \{w / \text{`}A\text{'} \notin V^- w\} \\ f_i w & \text{otherwise} \end{cases}$$

Let $f_i^L = \{f_i\} \cup \{f_i^A / \text{`}A\text{'} \in L\}$.

The following notion suggests formal means to capture the vagueness for epistemic accessibility.

Definition 3.3.7 (Sharpening, Jago (2014))**.** For an epistemic model M, let $a_M = \{\langle g_1, \dots, g_k \rangle \mid g_i \in f_i^L, i \leq k\}$. Each sequence $g \in a_M$ is a sharpening of epistemic accessibility for that model.

We can now give the truth clauses. Based on the above, $\langle f_1, \dots, f_k \rangle$ provides truth *simpliciter* whereas for what is *determinately* true, we demand truth in *all* alternatives in a_M.

Definition 3.3.8 (Relative truth/falsity, Jago (2014))**.** Let epistemic model M and alternative sequence $g \in a_M$ as above. Then we define g-relative truth and g-relative falsity in M, $\vDash_g$ and $\dashv_g$ respectively, as follows: For $w \in W^P$:

- $w \vDash_g p$ if and only if $p \in V^+ w$
- $w \vDash_g \neg A$ if and only if $w \nvDash_g A$
- $w \vDash_g A \wedge B$ if and only if $w \vDash_g A$ and $w \vDash_g B$
- $w \vDash_g A \vee B$ if and only if $w \vDash_g A$ or $w \vDash_g B$
- $w \vDash_g A \rightarrow B$ if and only if $w \nvDash_g A$ or $w \vDash_g B$
- $w \vDash_g K_i A$ if and only if $u \vDash_g A$ for all $u \in g^i w$
- $w \vDash_g \triangle A$ if and only if $w \vDash_h A$ for all $h \in a_M$
- $w \dashv_g A$ if and only if $w \nvDash A$

For $w \in W^I$:

- $w \vDash_g A$ if and only if $A \in V^+ w$
- $w \dashv_g A$ if and only if $A \in V^- w$

Definition 3.3.9 (n-entailment, Jago (2014))**.** A pointed model is a pair $M' = \langle M, w \rangle$ where M is as above and $w \in W^P$ in M. We define truth relative to M' as : $M' \vDash A$ if and only if $M, w \vDash_{\langle f_1 \ldots f_n \rangle} A$ where $f_1 \ldots f_n$ are the projection functions in M. For sets of sentences we say that: $M' \vDash \Gamma$ if and only if $M' \vDash A$ for all 'A'$\in \Gamma$. For any $n \in N \cup \{\omega\}$, logical n-entailment is then defined as: $\Gamma \vDash_n A$ if and only if, for every pointed model M of rank at least n, $M \vDash \Gamma$ only if $M \vDash A$.

Based on the above, Jago proposes his solution of what he calls the *problem of rational knowledge*. The framework allows for *epistemic blindspots*: since knowledge is not deductively closed, an agent might fail to know a trivial consequence of what she knows. But, as noted by Bjerring in Bjerring (2013), allowing for highly incomplete worlds delivers not only non-omniscient but also utterly irrational agents. Jago defends the existence of epistemic blindspots yet argues that they cannot be rationally precisified in one's reasoning. The philosophical discussion on vagueness and the sorites paradox might illuminate this insight: one cannot explicitly pinpoint the epistemic blindspot but there must surely be one, based on our conclusion. Such phenomena of unassertibility at the borderline support this stance on blindspots: they exist but they cannot be determinately asserted. If 'A' is a trivial truth, either agent i knows that A or it is indeterminate whether she knows. With A being trivial, it is never rational to say that determinately i does not know that A. Therefore, if A follows trivially from what i determinately knows then it is never determinate that i does not know that A. In what follows and based on the aforementioned framework, we obtain formal results to back this portrayal of rational but imperfect agents.

Theorem 3.3.1. (Main result, Jago (2014)) For any $n \in N \cup \{\omega\}$, if $\Gamma \vDash_n^e A$ then $\{\triangle K_i B \mid$ 'B' $\in \Gamma\} \vDash_n \neg \triangle \neg K_i A$.

Thus, no matter the sharpening, if Γ trivially infers A, then determinate knowledge of Γ entails that the agent i does not determinately lack knowledge that A.

As a result:

Corollary 3.3.1.1. For any $n \in N \cup \{\omega\}$, if $\Gamma \vDash_n^e A$ then $\{ `\triangle K_i B' \mid `B' \in \Gamma\} \cup \{`\neg K_i A'\} \vDash_n \triangledown K_i A$. Assuming $n \geq 2$, we get $\vDash_n \neg \triangle \neg K_i(A \vee \neg A)$ and $\neg K_i(A \vee \neg A) \vDash_n \triangledown K_i(A \vee \neg A)$.

Therefore, if an agent i does not know some trivial consequence 'A' of what she knows, then it is indeterminate whether she knows that A. This is in line with our original goal: a non-omniscient agent has epistemic blindspots; but we can never rationally pinpoint them. Indeed, according to the presented approach, if the agent is in an epistemic blindspot, it is indeterminate whether she is in that blindspot. This reflects how epistemic blindspots are analogous to the counterexamples of tolerance principles for vague predicates. Just as we obtain the failure of tolerance principles in instances of the sorites paradox, we obtain epistemic blindspots for agents' knowledge. But just as we cannot pinpoint sharp cut-offs in the sorites paradox, we cannot pinpoint where the blindspot lies. Overall, it is showed that the agents face some kind of failure and are indeed non-omniscient; but in order to respect the logical principles, as we would expect from rational agents, it is also showed that concrete counterexamples to them are disallowed.

Jago's take on the problem indeed balances between the two extremes. It also provides a plausible way of structuring the epistemic space and (carefully) discussing epistemic accessibility on these grounds. However, extensive criticisms against this attempt are suggested in Rasmussen and Bjerring (2015), mostly focusing on the notion of *Indeterminacy*, as employed by Jago. To begin with, the notion lacks motivation: there are no independent reasons, i.e. not related to avoiding logical incompetence, to accept the indeterminacy of blindspots. What Jago suggests is a structural similarity between the problem of rational knowledge and the sorites paradox. The authors of Rasmussen and Bjerring (2015) argue that structural similarity *presupposes* rather than *motivates* Jago's indeterminacy. What is more, there are potential counterexamples: consider a non-omniscient agent who knows just the following propositions: $p_1, p_1 \rightarrow p_2, \ldots, p_{k-1} \rightarrow p_k$. Consider that the agent is logically competent in the minimal sense of applying MP just once. Given the collapse result and that p_2 is the only trivial consequence of the agent's knowledge, it follows that the agent fails to know p_2. But then we end up with a logically competent agent who suffers from a *determinate* blindspot, that is, a rather unnatural result. Finally, this proposal lacks explanatory power: due to logical competence, the agent from the previous example should be able to realize that p_2 follows from p_1. But this, in combination with the fact that she does not know p_2, might seem strange. According to the current strategy, the agent's lack of knowledge in p_2 should be indeterminate. Even if it were, there is no explanation on why the agent assents to p_2 while not knowing it. Generalizing this point, there is no independently plausible story – i.e. other than the existence of a blindspot – on why we are not justified in attributing at least one piece of knowledge from the trivial consequences of an agent's knowledge to her. To sum up, the shortcomings of adopting indeterminacy as a means to tackle the issue might overshadow the explanatory potential of Jago's attempt.

3.3.3 Rasmussen & Bjerring

Another approach is suggested in Rasmussen and Bjerring (2015). It builds a *dynamic* doxastic impossible-worlds framework. Again, the diagnosis of the problem is as in Rasmussen (2015), and the goal is to reach a balance between non-omniscience and logical competence. Unlike Section 3.1.2 though, this response to the problem is not syntactic but it is rather based in the (dynamic) semantics provided through the impossible-worlds framework.

To begin with, and in order to suggest a solid criterion on what competence amounts to, a *behavioural test for logical competence* is established: for any p and q, with q a trivial consequence of p, if the agent believes p, then when asked whether q, she responds "yes" immediately. Logical competence seen as successful performance in the test, is attributed to the fact that the agent performs trivial chains of logical reasoning –namely, the trivial inference from p to q. It is not because a belief in q was in her belief state *prior* to being asked about it; in which case we would again face a collapse result as in Section 3.3.1. It is a *reasoning process* that updates her belief state with q, and this is why a *dynamic* impossible-worlds framework is needed. Next, in order to precisify the intended meaning of trivial reasoning, given a set $\mathcal{R}$ of inference rules, we characterize it in a step-wise manner:

Definition 3.3.10 (Trivial logical reasoning, Rasmussen and Bjerring (2015))**.** A chain of logical reasoning is trivial if and only if it involves at most n steps of logical reasoning using the rules in $\mathcal{R}$.

Next, the notation $\Gamma \vdash^n_{\mathcal{R}} \Gamma'$ reads: Γ proves Γ' within n steps of logical reasoning using the rules in $\mathcal{R}$. The varying value on n is such to reflect that the available computational resources determine what can count as "trivial". In addition, the following condition is imposed to express that addition of premises leaves the inference intact.

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

We continue with the main elements of this approach:

Definition 3.3.11 (Language $\mathcal{L}_R$, Rasmussen and Bjerring (2015))**.** The language $\mathcal{L}_R$ is defined inductively from a set of atomic sentences Φ, the doxastic operator B, and a countably infinite set of dynamic operators $\langle n \rangle$ and $[n]$:

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi \wedge \phi \quad | \quad B\phi \quad | \quad \langle n \rangle \phi \quad | \quad [n]\phi$$

with $p \in \Phi$, $n = 0, 1, 2, \ldots$.

The operators in $\mathcal{L}_R$ read as follows:

- $B\phi$: The agent believes ϕ.
- $\langle n \rangle \phi$: After some n steps of logical reasoning, ϕ is the case.
- $[n]\phi$: After any n steps of logical reasoning, ϕ is the case.

Definition 3.3.12 (Doxastic model, Rasmussen and Bjerring (2015))**.** Let W^P and W^I be non-empty sets of possible and impossible words respectively, and let $W := W^P \cup W^I$. A doxastic model for a single agent is a structure:

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

where $f : W \to \mathcal{P}(W)$ is an accessibility function, assigning to each world in W a set of worlds in W, and $V : W \to \mathcal{P}(\mathcal{L}_R)$ is a function, assigning to each world in W a set of sentences in $\mathcal{L}_R$.

Two useful remarks are necessary, regarding these definitions: (a) the valuation function V assigns the set of *atomic* sentences true at each *possible* world, while assigning the set of *all* sentences, atomic or not, at each *impossible* world, (b) Unlike possible worlds, which are complete and logically consistent, impossible worlds need neither be complete nor subject to any non-trivial closure conditions. However, they are considered *minimally consistent*, i.e. for any world $w \in W^I$ and sentence $\phi \in \mathcal{L}_R$, $\{\phi, \neg\phi\} \notin V(w)$. As a result, an agent cannot believe in explicit contradictions.

Now, the gist of the truth clauses, that eventually leads to the solution of the problem, is that those of $\langle n \rangle B\phi$, i.e. of sentences that encompass the dynamic process that resulted in a belief, should be weaker than of $B\phi$: the latter requires that ϕ is true at all doxastically accessible worlds while the former will merely require that ϕ *follows* from each doxastically accessible world within n steps of logical reasoning. Then and in order to capture the model change induced by reasoning processes, the following auxiliary definitions are introduced:

Definition 3.3.13 (n-radius, Rasmussen and Bjerring (2015))**.** The n-radius w^n of a world $w \in W$ is defined as:

$$w^n = \{w' \mid V(w) \vdash^n_{\mathcal{R}} V(w')\}$$

A member of w^n is called an n-expansion of w.

Then, given the remarks above:

- For $w \in W^P$: $w^n = \{w\}$ since possible worlds are deductively closed.
- For $w \in W^I$, the n-radius of w might contain many different n-expansions of w.

Definition 3.3.14 (Choice function, Rasmussen and Bjerring (2015))**.** Let $\mathcal{C} : \mathcal{P}(\mathcal{P}(W)) \to \mathcal{P}(\mathcal{P}(W))$ be a choice function that takes a set $\mathcal{W} = \{W_1, \cdots, W_n\}$ of sets of worlds as input and returns the set $\mathcal{C}(\mathcal{W})$ of sets of worlds which results from all the ways in which exactly one element can be picked from each $W_i \in \mathcal{W}$. A member of $\mathcal{C}(\mathcal{W})$ is called a choice of $\mathcal{W}$.

Based on these definitions, the authors construct a relation $\sim^n$ between pointed models (M, w) and (M', w'), capturing the transition from an agent's belief state characterized by (M, w) to a belief state characterized by (M', w'), that the agent enters after n-steps of logical reasoning. Therefore, we say that (M', w') is n-accessible from (M, w) just in case the set of doxastically accessible worlds from w in M is replaced in M' by a choice of n-expansions of w's accessible worlds in M. This update of doxastic accessibility is captured via:

Definition 3.3.15 (n-variation, Rasmussen and Bjerring (2015))**.** Let $M = \langle W^P, W^I, f, V \rangle$ be a model. $\mathcal{F}^n$ ($n = 0, 1, 2, \ldots$) is a function from pointed models to sets of accessibility functions defined as:

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

where $c \in \mathcal{C}(\{w'^n \mid w' \in f(w)\})$. A member of $\mathcal{F}^n$ is called an n-variation of f.

In order to capture that an agent's doxastic state changes as a result of performing n steps of logical reasoning, we finally need the following definition:

Definition 3.3.16 (n-accessibility, Rasmussen and Bjerring (2015))**.** Let $M = \langle W^P, W^I, f, V \rangle$ and $M' = \langle W^{P'}, W^{I'}, f', V' \rangle$ be models. $(M, w) \sim^n (M', w')$ if and only if $w' = w$, $W = W'$, $V' = V$, and $f' \in \mathcal{F}^n(M, w)$.

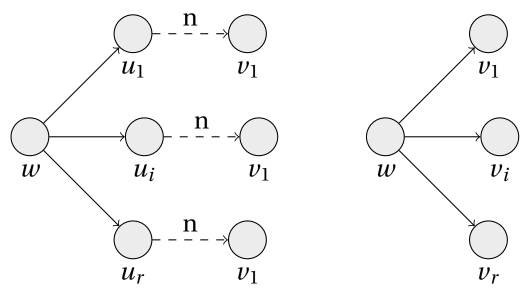


Figure 3.1: First, we have the pointed model (M, w) where $f(w) = \{u_1, \dots, u_r\}$ is the set of w's doxastically accessible worlds, represented by solid arrows. Dashed arrows represent that each v_i is an n-expansion of u_i. Then, a choice of n-expansions of u_i's replaces w's doxastically accessible worlds. Therefore, (M', w) is n-accessible from (M, w).

The semantics is then given as follows:

Definition 3.3.17 (Semantics for Section 3.3.3, Rasmussen and Bjerring (2015))**.** We use $M, w \vDash \phi$ to say that ϕ is true at w in M, and $M, w \Dashv \phi$ to say that ϕ is false at w in model M. For $w \in W^P$:

P1 $M, w \vDash p$ if and only if $p \in V(w)$, where $p \in \Phi$.

P2 $M, w \vDash \neg\phi$ if and only if $M, w \nvDash \phi$.

P3 $M, w \vDash \phi \wedge \psi$ if and only if $M, w \vDash \phi$ and $M, w \vDash \psi$.

P4 $M, w \vDash B\phi$ if and only if $M, w' \vDash \phi$ for all $w' \in f(w)$.

P5 $M, w \vDash \langle n \rangle \phi$ if and only if $M', w' \vDash \phi$ for some $(M', w') : (M, w) \sim^n (M', w')$.

P6 $M, w \vDash [n]\phi$ if and only if $M', w' \vDash \phi$ for all $(M', w') : (M, w) \sim^n (M', w')$.

P7 $M, w \Dashv \phi$ if and only if $M, w \nvDash \phi$.

For $w \in W^I$:

I1 $M, w \vDash \phi$ if and only if $\phi \in V(w)$

I2 $M, w \Dashv \phi$ if and only if $\neg\phi \in V(w)$

Validity is defined with respect to possible worlds only. That is, a formula is valid if and only if it is valid at all possible worlds in all models.

Back to the clauses, $\langle n \rangle B\phi$ is satisfied by (M, w) just in case $B\phi$ is satisfied by some n-accessible pointed model from (M, w). An agent comes to believe ϕ after a trivial chain of logical reasoning whenever there is a transition from the agent's doxastic state through n applications of the rules in $\mathcal{R}$ to a state in which she believes ϕ.

The key-point of these constructions is the next theorem:

Theorem 3.3.2 (Main result of Section 3.3.3, Rasmussen and Bjerring (2015))**.** If $\{\phi_1, \dots, \phi_k\} \vdash^n_{\mathcal{R}} \psi$ and $\langle m_i \rangle B\phi_i$ for $1 \le i \le k$, then $\langle \omega + n \rangle B\psi$, where $\omega = m_1 + \dots + m_k$.

That is, if a conclusion ψ follows within n steps of reasoning from a set of premises $\{\phi_1,\ldots,\phi_k\}$, and the agent can come to believe the i-th premise within m_i steps of reasoning ($1 \leq i \leq k$), then the agent can come to believe ψ within $n + m_1 + \ldots + m_k$ steps of reasoning. Two immediate consequences of this theorem provide the arguments for a balanced solution:

Corollary 3.3.2.1.

- If $\{\phi_1,\ldots,\phi_k\} \vdash^n_{\mathcal{R}} \psi$ and $B\phi_i$, for $1 \leq i \leq k$, then $\langle n \rangle B\psi$.
- If $\vdash^n_{\mathcal{R}} \phi$, then $\models \langle n \rangle B\phi$.

Comparing the two statements with (K): $B\phi \wedge B(\phi \rightarrow \psi) \rightarrow B\psi$. and (N): if $\models \phi$, then $\models B\phi$, we locate the crucial difference. The corollary merely says that the agent can *come to believe* anything that follows within n steps of reasoning from what she already believes, and she can also *come to believe* any logical truth that can be inferred within n steps of reasoning using the rules in $\mathcal{R}$.

To sum up: the desideratum on avoiding logical omniscience is satisfied; by simply allowing impossible worlds to be doxastically accessible and given the quantification on the truth clause of $B\phi$, it is easy to see how the closure principles can be destroyed. Also, the semantics for $\langle n \rangle B\phi$ does not commit us to the implausible claim that agents can trivially come to believe all logical consequences of what they believe, given the restrictive role of n. However, agents are still competent. Suppose $B\phi$ is true at w, for some $w \in W^P$, and consider any ψ that follows from ϕ within n steps of logical reasoning. By Corollary 3.3.2.1, $\langle n \rangle B\psi$ is true at w, and given the characterization of trivial reasoning, it follows that the agent can immediately come to believe any trivial logical consequence of what she believes.

Although this proposal seems to overcome the challenge of bridging non-omniscience and rationality, there are several remarks suggestive of improvements. To begin with, the establishment of the behavioural test and the characterization of trivial reasoning cast doubt on the proposal. It is natural to wonder how they can be independently motivated and on what empirical grounds they are based. For example, the non-specification of the n-steps that compose the definition of trivial reasoning, which the authors attribute to our choice of available computational resources, invites suspicion on how that number could have been sufficiently specified. In the absence of a determinate deciding method, such as an empirical indication, it is natural to wonder what differentiates n and $n+1$ steps in picking out a chain of reasoning as "too big" and thus what renders a belief unattainable. Inevitably, the vague parameter that Jago emphasized emerges here as well, although the authors have argued against it. In addition, there seems to be a discrepancy between the behavioural test and the result on which the authors rely for their way out of the problem. As we have already highlighted, (moderate) logical competence is a normative notion, thus the behavioural test argues on what should be expected from an agent. The main result in Theorem 3.3.2 (and its corollaries), though, does not represent the type of agent who satisfies the normative constraint but only describes the type of agent who *can* do it: it is possible to come to believe the conclusion after n steps of reasoning because this follows from the premises via n applications of inference rules. It therefore seems that along the way, the initial point for which the authors argued diverges from the goal they actually attain. Next, it is not clear how this proposal would deal with agents who believe in explicit contradictions. Of course, even belief in implicit contradictions, in combination with logical competence, is unaccommodated in this setting, given the definitions of the updated model and the behavioural test. Consider the combined beliefs in p, $p \rightarrow q$ and $\neg q$, i.e. an implicit contradiction: since no world can simultaneously verify both q and $\neg q$ and given the agent's belief in $\neg q$, no n-expansion of a doxastically accessible would verify q. As a result, the agent cannot come to believe q and would therefore not pass the behavioural test upon being asked whether q; she is considered logically incompetent, merely due to (indirectly) believing a contradiction. A full description of a rational, albeit fallible, agent should therefore engage in these questions as well.

3.4 Other remarks

We have discussed some prominent approaches to cope with the problem but the list is by no means exhaustive. It is therefore important to justify this selection. To be more specific, we focused on frameworks that target the problem of logical omniscience on its whole, instead of isolating particular special or weaker forms. Moreover, the emphasis is put on the approaches' potential to *model* knowledge and belief rather than to merely *represent* them. While still acknowledging the contribution of representation, it is in fact easier to accommodate real-life agents within such frameworks; however, no benefits can be reaped in the realm of prediction or philosophical investigation of propositional attitudes. Finally, it is reasonable to choose the most developed and fully unfolded variant(s) of a particular perspective to the problem.

In any case, though, it is useful to briefly summarize and comment on some other approaches. Besides, this discussion may be indicative of the remark made above. To begin with, *non-standard structures* are proposed in Fagin et al. (1995a). The gist of this proposal is the independence of truth values between a formula and its negation, that allows for *incoherent worlds* where both ϕ and $\neg\phi$ hold and *incomplete worlds* where neither holds. Although *Closure under Material Implication* fails, the problem persists (or can be retained) in other forms; additionally, logical competence is not secured. The *structures for implicit and explicit beliefs* constitute Levesque's own response to this decoupling (Levesque (1984)) and include two valuations, taking care of truth and falsehood. Criticisms on whether real, rational agents fit this setting (Fagin and Halpern (1987)) as well as on how the problem is unsolved with respect to relevance logic (Vardi (1986)), render other implicit-versus-explicit approaches, that build on this, more attractive. *Local reasoning structures*, presented in Fagin and Halpern (1987), modify the set of accessible worlds, by relativizing it to a *frame of reference*. However, forms of the problem, like *Closure under Valid Implication* and *Closure under Logical Equivalence* are not treated. In Fagin et al. (1995a), we find a discussion of *Montague-Scott semantics* that substitute the standard relational structures with *neighborhood structures* – a set *of sets* of worlds is assigned to each world by a neighborhood function. The modal truth clause is modified accordingly, yet *Closure under Logical Equivalence* eventually persists. Finally, it is worth mentioning Wansing's work in Wansing (1990), whereby it is shown that alternative logics for knowledge and belief boil down to special cases of impossible-worlds frameworks.

Chapter 4

Proposals for real-life agents

This chapter explores alternative ways out of the problem of logical omniscience that – among others – fix shortcomings revealed in Chapter 3. In addition, as we wish to approximate the real nature of reasoning, we propose solutions that explicitly account for the factors affecting it. More specifically, this attempt is grounded on the construction of a semantics for a dynamic epistemic logic of reasoning steps, in the spirit of Section 3.1.2. Two proposals – *rule-based worlds* (RW) and *impossible worlds* (IW) – emerge out of this investigation. These full frameworks will no longer be susceptible to the criticisms of partiality and ad-hocness, that appeared in the critical discussion of the previous chapter. Apart from alleviating such concerns, IW additionally allows us to gain a more concrete unraveling of a reasoning process and to also capture the (indirect) motivation of resource depletion behind, for instance, Section 3.3.3. This is better exemplified by IWe, a quantitative extension of IW, with which we avoid worries about the non-determination of the number of reasoning steps. What is more, these considerations are squared with the requirements of Artemov and Kuznets (2013) and their work on Justification Logic; they too stressed out the importance of having knowledge assertions equipped with indicators for the steps and the effort that resulted in them. The explanatory power of IWe also hints at the potential fix for another problem previously highlighted: reconciling a competent agent and beliefs in implicit contradictions. Next, and inspired by Chapter 2 and the dynamic extension in Section 3.2.4, we furnish IW with actions of external information, to draw an even more realistic picture of reasoning. To that end, we first integrate public announcements into the existing setting (IWPA). These developments are then translated into an impossible-worlds framework with plausibility models (IWp). This leaves room for softer actions such as radical upgrades, that can alter weaker attitudes, such as defeasible knowledge and belief, which are also accommodated in IWp. Finally, we suggest a reduction of models with impossible worlds to "awareness-like" structures, starting off with Rasmussen & Bjerring. It is then shown that the reduction illuminates how a complete logic for such frameworks can be built. Overall, many sub-tasks indicated by the critical survey are hereafter accommodated. In doing so, we bring together ideas from the literature and analyze real reasoning while, of course, avoiding the problem of logical omniscience.

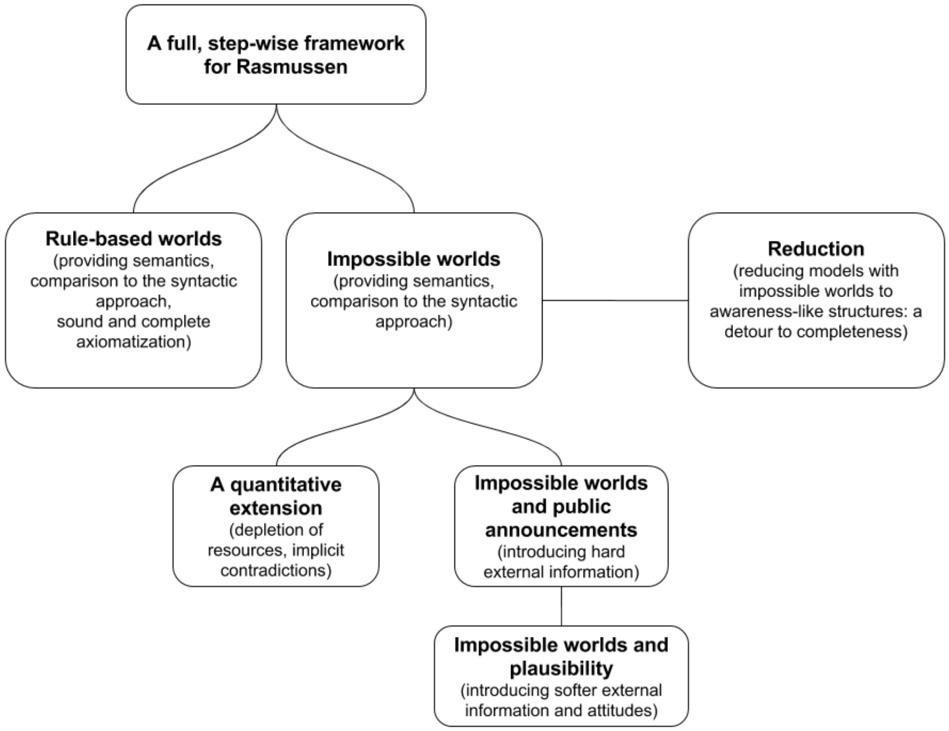


Figure 4.1: A guide to the content of the chapter and the tasks it undertakes.

4.1 A full framework for Rasmussen's dynamic epistemic logic

In this section, we aim at a detailed investigation of an agent's reasoning process, adopting the step-wise fashion, such as the one described in Section 3.1.2[1]. To achieve this, we too focus on (i) the inferences rules applied by the agent and (ii) the chronology of the applications. More specifically, we suggest two refinements of Rasmussen (2015), which we supplement with semantics that consequently allows for a richer study of the axioms proposed to escape the problem of logical omniscience. The first approach is inspired by Velázquez-Quesada (2011) and Jago (2009), while the second suggests rule-specific counterparts of the notions in Rasmussen and Bjerring (2015). In both cases the main exposition is concerned with epistemic systems. However, we also discuss how the constructions can be fine-tuned to yield doxastic settings.

[1] Henceforth, we might simply use "*Rasmussen*" to refer back to this description.

4.1.1 Rule-based worlds (RW)

Basic elements

To begin with, we modify Definition 3.1.2 and construct the language of this first approach, building on the following definitions:

Definition 4.1.1 (Propositional language and inference rules)**.** Given a countable set Φ of propositional atoms, the *propositional language* $\mathcal{L}_P$ is defined inductively as:

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi \wedge \phi$$

Given $\phi_1, \ldots, \phi_n, \psi \in \mathcal{L}_P$, an *inference rule* R_k is a formula of the form $\{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$, read as "whenever every formula in $\{\phi_1, \ldots, \phi_n\}$ is true, ψ is also true".

We then use:

- $pr(R_k)$ and $co(R_k)$ as abbreviations for the set of premises and the conclusion of R_k.
- $\mathcal{R}$ to denote the set of inference rules of $\mathcal{L}_P$.
- $\mathcal{L} := \mathcal{L}_P \cup \mathcal{R}$.

Definition 4.1.2 (Translations for elements of $\mathcal{L}$)**.** For $\phi \in \mathcal{L}_P$, its *translation* is defined as $Tr(\phi) := \phi$. For $R_k \in \mathcal{R}$, its *translation* is defined as $Tr(R_k) := \bigwedge_{\phi \in pr(R_k)} \phi \to co(R_K)$.

We now define the language of this approach:

Definition 4.1.3 (Language $\mathcal{L}_{\text{RW}}$)**.** Given a countable set of propositional atoms Φ, the language $\mathcal{L}_{\text{RW}}$ is defined inductively as follows:

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi \wedge \phi \quad | \quad K\chi \quad | \quad \langle R_k \rangle \phi$$

with $p \in \Phi$, $\chi \in \mathcal{L}$ and $R_k \in \mathcal{R}$.

Apart from the standard propositional language, $\mathcal{L}_{\text{RW}}$ contains knowledge assertions of the form $K\chi$, with χ being either a propositional formula or an inference rule. The former case serves as a means to express knowledge of *facts*, whereas the latter is to say which *rules* the agent knows (and is therefore capable of applying). Moreover, $\mathcal{L}_{\text{RW}}$ comprises formulas of the form $\langle R_k \rangle \phi$, with $\langle R_k \rangle$ seen as an operator for the application of the inference rule. Therefore, such formulas read "after some application of inference rule R_k, ϕ is true"[2]. Dual modalities of the form $[R_k]$ are defined as usual, as is the case with the remaining boolean connectives.

Next, the motivation behind our definition of a semantic model is that the reasoning steps, expressed through applications of inference rules, should be reflected *within* the model. This is why we introduce *rule-based worlds*, that are connected according to the effect of inference rules. These worlds practically work as sets of formulas. However, since reasoning steps affect our own understanding of the world (rather than factual truths), the usual valuation on (im)possible worlds is accompanied by another valuation that essentially yields the set of formulas that the agent knows at each such world. In this sense, the double function of rule-based

[2] At this point, it is worth mentioning that unlike Rasmussen (2015), we do not attach cognitive costs to the language.

worlds differentiates them from the usual manifestations of possible and impossible worlds[3]. Keeping this in mind, the formal definition is:

Definition 4.1.4 (RW-model)**.** An RW-model is a tuple $M = \langle W, T, V_1, V_2 \rangle$ where

- W is a non-empty set of rule-based worlds.
- $T : \mathcal{R} \to \mathcal{P}(W \times W)$ is a function such that a binary relation on W is assigned to each inference rule in $\mathcal{R}$. That is, for $R_i \in \mathcal{R}$, $T(R_i) = T_i \subseteq W \times W$, standing for the transition between worlds induced by the rule R_i.
- $V_1 : W \to \mathcal{P}(\Phi)$ is a labelling function assigning a set of propositional atoms to each world; intuitively those that are true at the world.
- $V_2 : W \to \mathcal{P}(\mathcal{L})$ is a labelling function assigning a set of formulas of $\mathcal{L}$ to each world; intuitively those that the agent knows at the world.

It is now clear that the function V_2 renders the worlds representations of the formulas that the agent knows, while V_1 works as usual. Each inference rule then triggers transitions between worlds. The idea, to anticipate what follows, is to structure the worlds in a way that captures the effect of applying inference rules. By performing an inference using R_k, its conclusion is added to the epistemic state of the agent, therefore a connection is established between the initial world and another world that contains this additional formula. These points allow us to think of the valuation V_2 as an indicator of "explicit knowledge", especially in combination with the factors that supported the introduction of this notion in some of the approaches in Chapter 3. In short, we want V_2 to progress in a step-wise manner across worlds, so that its elements are directly associated with the real knowledge the agent gains by reasoning.

Paving the way for the truth conditions, we emphasize that the language $\mathcal{L}_{\text{RW}}$ can be seen as including (a) the standard propositional part $\mathcal{L}_P$, whose primitive elements are the atoms of Φ, and (b) an epistemic part, whose primitive elements are of the form $K\chi$ with $\chi \in \mathcal{L}$. Introducing two labelling functions V_1, V_2 in the model is on a par with this distinction. The latter determines which primitive epistemic assertions are true (i.e. which propositional formulas/rules the agent knows) at each world, while the former determines which propositional atoms are true at each world. Then, indeed, each world does not only represent what is true at it, but also what the agent knows. Viewing worlds as enumerations of the known formulas naturally adds a syntactic flavour to our treatment of knowledge. Unlike other syntactically-oriented restrictions to escape the omniscience problem, we will see that this setting respects the expectation for rational and resource-bounded agents. Based on these remarks, we proceed with the truth clauses:

Definition 4.1.5 (RW-semantics)**.**

- $M, w \vDash p$ if and only if $p \in V_1(w)$ for $p \in \Phi$.
- $M, w \vDash K\phi$ if and only if $\phi \in V_2(w)$.
- $M, w \vDash \neg\phi$ if and only if $M, w \nvDash \phi$.

[3] Although, conceptually, one might argue that they fit under the spectrum of frameworks with both possible and impossible worlds, by appealing to whether they fulfill closure properties. Still, to avoid terminological misunderstandings, we resort to this new characterization.

- $M, w \vDash \phi \wedge \psi$ if and only if $M, w \vDash \phi$ and $M, w \vDash \psi$.
- $M, w \vDash \langle R_i \rangle \phi$ if and only if there exists some $u \in W$ such that $w T_i u$ and $M, u \vDash \phi$.

Validity is defined as usual.

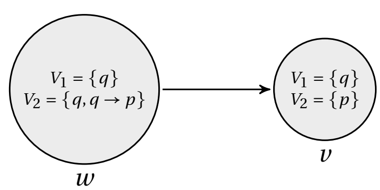
Of course, the two fragments of the language are not exhaustive. For example, $p \wedge Kp$ is well-formed in the language $\mathcal{L}_{\mathrm{RW}}$, yet it is not contained in either of the fragments. In fact, the first conjunct is part of the propositional fragment, and its truth is determined by V_1, while the second conjunct is part of the epistemic fragment, and its truth is determined by V_2. Providing meaningful interpretation to such combinations is indicative of the double function of worlds under the current RW-approach.

For notational convenience, also consider the abbreviation below:

> For given rules of inference $R_1, R_2, \ldots, R_n$, the n-step sequence[4] $\langle R_1 \rangle \ldots \langle R_n \rangle$ is denoted by $\langle \ddagger \rangle^n$. The symbol $\langle \dagger \rangle^m$ is also used to avoid confusion whenever two sequences are involved in one of the subsequent claims. The indices of the inference rules indicate their order of application. Sequences of the dual case work in the same way.

With these first building blocks in hand, it is not hard to see that certain conditions have to be imposed on this initial, general class of RW-models, if they are to capture properties of real reasoners. At this level, the labelling functions are allowed to behave arbitrarily so they do not necessarily reflect the motivation sketched above. Furthermore, formulas such as $\langle \ddagger \rangle^n K\phi \to \phi$ turn out to be invalid[5]. Indeed:

Example 1. Let MP be an instance of Modus Ponens. Suppose $M, w \vDash \langle MP \rangle Kp$, for some RW-model M and world w of the model. Then there is $v \in W$ such that $w T_{MP} v$ with $M, v \vDash Kp$, i.e. $p \in V_2(v)$. Consider the following counterexample:



Obviously, $M, w \nvDash p$ as $p \notin V_1(w)$. As a result, $\langle \ddagger \rangle^n Kp \to p$ is not valid.

[4] It is important to clarify why we deviate from the reading of sequence given in *Rasmussen*. The first, obvious difference is that cognitive costs are missing, which is a direct consequence of our choice of language and model. Secondly, one should keep in mind that the notion of an n-step sequence in $\langle \ddagger \rangle^n \phi$ might stand for a purely *existential* claim (i.e. there is some sequence of n-steps following which, ϕ is the case) *or* a claim involving a *specific*, determined array of rules (i.e. following some application of each rule that the sequence contains, in the intended order, ϕ is the case). It seems that the reading adopted by *Rasmussen* is the former, but it is not clear whether his axioms are best motivated by this. In the absence of semantics, these considerations were left aside. But since we are now augmenting an analogous step-wise approach with a semantical account, an ambiguity would not be constructive at all. Both our intuition on how to break down a reasoning process and the technical ramifications of our semantics call for the specific reading of a sequence. Finally, we only consider non-empty sequences, i.e. of length greater than 0. However, this difference is not substantial, as we will simply comment on the "empty" cases separately.

[5] This is in fact an analogue of Rasmussen's (A1) axiom. The difference, as observed, lies in the absence of cognitive costs in our setting and our diverging notation – and possibly intended meaning – of a sequence of reasoning.

Therefore, certain restrictions should be imposed on the plain semantic model to ensure that the set of worlds is structured in a way that does convey the original idea of transitions as enhancements of the epistemic state. Only then will we be in a position to argue for a full step-wise framework that preserves the attack to the problem, without denying the capabilities of real reasoners.

Definition 4.1.6 (Propositional truths)**.** Let M be an RW-model and $w \in W$ a world of the model. Its set of *propositional truths* is $V_1^*(w) = \{\phi \in \mathcal{L}_P \mid M, w \vDash \phi\}$.

Given the boolean clauses, it is safe to say that V_1^* is entirely determined by V_1.

We are now ready to fix an appropriate class of models, denoted by **M**. For any RW-model M, $M \in \mathbf{M}$ if and only if:

1. For any inference rule $R_i = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$, if $w \in W$ is such that $R_i \in V_2(w)$ and $\phi_1, \ldots, \phi_n \in V_2(w)$, then there exists a world $u \in W$ such that wT_iu and $V_2(u) = V_2(w) \cup \{\psi\}$.
2. For any $w, u \in W$ and inference rule $R_i = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$, if wT_iu then $R_i \in V_2(w)$, $\phi_1, \ldots, \phi_n \in V_2(w)$ and $V_2(u) = V_2(w) \cup \{\psi\}$.
3. For any $w \in W$ and $\phi \in \mathcal{L}$, if $\phi \in V_2(w)$ then $Tr(\phi) \in V_1^*(w)$.
4. For any $w, u \in W$ and inference rule R_i, if wT_iu then $V_1^*(w) = V_1^*(u)$.

The first condition says that if a world represents an epistemic state containing the premises of a known rule R_i, then this world is connected by the corresponding T_i with a world that extends it exactly by the conclusion. In this sense, it captures that transitions correspond to the ways an epistemic state is enriched by applications of inference rules. The second condition says that if w is connected to u via T_i, then it must be the case that u enriches the epistemic state of w in terms of R_i. This is to capture that each arrow drawn is indeed associated with some addition of a conclusion to an epistemic state. The third condition is imposed to guarantee the veridicality of knowledge within a world and the soundness of the known rules. The fourth condition states that T_i-connected worlds are propositionally indiscernible, therefore transitions stand for purely epistemic actions.

Comparison to *Rasmussen*

Once the foundations of this first attempt are laid, we compare its workings with *Rasmussen*. As commented in Section 3.1.2, providing alternative axioms is not itself sufficient to overcome the problem. It is in the face of a semantic model that this enterprise contributes to the solution as only then do we obtain a hands-on understanding on the credibility of axioms and therefore on the adequacy of the proposed solution. To be more precise on the grounds of the comparison, we test whether our interpretation matches *Rasmussen*-like axiomatizations, i.e. we examine the validity of analogues to the axioms of Definition 3.1.3 and Definition 3.1.4 with respect to **M**.

Theorem 4.1.1 (**M**-validity test of proposed axioms)**.**

1. $\langle\ddagger\rangle^n K\phi \rightarrow \phi$ is valid in the class **M**.
2. $\langle\ddagger\rangle^n K\phi \rightarrow \langle\ddagger\rangle^n [\dagger]^m K\phi$ is valid in the class **M**.

3. $\langle\ddagger\rangle^n\phi \wedge \langle\dagger\rangle^m\psi \to \langle\ddagger\rangle^n\langle\dagger\rangle^m(\phi\wedge\psi)$ is *not* valid in the class **M**.

4. $\langle\ddagger\rangle^n(\phi\wedge\psi) \to \langle\ddagger\rangle^n\phi$ is valid in the class **M**.

Proof.

1. Let arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n K\phi$. Unpacking the sequence according to the abbreviation, $M, w \vDash \langle R_1\rangle \dots \langle R_n\rangle K\phi$, for the inference rules $R_1, \dots, R_n$. Following Definition 4.1.5, there is a world $u_1 \in W$ such that wT_1u_1 and $M, u_1 \vDash \langle R_2\rangle \dots \langle R_n\rangle K\phi$. Continuing like that, there is a world $u_n \in W$ such that $u_{n-1}T_nu_n$ and $M, u_n \vDash K\phi$, which in turn amounts to $\phi \in V_2(u_n)$. Then, by condition 3, $\phi \in V_1^*(u_n)$. From condition 4, $\phi \in V_1^*(u_{n-1})$. Continuing this process backwards, $\phi \in V_1^*(w)$. Therefore $M, w \vDash \phi$. Given the arbitrariness of $M \in \mathbf{M}$ and $w \in W$, we finally conclude that the formula is valid in the class **M**.

2. Let arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n K\phi$. Unpacking the sequence according to the abbreviation, this amounts to $M, w \vDash \langle R_1\rangle \dots \langle R_n\rangle K\phi$. As in the previous case, we obtain a chain $wT_1u_1 \dots u_{n-1}T_nu_n$ such that $M, u_n \vDash K\phi$, which in turn amounts to $\phi \in V_2(u_n)$ [1]. It suffices to show that $M, u_n \vDash [\dagger]^m K\phi$, i.e., by repeating the unpacking, now for $[\dagger]^m = [R'_1] \dots [R'_m]$, that for every world $v_1 \in W$ such that $u_nT'_1v_1$, …, for every world $v_m \in W$ such that $v_{m-1}T'_mv_m$, $M, v_m \vDash K\phi$, i.e. $\phi \in V_2(v_m)$. Let arbitrary such $v_1, \dots, v_m$. Then due to condition 2 and [1], $\phi \in V_2(v_1)$ and continuing in the same fashion $\phi \in V_2(v_m)$. Therefore, $M, w \vDash \langle\ddagger\rangle^n[\dagger]^m K\phi$, hence $M, w \vDash \langle\ddagger\rangle^n K\phi \to \langle\ddagger\rangle^n[\dagger]^m K\phi$, as desired.

3. Consider the following counterexample. Let a model $M = \langle W, T, V_1, V_2\rangle$ in the class **M**, as depicted in 4.2, and with every $V_2(w)$ containing all instances of Double Negation Introduction and Modus Ponens.

   It is easy to check that $M, w \vDash \langle DNI\rangle(\neg Kq)$, $M, w \vDash \langle MP\rangle Kq$. However, we cannot obtain $M, w \vDash \langle DNI\rangle\langle MP\rangle(\neg Kq \wedge Kq)$ as this would have meant that there is some world z with $q \in V_2(z)$ and $q \notin V_2(z)$.

4. Let arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n(\phi\wedge\psi)$. It follows, as above, that there is a chain $wT_1u_1 \dots u_{n-1}T_nu_n$ such that $M, u_n \vDash \phi\wedge\psi$, so $M, u_n \vDash \phi$. It then immediately follows that $M, w \vDash \langle\ddagger\rangle^n\phi$, as desired.

□

Note that the validities obtained above also hold in case each of the sequences involved is taken as "empty" (i.e. if we simply delete its occurrences from the claims). The proof for these cases is either completely straightforward or follows as a special case of the given proofs.

We continue by checking whether analogues of the axioms envisaged by *Rasmussen* in Definition 3.1.4, and the result of Theorem 3.1.1, alluding to specific rules the agent is supplied with, correspond to **M**-valid sentences. As far as the former is concerned, we want to evaluate the general case, dubbed (R_D) (Definition 3.1.4). But first, we need to build our version of it. Apart from the standard, by now, modifications regarding sequences, we substitute the "arbitrary conjunction of sentences" Δ with an arbitrary single sentence ϕ. Given the semantics, this modification does not alter the purpose of Δ; it is just more economical in terms of presentation. In addition, we have to account for the fact that the rule is available to the agent. Specifically: let rule $R_k = \{\phi_1, \dots, \phi_n\} \rightsquigarrow \psi$ and ϕ any sentence in our language $\mathcal{L}_{\text{RW}}$. Then, it is

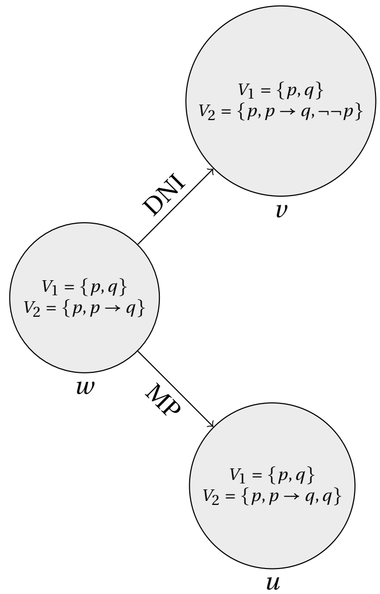


Figure 4.2: We use DNI and MP to talk about Double Negation Introduction and Modus Ponens respectively. We only write down the propositional elements of V_2 for brevity. Of course, given the conditions of **M**, more worlds and arrows should have been drawn but we omit those irrelevant for the purposes of the example for simplicity.

easy to see that $\langle\ddagger\rangle^n(\phi \wedge KR_k \wedge K\phi_1 \ldots \wedge K\phi_n) \to \langle\ddagger\rangle^n\langle R_k\rangle(\phi \wedge K\phi_1 \ldots \wedge K\phi_n \wedge K\psi)$ is *not* valid in the class **M**; simply consider $\phi := \neg K\psi$. Given that the motivating idea behind the introduction of such an axiom is to capture the effect of applying an inference rule R_k, this result should not come as a surprise. Not everything will survive the application of the inference rule. If we did not know ψ prior to this reasoning step, it is unnatural to ask that ψ remains to be unknown: this contradicts the very point of applying the inference rule. However, by imposing restrictions on ϕ, we can eventually reach a validity that captures the effect of applying the rule R_k. Similarly, we can "fix" the third case of the previous theorem and provide a validity that captures the effect of *Succession*, in terms of knowledge acquisition. We can additionally capture the desirable trait of a sequence of reasoning involving the rules Double Negation Elimination, Modus Ponens and Conjunction Introduction (as in Theorem 3.1.1). The next theorem gathers these results:

Theorem 4.1.2 (**M**-validities)**.**

1. $\langle\ddagger\rangle^n K\phi \wedge \langle\dagger\rangle^m K\psi \to \langle\ddagger\rangle^n\langle\dagger\rangle^m(K\phi \wedge K\psi)$ is valid in the class **M**.

2. Let rule $R_k = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$ and ϕ any sentence that is either propositional or a knowledge assertion of the form $K\chi$. Then:
$\langle\ddagger\rangle^n(\phi \wedge KR_k \wedge K\phi_1 \cdots \wedge K\phi_n) \to \langle\ddagger\rangle^n\langle R_k\rangle(\phi \wedge K\phi_1 \cdots \wedge K\phi_n \wedge K\psi)$ is valid in the class **M**.

3. Let the following instances of Double Negation Elimination, Modus Ponens and Conjunction Introduction denoted by DNE, MP and CI: $\{\neg\neg\phi\} \rightsquigarrow \phi$, $\{\phi, \phi \to \psi\} \rightsquigarrow \psi$, $\{\phi, \psi\} \rightsquigarrow$

$\phi \wedge \psi$. Then:

$\bigwedge_{R_k=DNE,MP,CI} KR_k \wedge K\neg\neg\phi \wedge K(\phi \to \psi) \to \langle DNE\rangle\langle MP\rangle\langle CI\rangle K(\phi \wedge \psi)$ is valid in the class **M**.

Proof.

1. Let arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n K\phi \wedge \langle\dagger\rangle^m K\psi$. So $M, w \vDash \langle\ddagger\rangle^n K\phi$ and $M, w \vDash \langle\dagger\rangle^m K\psi$. As above, we obtain a chain $wT_1u_1 \ldots u_{n-1}T_nu_n$ such that $M, u_n \vDash K\phi$, i.e. $\phi \in V_2(u_n)$, and a chain $wT'_1v_1 \ldots v_{m-1}T'_nv_m$ such that $M, v_m \vDash K\psi$, i.e. $\psi \in V_2(v_m)$. The rough idea of the proof is to make use of the conditions of **M** to merge the two chains. By condition 2, we know that $V_2(w) \subseteq V_2(u_n)$ and that $V_2(w)$ contains all the premises of rule R'_1, as well as the rule itself. Therefore, $V_2(u_n)$ in turn contains all the premises of rule R'_1 and the rule itself. By condition 1, there is a world z_1 such that $u_nT'_1z_1$ and $V_2(z_1) = V_2(u_n) \cup \{co(R'_1)\}$. Now again, by condition 2, $V_2(v_1) = V_2(w) \cup \{co(R'_1)\}$ and since $V_2(w) \subseteq V_2(u_n)$: $V_2(v_1) \subseteq V_2(z_1)$, so we know that z_1 contains the premises for R'_2 and the rule itself. Again by condition 1, there is a world z_2 such that $z_1T'_2z_2$ and $V_2(z_2) = V_2(z_1) \cup \{co(R'_2)\}$. Continuing like that, the alternations of condition 2 and condition 1, based on the initial assumptions, yield a world z_m such that $z_{m-1}T'_mz_m$ and $V_2(z_m) = V_2(z_{m-1}) \cup \{co(R'_m)\}$ with $V_2(v_m) \subseteq V_2(z_m)$. Therefore $\psi \in V_2(z_m)$. In addition, as the constructed chain is of the form $u_nT'_1z_1T'_2z_2 \ldots T'_mz_m$ and due to condition 2, $\phi \in V_2(z_m)$. So $M, z_m \vDash K\phi \wedge K\psi$, i.e. $M, u_n \vDash \langle\dagger\rangle^m(K\phi \wedge K\psi)$. So finally $M, w \vDash \langle\ddagger\rangle^n\langle\dagger\rangle^m(K\phi \wedge K\psi)$, as desired.

2. Let arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n(\phi \wedge KR_k \wedge K\phi_1 \wedge \ldots \wedge K\phi_n)$. Then, repeating the unpacking of the sequence as in the earlier examples, there is a chain $wT_1u_1 \ldots T_nu_n$ such that $M, u_n \vDash \phi \wedge KR_k \wedge K\phi_1 \wedge \ldots \wedge K\phi_n$. But then $M, u_n \vDash \phi$ and $R_k, \phi_1, \ldots, \phi_n \in V_2(u_n)$. Next, from condition 1, there is $v \in W$ such that $u_nT_{R_k}v$ and $V_2(v) = V_2(u_n) \cup \{\psi\}$. As a result, $M, v \vDash K\phi_1 \wedge \ldots \wedge K\phi_n \wedge K\psi$. If ϕ is propositional, then by condition 4, $\phi \in V_1^*(v)$, i.e. $M, v \vDash \phi$ too. If $\phi := K\chi$, then $\chi \in V_2(u_n)$ so immediately again $\chi \in V_2(v)$, i.e. $M, v \vDash K\chi$. Indeed, in both these cases $M, v \vDash \phi \wedge K\phi_1 \wedge \ldots \wedge K\phi_n \wedge K\psi$. Finally, wrapping this up backwards $M, w \vDash \langle\ddagger\rangle^n\langle R_k\rangle(\phi \wedge K\phi_1 \wedge \ldots \wedge K\phi_n \wedge K\psi)$, as desired.

3. Let arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \vDash \bigwedge_{R_k=DNE,MP,CI} KR_k \wedge K\neg\neg\phi \wedge K(\phi \to \psi)$. Then $DNE, MP, CI \in V_2(w)$, $\neg\neg\phi \in V_2(w)$ and $\phi \to \psi \in V_2(w)$. By condition 1, we get that there is $u \in W$ with $wT_{DNE}u$ and $M, u \vDash K(CI) \wedge K(MP) \wedge K(\phi \to \psi) \wedge K\phi$. Likewise, there is $v \in W$ with $uT_{MP}v$ and $M, v \vDash K(CI) \wedge K\phi \wedge K\psi$. Finally, there is $z \in W$ with $vT_{CI}z$ and $M, z \vDash K(\phi \wedge \psi)$. It is now easy to see that $M, w \vDash \langle DNE\rangle\langle MP\rangle\langle CI\rangle K(\phi \wedge \psi)$, as desired.

□

Again, the claims persist as special cases of the above, if we allow for "empty" sequences, as Rasmussen does.

We have thus far presented a semantics that improves *Rasmussen*-style syntactic proposals and demonstrates how knowledge is attained as the course of reasoning evolves. As a result, we managed to find a remedy from what was pointed out as a deficiency of this otherwise balanced conception, while explicitly spelling out reasoning steps in an intuitively plausible way. It is therefore expected that this fuller framework preserves (and in fact, completes) the attack against logical omniscience. Indeed, the closure principles are invalidated, without a simultaneous collapse to a state of utter ignorance. At the same time, suitable applications of inference rules, reflecting the effort to eventually reach a conclusion, ensure that an agent can *come to know* consequences of her knowledge.

These claims are justified as follows: both the full and weaker forms of omniscience are avoided because the values of knowledge assertions are essentially determined by the valuation V_2, which need not obey any closure principle. In any case, actual knowledge at a particular point of time carries no commitments for the agent. On the other hand, this does not mean that the agent is ignorant or incapable of acquiring knowledge. Once she follows the appropriate reasoning track, she can draw the consequences from what she already knows. Moreover, rationality standards are preserved because we unfold the actual process that resulted in knowledge and account for its dynamic nature. Thus ignorance is ruled out because of a more realistic modelling, and not because of additional restrictions imposed on an inflexible conception of knowledge. In fact, validities like the ones of Theorem 4.1.2 exemplify the importance of *reasoning* and help us avoid the implausible commitment to an automatic and effortless way to expand one's epistemic state, as the mainstream validity $K\phi_1 \wedge \ldots \wedge K\phi_n \to K\psi$ would dictate. Overall, we have completed the picture of *Rasmussen*, formally explaining how a real, resource-bounded agent, is able to exploit the available resources and acquire knowledge via reasoning.

A sound and complete axiomatization

Departing from the comparison with Rasmussen's approach and the interesting validities, we develop the logic Λ_{RW}. We can therefore argue for the superiority of this fuller framework in terms of explaining real reasoning processes, via logical modelling.

Definition 4.1.7 (Axiomatization of Λ_{RW})**.** The logic Λ_{RW} is axiomatized by the following axiom schemes:

(PC) All instances of classical propositional tautologies.

(T') $K\phi \to Tr(\phi)$.

(AX1) $[R_k](\phi \to \psi) \to ([R_k]\phi \to [R_k]\psi)$.

(AX2) $\bigwedge_{\phi \in pr(R_k)} K\phi \wedge K(R_k) \to \langle R_k \rangle (\bigwedge_{\phi \in pr(R_k)} K\phi \wedge Kco(R_k))$.

(AX3) $\langle R_k \rangle K\chi \to \bigwedge_{\phi \in pr(R_k)} K\phi \wedge KR_k \wedge K\chi$, for $\chi \neq co(R_k)$.

(AX4) $[R_k]Kco(R_k)$.

(AX5) $\langle R_k \rangle \phi \to \phi$, for $\phi \in \mathcal{L}_P$.

(AX6) $\phi \to [R_k]\phi$, for $\phi \in \mathcal{L}_P$.

(AX7) $K\chi \to [R_k]K\chi$.

and the rules:

(MP) From ϕ and $\phi \to \psi$, infer ψ.

(R_k-N) From ϕ infer $[R_k]\phi$.

Theorem 4.1.3 (Soundness)**.** The logic Λ_{RW} is sound with respect to the class **M**.

Proof. It suffices to show that the axioms of Definition 4.1.7 are valid in the class **M**, as our rules preserve validity as usual.

(PC) Trivial.

(T') Follows immediately from condition 3.

(AX1) Let arbitrary model $M \in \mathbf{M}$ and world $w \in W$ of the model such that $M, w \models [R_k](\phi \to \psi)$ and $M, w \models [R_k]\phi$. Then, for every $u \in W$ with $wT_k u$: $M, u \models \phi \to \psi$ and $M, u \models \phi$. It immediately follows that $M, u \models \psi$, so $M, w \models [R_k]\psi$, as desired.

(AX2) Follows as a special case of Theorem 4.1.2.

(AX3) Let any model $M \in \mathbf{M}$ and world $w \in W$ of the model such that $M, w \models \langle R_k \rangle K\chi$, for $R_k = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$. So there is $u \in W$ such that $wT_k u$ and $\chi \in V_2(u)$. By condition 2, $\phi_1, \ldots, \phi_n, R_k \in V_2(w)$ and since $V_2(u) = V_2(w) \cup \{\psi\}$, $\chi \in V_2(w) \cup \{\psi\}$. So either $\chi \in V_2(w)$ or $\chi = \psi$. Finally, $M, w \models K\phi_1 \wedge \ldots \wedge K\phi_n \wedge KR_k \wedge K\chi$, for $\chi \neq \psi$.

(AX4) Immediate due to condition 2.

(AX5) Let any model $M \in \mathbf{M}$ and world $w \in W$ of the model such that $M, w \models \langle R_k \rangle \phi$ for $\phi \in \mathcal{L}_P$. Then, there is $u \in W$ such that $wT_k u$ and $M, u \models \phi$, i.e. $\phi \in V_1^*(u)$. By condition 4, $\phi \in V_1^*(w)$, i.e. $M, w \models \phi$ as desired.

(AX6) Let any model $M \in \mathbf{M}$ and world $w \in W$ of the model such that $M, w \models \phi$. Let any $u \in W$ such that $wT_k u$. Then by condition 4, $\phi \in V_1^*(u)$, i.e. $M, u \models \phi$ so $M, w \models [R_k]\phi$, as desired.

(AX7) Let any model $M \in \mathbf{M}$ and world $w \in W$ of the model such that $M, w \models K\chi$, i.e. $\chi \in V_2(w)$. Let any $u \in W$ such that $wT_k u$. From condition 2, $\chi \in V_2(u)$, i.e. $M, u \models K\chi$. But then indeed $M, w \models [R_k]K\chi$.

□

Aiming at completeness, we follow the procedure of Blackburn et al. (2001)[6]. That is, we are going to show that Λ_{RW} is (strongly) complete with respect to **M**, through the construction of a *canonical model*.

Theorem 4.1.4 (Lindenbaum's Lemma)**.** If Γ is a Λ_{RW}-consistent set of formulas, then it can be extended to a maximal Λ_{RW}-consistent set Γ^+.

Proof. The proof goes as usual in these cases. After enumerating $\phi_0, \phi_1, \ldots$, the formulas of our language, one constructs the set Γ^+ as $\bigcup_{n \geq 0} \Gamma^n$ where: $\Gamma^0 = \Gamma$, $\Gamma^{n+1} = \Gamma^n \cup \{\phi_n\}$, if this is Λ_{RW}-consistent and $\Gamma^n \cup \{\neg\phi_n\}$ otherwise. The desired properties are easily obtained due to this construction. □

[6] The reader may consult the book for background details.

Definition 4.1.8 (Canonical Model)**.** The *canonical model* $\mathcal{M}$ for Λ_{RW} is a tuple $\langle \mathcal{W}, \mathcal{T}, \mathcal{V}_1, \mathcal{V}_2 \rangle$ where:

- $\mathcal{W} = \{w \mid w$ a maximal Λ_{RW}-consistent set$\}$.
- $\mathcal{T} : \mathcal{R} \to \mathcal{P}(\mathcal{W} \times \mathcal{W})$, such that for $R_i \in \mathcal{R}$, $\mathcal{T}(R_i) = \mathcal{T}_i$, where $w\mathcal{T}_i u$ if and only if $\{\langle R_i \rangle \phi \mid \phi \in u\} \subseteq w$.
- $\mathcal{V}_1 : \mathcal{W} \to \mathcal{P}(\Phi)$ such that $\mathcal{V}_1(w) = \{p \in \Phi \mid p \in w\}$.
- $\mathcal{V}_2 : \mathcal{W} \to \mathcal{P}(\mathcal{L})$ such that $\mathcal{V}_2(w) = \{\phi \in \mathcal{L} \mid K\phi \in w\}$.

It is easy to see that an equivalent formulation for the definition of $\mathcal{T}_i$ is $\{\phi \mid [R_i]\phi \in w\} \subseteq u$. Given the definition of the canonical model and our language $\mathcal{L}_{\text{RW}}$, we show:

Lemma 1 (Existence lemma)**.** For any formula ϕ in our language and $w \in \mathcal{W}$, if $\langle R_i \rangle \phi \in w$ then there is a world $u \in \mathcal{W}$ such that $w\mathcal{T}_i u$ and $\phi \in u$.

Proof. Suppose $\langle R_i \rangle \phi \in w$. Take $S = \{\phi\} \cup \{\psi \mid [R_i]\psi \in w\}$. This set is consistent. Were it inconsistent, there would be $\psi_1, \ldots, \psi_n$ such that $\vdash_{\Lambda_{\text{RW}}} \psi_1 \wedge \ldots \wedge \psi_n \to \neg\phi$. Using $[R_i]$-necessitation, distribution and propositional tautologies we obtain $\vdash_{\Lambda_{\text{RW}}} ([R_i]\psi_1 \wedge \ldots \wedge [R_i]\psi_n) \to [R_i]\neg\phi$. By the property of w as maximal consistent set and since $[R_i]\psi_1, \ldots, [R_i]\psi_n \in w$: $[R_i]\neg\phi \in w$. Therefore $\neg\langle R_i \rangle \phi \in w$. Indeed, we have reached a contradiction. Next, we extend S to S^+ according to Lindenbaum's lemma. Then, $\phi \in S^+$ and $[R_i]\psi \in w$ implies $\psi \in S^+$. Take $u := S^+$. As a result, $w\mathcal{T}_i u$ and $\phi \in u$. □

Lemma 2 (Truth lemma)**.** For any formula ϕ in our language and world $w \in \mathcal{W}$: $\mathcal{M}, w \vDash \phi$ if and only if $\phi \in w$.

Proof. The proof is by induction on the complexity of ϕ.

- Base cases: Consider $\phi := p$ with $p \in \Phi$. Then $\mathcal{M}, w \vDash p$ if and only if $p \in \mathcal{V}_1(w)$, and by definition, this is the case if and only if $p \in w$. Next, take $\phi := K\psi$ with $\psi \in \mathcal{L}$. Then $\mathcal{M}, w \vDash K\psi$ if and only if $\psi \in \mathcal{V}_2(w)$, and by definition, this is the case if and only if $K\psi \in w$.
- Consider $\phi := \neg\psi$ with Induction Hypothesis that the result holds for ψ. Then $\mathcal{M}, w \vDash \neg\psi$ if and only if $\mathcal{M}, w \nvDash \psi$ and by I.H. this is the case if and only if $\psi \notin w$. Since w is maximal consistent, this is the case if and only if $\neg\psi \in w$.
- Consider $\phi := \psi \wedge \chi$ with Induction Hypothesis that the result holds for ψ and χ. Then $\mathcal{M}, w \vDash \psi \wedge \chi$ if and only if $\mathcal{M}, w \vDash \psi$ and $\mathcal{M}, w \vDash \chi$. By I.H. this is the case if and only if $\psi \in w$ and $\chi \in w$ and again by the maximal consistency of w, $\psi \wedge \chi \in w$.
- Consider $\phi := \langle R_i \rangle \psi$ with Induction Hypothesis that the result holds for ψ. Then $\mathcal{M}, w \vDash \langle R_i \rangle \psi$ if and only if there is $u \in \mathcal{W}$ such that $w\mathcal{T}_i u$ and $\mathcal{M}, u \vDash \psi$. By I.H. this is the case if and only if $\psi \in u$, and by definition of $\mathcal{T}_i$, we get $\langle R_i \rangle \psi \in w$. The other direction follows immediately from the existence lemma.

□

Theorem 4.1.5 (Completeness)**.** For any set of formulas Γ and formula ϕ in our language: $\Gamma \vDash_{\mathbf{M}} \phi$ only if $\Gamma \vdash_{\Lambda_{\text{RW}}} \phi$.

Proof.

- We first expand Γ to a maximal Λ_{RW}-consistent set Γ^+. Then, let the canonical model $\mathcal{M}$ as constructed according to Definition 4.1.8. Then by Lemma 2, $\mathcal{M}, \Gamma^+ \vDash \Gamma$. It suffices to show that $\mathcal{M}$ fulfills the conditions of **M**.

- Condition 1 is satisfied.

  Let inference rule $R_i = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$ and $w \in \mathcal{W}$ with $R_i, \phi_1, \ldots, \phi_n \in \mathcal{V}_2(w)$, i.e. $K(R_i)$, $K\phi_1, \ldots, K\phi_n \in w$ [1]. We want to show that there is a world $u \in \mathcal{W}$ such that $w\mathcal{T}_i u$ and $\mathcal{V}_2(u) = \mathcal{V}_2(w) \cup \{\psi\}$. From [1], $K(R_i) \wedge K\phi_1 \wedge \ldots \wedge K\phi_n \in w$. But from (AX2), we also get that $\langle R_i \rangle (K\phi_1 \wedge \ldots \wedge K\phi_n \wedge K\psi) \in w$. Now, using the existence lemma, there is $u \in \mathcal{W}$ such that $w\mathcal{T}_i u$ and $K\phi_1 \wedge \ldots K\phi_n \wedge K\psi \in u$. It follows that $K\phi_1$, …, $K\psi \in u$, therefore $\phi_1, \ldots, \phi_n, \psi \in \mathcal{V}_2(u)$. Then, take any $\chi \in \mathcal{V}_2(w)$, so $K\chi \in w$. By (AX7), $[R_i]K\chi \in w$ and by definition of $\mathcal{T}$ we get $K\chi \in u$, i.e. $\chi \in \mathcal{V}_2(u)$. Hence, $\mathcal{V}_2(w) \cup \{\psi\} \subseteq \mathcal{V}_2(u)$. Finally, take any $\chi \in \mathcal{V}_2(u)$ with $\chi \neq \psi$. Then $\langle R_i \rangle K\chi \in w$, by definition of $\mathcal{T}$, and by (AX3): $K\chi \in w$, i.e. $\chi \in \mathcal{V}_2(w)$ too. So indeed, if $\phi \in \mathcal{V}_2(u)$ then $\phi = \psi$ or $\phi \in \mathcal{V}_2(w)$, i.e. $\mathcal{V}_2(u) \subseteq \mathcal{V}_2(w) \cup \{\psi\}$. Therefore, $\mathcal{V}_2(u) = \mathcal{V}_2(w) \cup \{\psi\}$.

- Condition 2 is satisfied.

  Suppose that $w\mathcal{T}_i u$ with $R_i = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$, i.e. if $\phi \in u$ then $\langle R_i \rangle \phi \in w$. Let arbitrary $\chi \in \mathcal{V}_2(u)$. That is, $K\chi \in u$. Therefore, $\langle R_i \rangle K\chi \in w$. From (AX3), indeed $\phi_1, \ldots, \phi_n, R_k \in \mathcal{V}_2(w)$. From (AX4) and definition of $\mathcal{T}_i$, $K\psi \in u$, i.e. $\psi \in \mathcal{V}_2(u)$. Furthermore again by this definition and (AX7) we obtain that $\mathcal{V}_2(w) \subseteq \mathcal{V}_2(u)$. Therefore, $\mathcal{V}_2(w) \cup \{\psi\} \subseteq \mathcal{V}_2(u)$. Next suppose that there is $\phi \in \mathcal{V}_2(u)$ and $\phi \neq \psi$. Then $\langle R_k \rangle K\phi \in w$. From (AX3), $K\phi \in w$. As a result, $\phi \in \mathcal{V}_2(w)$. Clearly then, $\mathcal{V}_2(u) = \mathcal{V}_2(w) \cup \{\psi\}$.

- Condition 3 is satisfied.

  Let $\phi \in \mathcal{L}$. Suppose that $\phi \in \mathcal{V}_2(w)$. That is, $K\phi \in w$. Then by (T') we obtain, $Tr(\phi) \in w$, that is $\mathcal{M}, w \vDash Tr(\phi)$ and therefore $Tr(\phi) \in \mathcal{V}_1^*(w)$.

- Condition 4 is satisfied.

  Take $w, u \in \mathcal{W}$ and $w\mathcal{T}_i u$. By definition of $\mathcal{T}_i$, if $\phi \in u$ then $\langle R_i \rangle \phi \in w$. Now take arbitrary $\phi \in \mathcal{L}_P$ such that $\mathcal{M}, u \vDash \phi$, i.e. $\phi \in \mathcal{V}_1^*(u)$. This means that $\phi \in u$, therefore $\langle R_i \rangle \phi \in w$. From (AX5), we obtain $\phi \in w$, i.e. $\mathcal{M}, w \vDash \phi$ so $\phi \in \mathcal{V}_1^*(w)$. As ϕ was arbitrary, $\mathcal{V}_1^*(u) \subseteq \mathcal{V}_1^*(w)$. For the other inclusion, take arbitrary $\phi \in \mathcal{L}_P$ such that $\mathcal{M}, w \vDash \phi$, i.e. $\phi \in \mathcal{V}_1^*(w)$. This means that $\phi \in w$. From (AX6), we get that $[R_i]\phi \in w$ too. Then we exploit the alternative definition of $\mathcal{T}_i$; since $[R_i]\phi \in w$, $\phi \in u$, i.e. $\mathcal{M}, u \vDash \phi$ so $\phi \in \mathcal{V}_1^*(u)$. As ϕ was arbitrary, $\mathcal{V}_1^*(w) \subseteq \mathcal{V}_1^*(u)$. Overall, $\mathcal{V}_1^*(w) = \mathcal{V}_1^*(u)$.

□

On a final note, we describe how this attempt can be adjusted for doxastic frameworks. Given that worlds are representations of knowledge and also given the (indirect) syntactic flavour of the resulting framework, we simply propose an analogous idea: worlds are utilized as representations of belief as well, via attaching another labelling function to the model, which fixes what the agent believes. Nevertheless, not everything works in complete analogy. It is usually assumed that belief is not factive and as a result the status of the third condition suggested for the class **M** is debatable. In addition, we may not impose that doxastic states are preserved after each step, because the agent's beliefs are revizable. Given our purposes, it is indeed reasonable to drop these conditions under a realistic doxastic framework.

4.1.2 Impossible worlds (IW)

This second approach introduces an impossible-worlds framework, similar to the one of Section 3.3.3 (henceforth, *Rasmussen*&*Bjerring*). This enables us to reap the harvest of destroying closure principles via allowing impossible worlds to be epistemically accessible. Yet instead of abstracting away from the inner details of the reasoning process, substituting the arrays of inference rules with generic "inference steps", this approach captures the effect of the application of specific inference rules, their order and their cognitive significance. By spelling out the reasoning process in a more elaborate way, we do not have to rely on a debatable initial cutoff of n steps. Rather, our goal is to make explicit the intuition underlying the existence of this cutoff: it is the depletion of resources that eventually brings the pursuit of knowledge to a halt.

We present our full-fledged account in a gradual manner. First, we provide the basic material of this second approach and the comparison to the syntactic view, and only then do we reveal a quantitative extension that encapsulates the motivation given above.

Basic elements

To begin with, we define the language in the spirit of *Rasmussen*&*Bjerring*:

Definition 4.1.9 (Language $\mathcal{L}_{\text{IW}}$)**.** The language $\mathcal{L}_{\text{IW}}$ is defined inductively from a countable set of propositional atoms Φ, the epistemic operator K and dynamic operators $\langle R_k \rangle$ and $[R_k]$ as:

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi\wedge\phi \quad | \quad K\chi \quad | \quad \langle R_k\rangle\phi \quad | \quad [R_k]\phi$$

with $p \in \Phi$, $\chi \in \mathcal{L}_P$ where $\mathcal{L}_P$ is the standard propositional language, and $R_k \in R$, for a given countable set R.

Intuitively, R is a countable set of inference rules available to the agent[7]. Then, $\langle R_k\rangle\phi$ stands for "after some application of inference rule R_k, ϕ is true", while "any" replaces "some" for $[R_k]\phi$. It is worth noticing that unlike *Rasmussen*, we do not attach cognitive costs to the language. Still, we deal with the part cognitive effort plays through the updates of the models, which we now define:

Definition 4.1.10 (IW-model)**.** An IW-*model* is a tuple $M = \langle W^P, W^I, f, V, R, Res, C, cp\rangle$ where:

- W^P, W^I, f, V are as in Definition 3.3.12; note that f now reflects epistemic accessibility.
- R is the countable set of inference rules available to the agent.
- Res is a finite set of resources, such as *memory*, *time* etc. Let $r := |Res|$.
- $C : R \to \mathbb{N}^r$ is a function such that every inference rule $R_k \in R$ is assigned a particular *cognitive cost* for each resource.
- cp denotes the agent's cognitive capacity, i.e. $cp \in \mathbb{N}^r$, intuitively standing for what the agent is able to afford with regard to each resource.

[7] It is natural to think of R containing *sound* rules, at least in a purely epistemic system. In particular, R can be thought as a set of schemes in the context of this attempt.

Again, impossible worlds only comply with (Minimal Consistency) while possible worlds are complete and consistent entities. In addition, the valuation V from $W := W^P \cup W^I$ to $\mathcal{P}(\mathcal{L}_{\text{IW}})$ assigns the set of true *atomic* sentences at *possible* worlds, whilst assigning the set of *all* true sentences, atomic or composite, at *impossible* worlds. Furthermore, $\Gamma \vdash_{R_k} \Gamma'$ means that Γ proves Γ' within an application of R_k (recall $\Gamma \vdash^n_{\mathcal{R}} \Gamma'$, meaning that Γ proves Γ' within n-steps of reasoning from $\mathcal{R}$, in Section 3.3.3 [8]). Next, we adapt several definitions to account for individual and rule-specific changes of our models.

Definition 4.1.11 (Rule-specific radius)**.** Given an inference rule $R_k \in R$, the R_k-radius of a world $w \in W$ is $w^{R_k} = \{w' \mid V(w) \vdash_{R_k} V(w')\}$.

Then, a member of w^{R_k} is an R_k*-expansion* of w. In our view, $\vdash_{R_k}$ then requires that $V(w')$ preserves $V(w)$ and extends it just by a conclusion of R_k. This is also how we capture the monotonic property: R_k-expansions, as the name indicates, enrich the state from which they originate, in terms of R_k. In this sense, inferences are not defeated as reasoning steps are taken, thereby providing (Monotonicity), in our reading too. As before, $w^{R_k} = \{w\}$ for $w \in W^P$ (due to the deductive closure of possible worlds[9]) while the R_k-radius of impossible worlds can contain different R_k-expansions. Analogously:

Definition 4.1.12 (Rule-specific variation)**.** Let an IW-model $M = \langle W^P, W^I, f, V, R, Res, C, cp\rangle$ and inference rule $R_k \in R$. Then $\mathcal{F}^{R_k}$ is a function from pointed models to sets of accessibility functions defined as:

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

where $c \in \mathcal{C}(\{w'^{R_k} \mid w' \in f(w)\})$. A member of $\mathcal{F}^{R_k}$ is called an R_k*-variation* of f.

In other words, for an IW-model M and world w, an accessibility function g is an R_k-variation if $g(w)$ is the outcome of a choice, originating from the R_k-radii of w's accessible worlds.

Definition 4.1.13 (rule-specific accessibility)**.** Let $M = \langle W^P, W^I, f, V, R, Res, C, cp\rangle$ and $M' = \langle W^{P'}, W^{I'}, f', V', R', Res', C', cp'\rangle$ be IW-models and $R_k \in R$. Then, $(M,w) \sim^{R_k} (M',w')$ if and only if $w' = w$, $W = W'$, $V' = V$, $Res' = Res$, $R = R'$, $C = C'$, $f' \in \mathcal{F}^{R_k}(M,w)$ and $cp' = cp - C(R_k)$.

That is, (M',w') is R_k-accessible from (M,w), just in case (a) the set of all epistemically accessible worlds from w in M is replaced in M' by a choice of R_k-expansions of w's accessible worlds in M and (b) the cognitive capacity is reduced by the cost of performing an R_k-step.

The semantics is then given by:

Definition 4.1.14 (IW-semantics)**.** We use $M,w \vDash \phi$ to say that ϕ is true at w in M, and $M,w \dashv \phi$ to say that ϕ is false at w in model M. For $w \in W^P$:

[8] At this point, a clarification may be needed. In Section 3.3.3, we did not explicitly write down a precise characterization of what is meant by the last sentence, to remain faithful to the presentation of the authors, who did not provide an explicit definition. Of course, given the context of their work, one can imagine what is meant by it, i.e. there seem to be some implicit assumptions on their understanding of the notation. In any case, it does no harm for us to explicitly state such assumptions, as we continue, at least regarding our own system (that is, on the properties of $\vdash_{R_k}$) and prevent any possible confusion.

[9] Again, accepting this in the lines of *Rasmussen*&*Bjerring* seems to imply that possible worlds are valuation-wise unique.

1. $M, w \vDash p$ if only if $p \in V(w)$, where $p \in \Phi$.
2. $M, w \vDash \neg\phi$ if and only if $M, w \nvDash \phi$.
3. $M, w \vDash \phi \wedge \psi$ if and only if $M, w \vDash \phi$ and $M, w \vDash \psi$.
4. $M, w \vDash K\phi$ if and only if $M, w' \vDash \phi$ for all $w' \in f(w)$.
5. $M, w \vDash \langle R_k \rangle \phi$ if and only if $M', w' \vDash \phi$ for some $(M', w') : (M, w) \sim^{R_k} (M', w')$.
6. $M, w \vDash [R_k]\phi$ if and only if $M', w' \vDash \phi$ for all $(M', w') : (M, w) \sim^{R_k} (M', w')$.
7. $M, w \dashv \phi$ if and only if $M, w \nvDash \phi$.

For $w \in W^I$:

1. $M, w \vDash \phi$ if and only if $\phi \in V(w)$.
2. $M, w \dashv \phi$ if and only if $\neg\phi \in V(w)$.

Validity is defined as usual, with respect to the possible worlds.

Again, for notational convenience, it is useful to abbreviate sequences of inference rules as in the first approach: namely, we will use $\langle \ddagger \rangle^n \phi$ instead of $\langle R_1 \rangle \ldots \langle R_n \rangle$. Based on our semantics, we now give the truth conditions for sentences prefixed by a sequence[10], given through successive model changes activated by the inference rules involved.

Let model $M = \langle W^P, W^I, f, V, R, Res, C, cp \rangle$ and world w of the model, such that $M, w \vDash \langle \ddagger \rangle^n \phi$. Unpacking the sequence: $M, w \vDash \langle R_1 \rangle \ldots \langle R_n \rangle \phi$. According to Definition 4.1.14, $M', w' \vDash \langle R_2 \rangle \ldots \langle R_n \rangle \phi$, for some (M', w'): $(M, w) \sim^{R_1} (M', w')$. According to Definition 4.1.13, this means that $M', w \vDash \langle R_2 \rangle \ldots \langle R_n \rangle \phi$ for some $\langle W^P, W^I, f', V, R, Res, C, cp' \rangle$, where $f' \in \mathcal{F}^{R_1}(M, w)$ and $cp' = cp - C(R_1)$. By Definition 4.1.12, $f'(w) = c'$ for some choice $c' \in C(\{v^{R_1} \mid v \in f(w)\})$. Likewise, there is (M'', w) where $f'' \in \mathcal{F}^{R_2}(M', w)$ and $cp'' = cp' - C(R_2)$ such that $M'', w \vDash \langle R_3 \rangle \ldots \langle R_n \rangle \phi$. That is, there is model M'' with $f''(w) = c''$ for some choice $c'' \in C(\{v^{R_2} \mid v \in f'(w)\})$ such that $M'', w \vDash \langle R_3 \rangle \ldots \langle R_n \rangle \phi$. Continuing like that we get a model M^n with $f^n(w)$ for some $c^n \in C(\{v^{R_n} \mid v \in f^{n-1}(w)\})$ and $M^n, w \vDash \phi$. Bringing these steps together, we get that there is a model M^* with $f^*(w) = c^*$ for some $c^* \in C(\{v^{R_1,\ldots,R_n} \mid v \in f(w)\})$ and $M^*, w \vDash \phi$. Note that: $v^{R_1,\ldots,R_n} = \{w' \mid V(v) \vdash_{R_1,\ldots,R_n} V(w')\}$ with $V(v) \vdash_{R_1,\ldots,R_n} V(w')$ denoting that $V(w')$ follows by successive applications of $R_1, \ldots, R_n$ (in *this* order) from $V(v)$. This sort of notation, involving an array of rules, can be generalized accordingly for $\mathcal{F}$ and thus $\sim$.

Comparison to ***Rasmussen***

As with the first RW approach, we adapt the axioms put forward by Rasmussen to our language and check their compatibility with IW-semantics.

[10] This analysis is spelled out here as it is useful for the subsequent sections, where, for brevity, its outcome will be directly retrieved. That is, we will not write down the whole procedure that gives rises to an updated model M^*. Nor will we write down every component of it, other than specifying the value of $f^*(w)$, which, at this level, monopolizes our interest in the updated model. Finally, note that for cases of “empty” sequences, the claims that follow can be simply obtained as special cases.

Theorem 4.1.6 (IW-validity test of proposed axioms)**.**

1. $\langle\ddagger\rangle^n K\phi \to \phi$ is *not* valid in the class of *all* models.
2. $\langle\ddagger\rangle^n K\phi \to \langle\ddagger\rangle^n[\dagger]^m K\phi$ is valid in the class of all models.
3. $\langle\ddagger\rangle^n \phi \wedge \langle\dagger\rangle^m \psi \to \langle\ddagger\rangle^n\langle\dagger\rangle^m(\phi\wedge\psi)$ is *not* valid in the class of all models.
4. $\langle\ddagger\rangle^n(\phi\wedge\psi) \to \langle\ddagger\rangle^n\phi$ is valid in the class of all models.

Proof.

1. If the choice c^*, obtained from the analysis above, does not contain the world of evaluation itself, constructing counterexamples is a trivial task[11].

2. Let arbitrary model M and world $w \in W^P$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n K\phi$. Then, there is a model M^* with $f^*(w) = c^*$ for some choice $c^* \in C(\{v^{R_1,\dots,R_n} \mid v \in f(w)\})$ such that $M^*, w \vDash K\phi$. As a result, $M^*, u \vDash \phi$ for all $u \in c^*$ [1]. We want to show that $M, w \vDash \langle\ddagger\rangle^n[\dagger]^m K\phi$. So it suffices to show that $M^*, w \vDash [\dagger]^m K\phi$, i.e. that for all (M^{**}, w) with $f^{**}(w) = c^{**} \in C(\{v^{R'_1,\dots,R'_m} \mid v \in f^*(w)\})$: $M^{**}, w \vDash K\phi$. Let arbitrary such M^{**}. From [1] and (Monotonicity), $M^{**}, w \vDash K\phi$. As a result, $M^*, w \vDash [\dagger]^m K\phi$, as desired.

3. Consider the following counterexample. Let model $M = \langle W^P, W^I, f, V, R, Res, C, cp\rangle$ and world $w \in W^P$ such that $M, w \vDash \langle R_k\rangle Kp$ [1] and $M, w \vDash \langle R_l\rangle K\neg p$ [2] for inference rules R_k and R_l. Due to the semantics, from [1] we obtain that there is M' with $f'(w) = c'$ for some choice $c' \in C(\{v^{R_k} \mid v \in f(w)\})$ such that $M', w \vDash Kp$, i.e. $M', u \vDash p$ for all $u \in c'$. Likewise, from [2], there is M'' with $f''(w) = c''$ for some choice $c'' \in C(\{v^{R_l} \mid v \in f(w)\})$ such that $M'', w \vDash K(\neg p)$, i.e. $M'', u \vDash \neg p$ for all $u \in c''$. But then, it cannot be that there is a choice $c^* \in C(\{v^{R_l} \mid v \in f'(w)\})$ such that for M^* with $f^*(w) = c^*$, $M^*, w \vDash Kp \wedge K\neg p$, because then $M^*, u \vDash p$ and $M^*, u \vDash \neg p$ for all $u \in c^*$ which violates (Minimal Consistency).

4. Let arbitrary model M and world $w \in W^P$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n(\phi\wedge\psi)$. Then, there is a model M^* with $f^*(w) = c^*$ for some $c^* \in C(\{v^{R_1,\dots,R_n} \mid v \in f(w)\})$ such that $M^*, w \vDash \phi\wedge\psi$ and thus $M^*, w \vDash \phi$. But then clearly $M, w \vDash \langle\ddagger\rangle^n\phi$.

□

Again, we continue by checking whether analogues of (R_D) axiom and Theorem 3.1.1 are valid under the semantics of this approach. It is easy to see that for $\Delta := K\neg\psi$, from (Minimal Consistency), we obtain a counterexample for (R_D). In addition, we can fix the cases of failure demonstrated in the previous theorem, by modifying Rasmussen's axioms, arguing on why the new view constitutes a more credible reading. In particular, we will see that imposing reflexivity secures veridicality of knowledge. Since this is a desirable trait, as in our familiar systems, we will give the other validities with respect to the class of reflexive models. Overall, our results will naturally hint at a remarkable consonance of RW- and IW- results, in terms of the validity of the axioms in question.

Theorem 4.1.7 (IW-validities)**.**

1. $\langle\ddagger\rangle^n K\phi \to \phi$ is valid in the class of all *reflexive* models.

[11] Counterexamples can be easily constructed for the "empty" case.

2. $\langle\ddagger\rangle^n K\phi \wedge \langle\dagger\rangle^m K\psi \rightarrow \langle\ddagger\rangle^n\langle\dagger\rangle^m(K\phi \wedge K\psi)$, with $\phi \neq \neg\psi$, is valid in the class of all reflexive models.

3. If $\{\phi_1,\ldots,\phi_n\} \vdash_{R_k} \psi$ then: $\langle\ddagger\rangle^n(\phi \wedge K\phi_1 \cdots \wedge K\phi_n) \rightarrow \langle\ddagger\rangle^n\langle R_k\rangle(\phi \wedge K\phi_1 \cdots \wedge K\phi_n \wedge K\psi)$ is valid in the class of all reflexive models, for $\phi \in \mathcal{L}_P$.

4. $K\neg\neg\phi \wedge K(\phi \rightarrow \psi) \rightarrow \langle DNE\rangle\langle MP\rangle\langle CI\rangle K(\phi \wedge \psi)$ is valid in the class of all reflexive models.

Proof.

1. Let arbitrary reflexive model M and world $w \in W^P$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n K\phi$. Then, there is a model M^* with $f^*(w) = c^* \in C(\{v^{R_1,\ldots,R_n} \mid v \in f(w)\})$ such that $M^*, w \vDash K\phi$. As a result, $M^*, u \vDash \phi$ for all $u \in c^*$ [1]. Due to reflexivity, $w \in f(w)$ and since w is a possible world, any expansion of it would amount to itself, i.e. $w^{R_1,\ldots,R_n} = \{w\}$. As a result, in any case, w is contained in the choice c^*, that is $w \in f^*(w)$. Then, from [1] we conclude $M^*, w \vDash \phi$. Finally since $\phi \in \mathcal{L}_P$ and M^* deviates from M only in terms of the accessibility relation, $M, w \vDash \phi$.

2. Let arbitrary model M and world $w \in W^P$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n K\phi \wedge \langle\dagger\rangle^m K\psi$. Then $M, w \vDash \langle\ddagger\rangle^n K\phi$ and $M, w \vDash \langle\dagger\rangle^m K\psi$. That is, there is a model M^* with $f^*(w) = c^*$ for some choice $c^* \in C(\{v^{R_1,\ldots,R_n} \mid v \in f(w)\})$ such that $M^*, w \vDash K\phi$ and there is M^{**} with $f^{**}(w) = c^{**}$ for some $c^{**} \in C(\{v^{R'_1\ldots R'_m} \mid v \in f(w)\})$ such that $M^{**}, w \vDash K\psi$. This means that $M^*, u \vDash \phi$ for all $u \in c^*$ and $M^{**}, v \vDash \psi$ for all $v \in c^*$. But then by (Monotonicity), there is a choice $c^{***} \in C(\{v^{R_1,\ldots,R_n,R'_1,\ldots,R'_m} \mid v \in f(w)\})$ such that if M^{***} has f^{***} where $f^{***}(w) = c^{***}$, then $M^{***}, z \vDash \phi$ and $M^{***}, z \vDash \psi$ for all $z \in c^{***}$. This amounts to $M^{***}, w \vDash K\phi \wedge K\psi$. Since $f^{***} \in \mathcal{F}^{R_1,\ldots,R_n,R'_1,\ldots,R'_m}(M, w)$, $(M, w) \sim^{R_1,\ldots,R_n,R'_1,\ldots,R'_m} (M^{***}, w)$. So for some model (M^{***}, w): $(M, w) \sim^{R_1,\ldots,R_n,R'_1,\ldots,R'_m} (M^{***}, w)$: $M^{***}, w \vDash K\phi \wedge K\psi$, hence $M, w \vDash \langle\ddagger\rangle^n\langle\dagger\rangle^m(K\phi \wedge K\psi)$, as desired.

3. Let arbitrary model M and world $w \in W^P$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^n(\phi \wedge K\phi_1 \ldots \wedge K\phi_n)$ [1] and $\{\phi_1,\ldots,\phi_n\} \vdash_{R_k} \psi$ [2]. We want to show that $M, w \vDash \langle\ddagger\rangle^n\langle R_k\rangle(\phi \wedge K\phi_1 \ldots \wedge K\phi_n \wedge K\psi)$. From [1]: there is a model M^* with $f^*(w) = c^*$ for some $c^* \in C(\{v^{R_1,\ldots,R_n} \mid v \in f(w)\})$ such that $M^*, w \vDash (\phi \wedge K\phi_1 \ldots \wedge K\phi_n)$. Then it easily follows that $M^*, u \vDash \phi_1 \wedge \ldots \wedge \phi_n$ for all $u \in c^*$. But from [2], we get that there will be a choice $c^{**} \in C(\{v^{R_1,\ldots,R_n,R_k} \mid v \in f(w)\})$ such that if M^{**} has $f^{**}(w) = c^{**}$ then $M^{**}, u \vDash \psi$ for all $u \in c^{**}$. As a result, $M^{**}, w \vDash K\psi$. By (Monotonicity), $M^{**}, w \vDash K\phi_1 \wedge \ldots \wedge K\phi_n$, too. Finally, since ϕ is propositional, it is not affected by the change of the accessibility relation, therefore $M^{**}, w \vDash \phi$. As a result, $M^{**}, w \vDash (\phi \wedge K\phi_1 \ldots \wedge K\phi_n \wedge K\psi)$. By the construction of M^{**}, $M, w \vDash \langle\ddagger\rangle^n\langle R_k\rangle(\phi \wedge K\phi_1 \ldots \wedge K\phi_n \wedge K\psi)$, as desired.

4. Let arbitrary model M and world $w \in W$ of the model. Suppose $M, w \vDash K\neg\neg\phi \wedge K(\phi \rightarrow \psi)$. We also have that $\{\neg\neg\phi\} \vdash_{DNE} \phi$, $\{\phi, \phi \rightarrow \psi\} \vdash_{MP} \psi$, $\{\phi, \psi\} \vdash_{CI} \phi \wedge \psi$. We get that $M, w \vDash \langle DNE\rangle K\phi$ as a corollary from the previous validity. By this result, (Monotonicity) and assumption we get that $M, w \vDash \langle DNE\rangle(K\phi \wedge K(\phi \rightarrow \psi))$. Then again as a corollary of the previous validity, we obtain $M, w \vDash \langle DNE\rangle\langle MP\rangle K\psi$. By this, (Monotonicity) and assumption, $M, w \vDash \langle DNE\rangle\langle MP\rangle(K\psi \wedge K\phi)$. A final application of the previous validity results in $M, w \vDash \langle DNE\rangle\langle MP\rangle\langle CI\rangle K(\phi \wedge \psi)$, as desired.

□

The second validity and the fourth validity of Theorem 4.1.6 and Theorem 4.1.1 allow us to spot an initial convergence between Rasmussen and the two semantic interpretations (RW and IW). This can be said regarding the first validity as well; imposing reflexivity under the second approach should not be thought as a radical intervention that hints at some kind of substantial divergence between the core of the two interpretations. On the contrary, reflexivity is the standard veridicality-preserving condition in frameworks that employ accessibility relations. As these were simply absent in the first approach, no reflexivity condition was needed to ensure veridicality. In fact, this was accommodated via condition 3 of the class **M**. Interestingly, $\langle\ddagger\rangle^n\phi \wedge \langle\dagger\rangle^m\psi \rightarrow \langle\ddagger\rangle^n\langle\dagger\rangle^m(\phi\wedge\psi)$ fails to be valid both under RW-semantics and IW-semantics. The fact that these results coincide, along with the counterexamples provided, can be seen as indicative of a problematic motivation behind the inclusion of the analogous axiom in *Rasmussen*. The similarity extends further; consider the limited version of (R_D) as in Theorem 4.1.2 and Theorem 4.1.7. We can thus speculate that what was called (R_D)-*Success* does not really deliver the expected result.

The way out of the problem of logical omniscience under this new, full framework can be additionally reflected on the next theorem:

Theorem 4.1.8 (Reasoning from rules)**.** If $\{\phi_1,\ldots,\phi_k\} \vdash_{R_1,\ldots,R_n} \psi$ and $\langle\ddagger\rangle^{m_i}K\phi_i$ for $1 \leq i \leq k$, where each $\langle\ddagger\rangle^{m_i}$ is a sequence of m_i-many inference rules, then $\langle\ddagger\rangle^{m_1}\ldots\langle\ddagger\rangle^{m_k}\langle\ddagger\rangle^n K\psi$.

Proof. Let arbitrary model M and world $w \in W^P$ of the model. Suppose $M, w \vDash \langle\ddagger\rangle^{m_i}K\phi_i$, for $1 \leq i \leq k$. As a result, for each i, there is M_i^* with $f_i^*(w) = c_i^*$ for some $c_i^* \in C(\{v^{R_{i1},\ldots,R_{im_i}} \mid v \in f(w)\})$ such that $M_i^*, w \vDash K\phi_i$. That is, for all $v \in f_i^*(w)$, $M_i^*, v \vDash \phi_i$. Due to (Monotonicity), there is $c^* \in C(\{v^{R_{ij}} \mid v \in f(w)\})$ with R_{ij} abbreviating $R_{11},\ldots,R_{1m_1},\ldots,R_{k1},\ldots,R_{km_k}$ such that if a model M^* has $f^*(w) = c^*$, then $M^*, u \vDash \phi_i$ for all $u \in c^*$. Since $\{\phi_1,\ldots,\phi_k\} \vdash_{R_1,\ldots,R_n} \psi$, there will be $c^{\circledast} \in C(\{v^{R_{ij},R_1,\ldots,R_n} \mid v \in f(w)\})$ such that if a model $M^{\circledast}$ has $f^{\circledast}(w) = c^{\circledast}$, then $M^{\circledast}, u \vDash \psi$ for every $u \in c^{\circledast}$. This immediately results in $M^{\circledast}, w \vDash K\psi$. But then due to construction of $M^{\circledast}$, $M, w \vDash \langle\ddagger\rangle^{m_1}\ldots\langle\ddagger\rangle^{m_k}\langle\ddagger\rangle^n K\psi$. □

This second approach naturally extends both *Rasmussen* and *Rasmussen&Bjerring*. It extends Rasmussen by importing semantics, thereby offering a concrete insight on the adequacy of the axioms. It extends *Rasmussen&Bjerring* by unraveling the reasoning process and formally introducing the effect of cognitive effort. These results enable us to overcome the problem of logical omniscience, while still preserving an account of how reasoning ensures that we can perform inferences lying within suitable applications of rules. In particular, the argument of impossible worlds suffices to invalidate the closure principles: since the clause for knowledge quantifies over both possible and impossible worlds, it can easily be the case that an agent knows ϕ but not ψ, even though the latter is logically entailed by the former. In addition, the truth clause for $\langle R_k\rangle K\phi$, and subsequently Theorem 4.1.8, demonstrate that an agent can only come to know ψ via suitable applications of inference rules. This also manifests how logical competence is preserved.

In fact, the rule-sensitivity, the measure on cognitive capacity and the way it is updated allow us to practically witness to which extent reasoning evolves and thus to which extent consequences of what we already know can come to be in turn known. Besides, running out of resources depends not only on the number but also on the kind and chronology of rules. The IW-approach takes these factors into account and can therefore explain that the agent exhausts her resources while reasoning. Overall, it does justice to the idea that consumption of specific resources is responsible for the failure to acquire knowledge from some point onwards.

In the next section, we will actually incorporate this explanation *into* the logical framework itself.

Before delving into the anticipated extension of IW, it is useful to make some comparative remarks between our two approaches. As we have already observed, despite the use of different equipment, the two views converge substantially in (in)validating certain suggested axioms. This serves as a strong indication on the adequacy of the axioms and as a criterion on what actually is credible enough to formally describe a reasoning process. Yet it is straightforward to see that RW-semantics is more syntactically inclined; worlds essentially constitute a convenient way to talk about (factive) truths and epistemic states in tandem with transitions among the latter, triggered by reasoning steps. The second view is still committed to the mainstream (im)possible worlds semantics, only now augmented by dynamic clauses. More specifically, it still employs accessibility relations and model transformations (in the sense of DEL) to account for the progress of reasoning, taking note of, among others, the agent's cognitive effort. The first attempt eluded the need of model change due to its "temporal" portrayal of reasoning and the way the set of rule-based worlds got *itself* structured in accordance with the inference rules. This is why we observed that differences in the obtained validities can be attributed to technical ramifications; they do not really affect the gist of the line of argumentation on escaping the problem – besides, both approaches provide the semantics underlying the very same conception. An important difference that will manifest itself more clearly in what follows is that the second approach offers the breeding ground to talk about resource depletion in a quantitative manner and adopt mainstream techniques regarding external information. To summarize, both approaches strike the balance between non-omniscience and ignorance and manage to do so by acknowledging that the progress of reasoning is the distinguishing factor. Capturing this characteristic, albeit in different ways, is what increases their explanatory power and successfully addresses the criticisms against existing attempts.

A quantitative extension of IW

This section elucidates the idea of resource consumption, incorporated into the logical language and the IW-semantics. The goal is to show that a suitable quantitative fragment, once appended to these basic elements, yields a thorough account of how real-life agents expand their epistemic states, given certain preconditions, and why this eventually stops. By establishing a concrete connection between resource consumption and the evolution of reasoning, we wish to soothe the objections about the broad use of n reasoning steps, as discussed in Section 3.3.3. Instead of relying on some vague impression of some reasoning steps as "too many", one may now gain insight on the agent's inferential ability, by looking at "snapshots" of the available resources following the applications of inference rules.

To begin with, the new language includes an additional component (along the lines of Fagin and Halpern (1994)):

Definition 4.1.15 (Language $\mathcal{L}_{\text{IWe}}$)**.** Let Φ be the standard set of propositional atoms and R, I countable sets. The language $\mathcal{L}_{\text{IWe}}$ consists of a set T of constants and a set F of formulas. In particular, $T = \{c_{R_k} \mid R_k \in R\} \cup \{cp_i \mid i \in I\}$. Then, the set of formulas is defined inductively as follows:

$$\phi \quad ::= \quad p \quad | \quad s \geq t \quad | \quad \neg\phi \quad | \quad \phi \wedge \phi \quad | \quad K\chi \quad | \quad \langle R_k \rangle \phi \quad | \quad [R_k]\phi$$

where $s, t \in T$, and as before $p \in \Phi$, $\chi \in \mathcal{L}_P$ and $R_k \in R$.

The intended reading of the constants is that those of the form c_{R_k} correspond to the cognitive costs of inference rules whereas those of the form cp_i correspond to the agent's cognitive capacity at the several stages of reasoning. The inequalities are then introduced to express the comparisons between costs and capacity.

Definition 4.1.16 (Interpretation of constants)**.** Given a model $M = \langle W^P, W^I, f, V, R, Res, C, cp \rangle$, the constants of T are interpreted as follows: $cp_i^M = cp$ and $c_{R_k}^M = C(R_k)$.

Leaving aside the logical language, it is important to clarify our intended reading of $\geq$. Keep in mind that, given M, both cp and $C(R_k)$ are ordered r-tuples. Then for α, β such tuples, our reading of $\alpha \geq \beta$ is that *every* i-th component of α is greater or equal than the i-th component of β [12]. Now, the motivation behind the new clauses is that an application of an inference rule should be "affordable"; the agent's cognitive capacity must be enough to "endure" the resource consumption caused whenever some inference rule is fired. Otherwise, the action of applying the inference rule is not executable. In short, the agent's capacity with regard to each resource should be greater or equal than the cost of the rule with regard to the respective resource.

With our new extended language $\mathcal{L}_{\text{IWe}}$, we can formally express this requirement.

Definition 4.1.17 (IWe-semantics)**.** For $w \in W^P$:

1. $M, w \vDash p$ if and only if $p \in V(w)$, where $p \in \Phi$.
2. $M, w \vDash s \geq t$ if and only if $s^M \geq t^M$.
3. $M, w \vDash \neg\phi$ if and only if $M, w \nvDash \phi$.
4. $M, w \vDash \phi \wedge \psi$ if and only if $M, w \vDash \phi$ and $M, w \vDash \psi$.
5. $M, w \vDash K\phi$ if and only if $M, w' \vDash \phi$ for all $w' \in f(w)$.
6. $M, w \vDash \langle R_k \rangle \phi$ if and only if $M, w \vDash (cp_i \geq c_{R_k})$ and $M', w' \vDash \phi$ for some $(M', w') : (M, w) \sim^{R_k} (M', w')$.
7. $M, w \vDash [R_k]\phi$ if and only if $M, w \vDash (cp_i \geq c_{R_k})$ implies $M', w' \vDash \phi$ for all (M', w') : $(M, w) \sim^{R_k} (M', w')$.
8. $M, w \dashv \phi$ if and only if $M, w \nvDash \phi$.

For $w \in W^I$:

1. $M, w \vDash \phi$ if and only if $\phi \in V(w)$.
2. $M, w \dashv \phi$ if and only if $\neg\phi \in V(w)$.

It is important to mention that viewing values of resources as natural numbers is not the only option. While it does seem as an adequate way to express the availability and cost of –for instance– *time*, other resources might be better captured via other quantitative assignments. The modelling choices behind $C(R_k)$ and cp and the reading of $\geq$ are mostly indicative of a simple way to formalize the idea of resource consumption, that bounds real agents' performance. Of course, it is empirical research that sheds light on the units that best describe resources, the values corresponding to each inference rule etc. Once we have the tools

[12] This can be easily expressed with the help of projection functions $proj_i$ for $i \in r$. Each $proj_i$ maps a tuple α to its i-th component.

to formalize the core idea, it is possible to manipulate the relevant components and fit other modelling choices. What we have showed is that it is altogether achievable to spell out the cognitive effort under such step-wise proposals, without claiming that our own numerical way of evaluating resources must monopolize this line of investigation.

Next, we give an example to illustrate the subtleties of this new extension, and then point towards a potential application.

Example 2.

Recall Theorem 4.1.7. We will now construct a model that demonstrates a failure to apply Conjunction Introduction, following any application of Double Negation Elimination and Modus Ponens, attributed to the fact that time expires before the agent manages to apply this last step.

- Let model $M = \langle W^P, W^I, f, V, R, Res, C, cp\rangle$ with $R = \{DNE, MP, CI\}$, $Res = \{time, memory\}$ $C(MP) = CI = (2,2), C(DNE) = (3,1)$ while $cp = (5,10)$. In addition, suppose that for world $w \in W^P$: $M, w \vDash K\neg\neg\phi \wedge K(\phi \to \psi)$.
- Then, $M, u \vDash \neg\neg\phi$ and $M, u \vDash \phi \to \psi$ for all $u \in f(w)$. First, we get $M, w \vDash cp_i \geq c_{DNE}$. Second, there is a choice $c' \in C(\{u^{DNE} \mid u \in f(w)\})$ such that if $f'(w) = c'$ and $cp' = cp - C(DNE) = (2,9)$ for a model M', then $M', w \vDash K\phi$.
- Following the same procedure for MP: $M', w \vDash cp_i \geq c_{MP}$ and there is a choice $c'' \in C(\{v^{MP} \mid v \in f'(w)\})$ such that if $f''(w) = c''$ and $cp'' = cp' - C(MP) = (2,9) - (2,2) = (0,7)$ for a model M'', then $M'', w \vDash K\psi$.
- But then $M'', w \nvDash cp_i \geq c_{CI}$.

  So finally, $M'', w \nvDash \langle CI\rangle K(\phi \wedge \psi)$, therefore $M'', w \vDash \neg\langle CI\rangle K(\phi \wedge \psi)$. But this means that $M', w \vDash \langle MP\rangle\neg\langle CI\rangle K(\phi \wedge \psi)$.

  In turn $M, w \vDash \langle DNE\rangle\langle MP\rangle\neg\langle CI\rangle K(\phi \wedge \psi)$.

  As a result, indeed $M, w \nvDash [DNE][MP]\langle CI\rangle K(\phi \wedge \psi)$.

Remark:(Doxastic setting and the challenge of implicit contradictions). Unsurprisingly, the doxastic analogue of the system hitherto discussed can be obtained by adjusting the constructions of Definition 4.1.9, Definition 4.1.10, Definition 4.1.11, Definition 4.1.12, Definition 4.1.13 and Definition 4.1.14, now based on an initial *doxastic* accessibility function f. However, it is worth emphasizing some other potential changes, that better embody the notion of belief and differentiate the propositional attitudes. For instance, we hinted at the suitability of reflexive models to grasp properties of knowledge, such as veridicality. It is natural to drop this requirement in a doxastic setting. Most importantly, in Section 3.3.3, we underlined the insufficiency of Rasmussen's doxastic framework to accommodate belief in implicit contradictions, an unwelcome side-effect of (Monotonicity), (Minimal Consistency) and the way the updated model of Definition 3.3.16 is defined. Should we stick to a purely epistemic framework, we actually escape this worry, given the veridicality of knowledge. The objection persists with belief, though; a real-life, fallible agent might believe an implicit contradiction, therefore our doxastic alternative should do justice to such cases too.

The solution to this sub-task, as given below, focuses on *memory*, a resource that always seems a plausible candidate-element of the set *Res* and crucial for failures such as belief in implicit contradictions[13]. It is natural to assume that the more information stored, the larger the chances of ending up believing something that contradicts an existing belief. As a result, with memory wearing out, (Monotonicity) may be dropped.

Of course, the extension IWe puts forward an extra constraint that disallows action once at least one resource is completely exhausted. Therefore, dropping (Monotonicity) takes place at some intermediate point. Given that the agent's capacity regarding memory gets its maximum value at the initial model, and is completely depleted after some updates, then (Monotonicity) must be dropped at some prior point. The idea is that before the resource is totally exhausted, the increasing burden forces the agent's performance to gradually decline. This can be even formally expressed by equipping our model M with a threshold for this "intermediate" failure and comparing it with the value of the agent's memory capacity. More specifically, the latter may be formalized and interpreted by (a) introducing a new kind of constants and (b) assigning the appropriate projection of the ordered pair *cp* to each such constant, as interpretation at the model M. This is then compared to the value of the threshold. If the memory capacity has fallen below the acceptable limit, then (Monotonicity) is dropped and belief in implicit contradictions is not in principle disallowed.

4.2 Impossible worlds and public announcements (IWPA)

Knowledge is acquired not only because of internal mental processes, but also due to external information. For example, truthful public announcements supply additional pieces of information and these are actually obtained for "free", in terms of cognitive effort. We therefore enrich the repertoire of actions to account for these cases. Truthful public announcements and their effect on our models are hereafter examined:

Definition 4.2.1 (Language $\mathcal{L}_{\text{IWPA}}$)**.** The language $\mathcal{L}_{\text{IWPA}}$ is defined inductively from a countable set of propositional atoms Φ:

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi\wedge\phi \quad | \quad K\psi \quad | \quad \langle R_k\rangle\phi \quad | \quad [R_k]\phi \quad | \quad [\psi!]\phi$$

with $p \in \Phi$, $\psi \in \mathcal{L}_P$, where $\mathcal{L}_P$ is the usual propositional language[14], and $R_k \in R$ for a given countable set R.

Then, $[\psi!]\phi$ reads "after the public announcement of ψ, ϕ is true", all else being as in Definition 4.1.9. We extend our semantics to account for the truth of such sentences. To do so, the following definition (along the lines of Section 2.1) is needed to establish the change generated by the announcement.

Definition 4.2.2 (IW-model transformation by public announcement)**.** Given an IW-model $M = \langle W^P, W^I, f, V, R, Res, C, cp\rangle$, its transformation by $\psi!$ is
$M^{\psi!} = \langle (W^P)^{\psi!}, (W^I)^{\psi!}, f^{\psi!}, V^{\psi!}, R^{\psi!}, Res^{\psi!}, C^{\psi!}, cp^{\psi!}\rangle$ where:

- $(W^P)^{\psi!} = \{w \in W^P \mid M, w \vDash \psi\}$
- $(W^I)^{\psi!} = \{w \in W^I \mid M, w \vDash \psi\}$

[13] Although we do stress the part memory plays, the solution can be adapted for other resources too. In this sense, the sketch is independent of the choice of resource and employing memory is mostly due to illustrative purposes.

[14] As we have so far focused on the agent's knowledge about the world, rather than, for example, higher-order knowledge or knowledge of her own reasoning processes, a restriction on announcements seems reasonable.

- $f^{\psi!}(w) = f(w) \cap W^{\psi!}$, for $w \in W^{\psi!} := (W^P)^{\psi!} \cup (W^I)^{\psi!}$
- $V^{\psi!}(w) = V(w)$, for $w \in W^{\psi!}$
- $R^{\psi!} = R$
- $Res^{\psi!} = Res$
- $C^{\psi!} = C$
- $cp^{\psi!} = cp$

Then:

Definition 4.2.3 (IWPA-semantics)**.**

1. The clauses of Definition 4.1.14.
2. For $w \in W^P$: $M, w \models [\psi!]\phi$ if and only if $M, w \not\models \psi$ or $M^{\psi!}, w \models \phi$.

A full-scale analysis of a real agent's reasoning should encompass the effect of both external and internal information as well as of their combination. Theorem 4.1.8 provided us with a result indicating how purely internal reasoning progresses by drawing on successive applications of inference rules. The next goal is to show how external information, in the form of public announcements, can enhance our epistemic state. Finally, we combine these to account for the result of both applying inference rules and utilizing interaction, as usually is the case in the real world.

Theorem 4.2.1 (Reasoning from announcements)**.** If $\{\phi_1, \ldots, \phi_k\} \vdash_{R_1,\ldots,R_n} \chi$ and $[\psi!]K\phi_i$ for $1 \leq i \leq k$, then $[\psi!]\langle\ddagger\rangle^n K\chi$.

Proof. Let arbitrary model M and world $w \in W^P$ of the model. Suppose that $M, w \models [\psi!]K\phi_i$ for all $1 \leq i \leq k$. Then, it follows from Definition 4.2.3 that $M, w \models \psi$ implies $M^{\psi!}, w \models K\phi_i$ for all $1 \leq i \leq k$. Suppose $M, w \models \psi$. Then $M^{\psi!}, u \models \phi_i$ for all $1 \leq i \leq k$ and all $u \in f^{\psi!}(w)$. But since $\{\phi_1, \ldots, \phi_k\} \vdash_{R_1,\ldots,R_n} \chi$, there is a choice $c^* \in C(\{u^{R_1,\ldots,R_n} \mid u \in f^{\psi!}(w)\})$ such that if $f^*(w) = c^*$ for a model M^* then $M^*, v \models \chi$ for all $v \in c^*$. Therefore, $M^*, w \models K\chi$. As a result, if $M, w \models \psi$ then $M^{\psi!}, w \models \langle\ddagger\rangle^n K\chi$. Overall, $M, w \models [\psi!]\langle\ddagger\rangle^n K\chi$, as desired. □

Theorem 4.2.2 (Reasoning from announcements and rules)**.**

1. If $\{\phi_1, \phi_2\} \vdash_{R_1,\ldots,R_n} \chi$ and $[\psi!](K\phi_1 \wedge \langle\ddagger\rangle^m K\phi_2)$, then $[\psi!]\langle\ddagger\rangle^m\langle\dagger\rangle^n K\chi$.
2. If $\{\phi_1, \phi_2\} \vdash_{R_1,\ldots,R_n} \chi$ and $\langle\ddagger\rangle^m(K\phi_1 \wedge [\psi!]K\phi_2)$, then $\langle\ddagger\rangle^m[\psi!]\langle\dagger\rangle^n K\chi$.

Proof.

1. Let arbitrary model M and world $w \in W^P$ of the model. Suppose that $M, w \models [\psi!](K\phi_1 \wedge \langle\ddagger\rangle^m K\phi_2)$. Then $M, w \models \psi$ implies $M^{\psi!}, w \models (K\phi_1 \wedge \langle\ddagger\rangle^m K\phi_2)$, i.e. $M^{\psi!}, w \models K\phi_1$ and $M^{\psi!}, w \models \langle\ddagger\rangle^m K\phi_2$. That is, $M^{\psi!}, u \models \phi_1$ for all $u \in f^{\psi!}(w)$. Additionally, there is model M^* such that $f^*(w) = c^*$ for some $c^* \in C(\{v^{R_1,\ldots,R_m} \mid v \in f^{\psi!}(w)\})$ and $M^*, u \models \phi_2$ for all $u \in c^*$. Due to (Monotonicity), for all $u \in c^*$, $M^*, u \models \phi_1 \wedge \phi_2$. Since $\{\phi_1, \phi_2\} \vdash_{R_1,\ldots,R_n} \chi$, there is a choice $c^{**} \in C(\{v^{R_1,\ldots,R_n} \mid v \in f^*(w)\})$ such that if $f^{**}(w) = c^{**}$ for a model M^{**} then $M^{**}, u \models \chi$ for all $u \in c^{**}$. But then $M^{**}, w \models K\chi$, and in turn $M^*, w \models \langle\dagger\rangle^n K\chi$. Likewise $M^{\psi!}, w \models \langle\ddagger\rangle^m\langle\dagger\rangle^n K\chi$. Overall, $M, w \models \psi$ implies $M^{\psi!}, w \models \langle\ddagger\rangle^m\langle\dagger\rangle^n K\chi$, so we finally conclude $M, w \models [\psi!]\langle\ddagger\rangle^m\langle\dagger\rangle^n K\chi$.

2. Let arbitrary model M and world $w \in W^P$ of the model. Suppose that $M, w \vDash \langle \ddagger \rangle^m (K\phi_1 \wedge [\psi!]K\phi_2)$. So there is model M^* such that $f^*(w) = c^*$ for some $c^* \in C(\{v^{R_1,\ldots,R_m} \mid v \in f(w)\})$ and $M^*, w \vDash K\phi_1 \wedge [\psi!]K\phi_2$, i.e. $M^*, w \vDash K\phi_1$ and if $M^*, w \vDash \psi$ then $(M^*)^{\psi!}, w \vDash K\phi_2$. It then follows that $M^*, w \vDash \psi$ implies $(M^*)^{\psi!}, w \vDash (K\phi_1 \wedge K\phi_2)$, i.e. for all $u \in (f^*)^{\psi!}$, we get $(M^*)^{\psi!}, u \vDash \phi_1 \wedge \phi_2$. Since $\{\phi_1, \phi_2\} \vdash_{R_1,\ldots,R_n} \chi$, there is a choice $c^{**} \in C(\{v^{R_1,\ldots,R_n} \mid v \in (f^*)^{\psi!}(w)\})$ such that if $f^{**}(w) = c^{**}$ for a model M^{**} then $M^{**}, v \vDash \chi$ for all $v \in c^{**}$. But then $M^{**}, w \vDash K\chi$ and in turn $(M^*)^{\psi!}, w \vDash \langle \dagger \rangle^n K\chi$. Therefore, $M^*, w \vDash \psi$ implies $(M^*)^{\psi!}, w \vDash \langle \dagger \rangle^n K\chi$, that is $M^*, w \vDash [\psi!]\langle \dagger \rangle^n K\chi$. So finally, $M, w \vDash \langle \ddagger \rangle^m [\psi!]\langle \dagger \rangle^n K\chi$.

□

The foregoing can be accordingly generalized for more announcements, applications of rules and thus number of premises. Theorem 4.2.2 also exemplifies the order-sensitivity of a reasoning orocess that is assisted by a combination of external and internal tools. The explanatory power of the theorems is also evident in the *restaurant scenario*:

Example 3.
Consider the restaurant scenario from van Benthem (2008):

> You are in a restaurant with your parents, and you have ordered three dishes: Fish, Meat, and Vegetarian. Now a new waiter comes back from the kitchen with three dishes. What will happen?

Drawing on our experience, we expect that the waiter only needs two announcements and one inference to distribute the dishes correctly. We use atoms of the form f_i $(i = 1,2,3)$ to express “fish to person i” and likewise for atoms m_i and v_i. Let R be the set of rules containing CI and MP. We can now see the contribution of our constructions in sketching the waiter's reasoning. For instance, assume that $[v_1!][f_2!](Kv_1 \wedge Kf_2)$ and $K((v_1 \wedge f_2) \rightarrow m_3)$. Then, since $\{v_1, f_2\} \vdash_{CI} v_1 \wedge f_2$, following the method of Theorem 4.2.1 we obtain $[v_1!][f_2!]\langle CI \rangle K(v_1 \wedge f_2)$. Next, due to $\{v_1 \wedge f_2 \rightarrow m_3, v_1 \wedge f_2\} \vdash_{MP} m_3$ and the methods of Theorem 4.2.2, we obtain $[v_1!][f_2!]\langle CI \rangle \langle MP \rangle Km_3$, as expected.

On a final note, the construction of a doxastic counterpart of this new system should comply with the considerations of the Remark. An extension of the language and the model to formalize the idea of resource depletion follows directly from Definition 4.1.15, Definition 4.1.16 and Definition 4.1.17.

4.3 Impossible worlds and plausibility (IWp)

Until now, the only tools that an agent can employ are inference rules and public announcements. Still, there are other actions influencing the progress of reasoning. Merging our attempt with the considerations of Chapter 2 allows us to bring several grades of knowledge and belief under the same roof, and include events of incoming information, softer than the public announcement. To that end, we combine the plausibility models and IWPA in a new setting, called IWp. In particular, we will now develop a plausibility-sensitive counterpart of the previous IWPA setting and incorporate the action $\phi \Uparrow$ of (radically) *upgrading* with ϕ.

Definition 4.3.1 (Language $\mathcal{L}_{IWp}$)**.** The language $\mathcal{L}_{\text{IWp}}$ is defined inductively from a countable set of propositional atoms Φ:

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi \wedge \phi \quad | \quad K\psi \quad | \quad B\psi \quad | \quad \Box\psi \quad | \quad \langle R_k \rangle \phi \quad | \quad [\psi!]\phi \quad | \quad [\psi \Uparrow]\phi$$

with $p \in \Phi$, $\psi \in \mathcal{L}_P$, where $\mathcal{L}_P$ is the propositional language, and $R_k \in R$ for a given countable set R.

The intended reading of $[\psi \Uparrow]\phi$ is "after (radically) upgrading with ψ, ϕ is true", while B and $\Box$ correspond to belief and defeasible knowledge respectively.

Next, we build our own IWp-plausibility models. Unlike the standard plausibility models, we use a mapping to the class of ordinals Ω to derive the plausibility ordering, inspired by Spohn (1988). This is instrumental for a sharp account on model-transformation induced by applications of inference rules.

Definition 4.3.2 (IWp-plausibility model)**.** An IWp-*plausibility model* is a tuple $M = \langle W^P, W^I, ord, V, R, Res, C, cp \rangle$ where:

- W^P, W^I are countable non-empty sets of possible and impossible worlds respectively.
- ord is a function from $W := (W^P \cup W^I)$ to the class of ordinals Ω such that an ordinal number is assigned to each world. Intuitively, the smallest this ordinal is, the more plausible the world.
- V, R, Res, C, cp are as in Definition 4.1.10.

We can see that the function suffices to extract a plausibility ordering in the usual sense. Specifically, for $w, u \in W$: $w \geq u$ if and only if $ord(w) \geq ord(u)$. Hence, indeed, the ranking of worlds is reflected upon the ordering of ordinals. Based on the definition, it is easy to verify that the intended reading is "w is no more plausible than u". In addition, reflexivity, transitivity, connectedness and converse well-foundedness directly follow from this sort of definition. Finally, we can retrieve the usual equivalence relation $\sim$, representing epistemic indistinguishability, as follows: $w \sim u$ if and only if either $w \geq u$ or $u \geq w$. Furthermore, the following abbreviations are still used (a) $C_{\geq}(w) = \{u \in W \mid w \geq u \text{ or } u \geq w\}$ (which amounts to W, given the properties of $\geq$),and (b)$min(C_{\geq}(w)) = \{u \in C_{\geq}(w) \mid \text{ for every } v \in C_{\geq}(w) : v \geq u\}$.

Once the model is devised, the wider project is to fully break down a reasoning process that involves (i) applications of inference rules, (ii) public announcements (as actions of hard information), and (iii) upgrades (as actions of softer information). While the effect of (ii) and (iii) may be captured based on the literature of Chapter 2, this is not the case for (i), whose account we have thus far explored only under the relational scope. In particular, the challenge lies in unveiling what $\langle R_i \rangle \phi$ means in terms of plausibility, and ensuring that this interpretation does not inhibit our familiar modelling of (ii) and (iii). Essentially, to bring together the impossible-worlds, rule-sensitive setting and its perks (avoiding logical omniscience while still securing that agents are non-ignorant) and the rich insights provided by plausibility models, we have to revisit the notion of rule-specific accessibility (as presented in Definition 4.1.13).

First, we give the account of model transformation regarding (ii) and (iii):

Definition 4.3.3 (IWp-plausibility model transformation by public announcement)**.** Given an IWp-model $M = \langle W^P, W^I, ord, V, R, Res, C, cp \rangle$, its transformation by $\psi!$ is a model $M^{\psi!} = \langle (W^P)^{\psi!}, (W^I)^{\psi!}, ord^{\psi!}, V^{\psi!}, R^{\psi!}, Res^{\psi!}, C^{\psi!}, cp^{\psi!} \rangle$ where:

- $(W^P)^{\psi!} = \{w \in W^P \mid M, w \vDash \psi\}$
- $(W^I)^{\psi!} = \{w \in W^I \mid M, w \vDash \psi\}$
- $ord^{\psi!}(w) = ord(w)$, for $w \in W^{\psi!} := (W^P)^{\psi!} \cup (W^I)^{\psi!}$
- $V^{\psi!}(w) = V(w)$, for $w \in W^{\psi!}$
- $R^{\psi!} = R$
- $Res^{\psi!} = Res$
- $C^{\psi!} = C$
- $cp^{\psi!} = cp$

Definition 4.3.4 (IWp-plausibility model transformation by radical upgrade)**.** Given an IWp-model $M = \langle W^P, W^I, f, V, R, Res, C, cp\rangle$, its transformation by $\psi \Uparrow$ is a model $M^{\psi\Uparrow} = \langle (W^P)^{\psi\Uparrow}, (W^I)^{\psi\Uparrow}, ord^{\psi\Uparrow}, V^{\psi\Uparrow}, R^{\psi\Uparrow}, Res^{\psi\Uparrow}, C^{\psi\Uparrow}, cp^{\psi\Uparrow}\rangle$ where:

- $(W^P)^{\psi\Uparrow} = W^P$
- $(W^I)^{\psi\Uparrow} = W^I$
- $ord^{\psi\Uparrow}$ can be every function from the set[15] $\{f : W \to \Omega \mid$ for any $w, u \in W^{\psi\Uparrow} : f(w) \geq f(u)$ if and only if $w \geq^{\psi\Uparrow} u\}$
- $V^{\psi\Uparrow} = V$
- $R^{\psi\Uparrow} = R$
- $Res^{\psi\Uparrow} = Res$
- $C^{\psi\Uparrow} = C$
- $cp^{\psi\Uparrow} = cp$

While the definition dealing with public announcements is quite straightforward, an explanation might be useful regarding the transformation by upgrades. More specifically, our interest is to preserve the original intuition of all ψ-worlds becoming more plausible than $\neg\psi$ ones, keeping the previous ordering within the two zones. The characterization via ordinals does not interfere with radical upgrades, but rather, as will see, with inference rules. In all cases, we will not be interested in the assigned number *per se*, but on the action-induced re-arrangement (i.e. plausibility of worlds relative to other worlds). This is why all functions from $\{f : W \to \Omega \mid$ for any $w, u \in W^{\psi\Uparrow} : f(w) \geq f(u)$ if and only if $w \geq^{\psi\Uparrow} u\}$ work for our purposes.

Now, we move to the effect of applying inference rules. The backbone of this new account is that a pointed IWp-plausibility model (M', w') is R_k*-accessible* from a given pointed IWp-plausibility model for an inference rule R_k, whenever the set $P_{\geq}(w) = \{u \in W \mid w \geq u\}$ of worlds at least as plausible as w is replaced by a choice of worlds reachable by an application of R_k from the elements of $P_{\geq}(w)$, while the remaining ordering is accordingly adapted. That is,

[15] To determine $ord^{\psi\Uparrow}$, first consider the relation $\geq$ that can be derived from it. As an auxiliary step take: $\geq^{\psi\Uparrow} = (\geq \cap (W \times [[\psi]])) \cup (\geq \cap ([[\neg\psi]] \times W)) \cup (\sim \cap ([[\neg\psi]] \times [[\psi]]))$, that is our familiar re-arrangement due to the upgrade.

we focus on the more (or equally) plausible worlds, as these would be the prioritized cases whenever one applies an inference rule. From Chapter 2, we know that the set of worlds at least as plausible as w is what determines defeasible knowledge at w. Given that our attention is restricted to what follows from this set (in terms of R_k), one can safely say that applications of inference rules are viewed as potential modifiers of (at least) the agent's defeasible knowledge.

This initial description on the effect of R_k continues as follows: if a world u was considered more (or equally) plausible than w, but after an application of R_k does not "survive" as such, then it is natural to eliminate it. This world must have been an impossible world[16] and now its impossibility has been uncovered; there is no reason to keep it in our set of worlds.

Once we rule out such worlds, if any, the question arises: how are the remaining worlds ordered? A simple response is that the initial ordering should be preserved to the extent that it is unaffected by the application of the inference rule.

To be more precise, we use the ordinal function. Note that the notion of rule-specific radius, introduced in Definition 4.1.11, is adopted as it is. Let $M = \langle W^P, W^I, ord, V, R, Res, C, cp \rangle$ an IWp-plausibility model. We spell out the attempt in steps:

Step 1 Let (M, w) be a pointed model. Then, given an inference rule R_k, let $P^{R_k}(w) := c$ where c is some choice in $C(\{v^{R_k} \mid v \in P_{\geq}(w)\})$. In words, a choice of R_k-expansions of the worlds initially considered at least as plausible as w.

Step 2 Based on the argument used above, if $u \in P_{\geq}(w)$ but $u \notin P^{R_k}(w)$, then u must be excluded from the new model. So in any case, the R_k-accessible pointed model (M', w') should be such that its set of worlds is $W^{R_k} = W \setminus \{u \in W \mid u \in P_{\geq}(w) \setminus P^{R_k}(w)\}$ and $w' = w$. As observed above, this elimination of worlds is in fact an elimination affecting the set W^I.

Step 3 We now develop the new ordering ord^{R_k} following the application of the inference rule. Let $u \in W^{R_k}$:

(a) If $u \notin P_{\geq}(w) \cup P^{R_k}(w)$, then $ord^{R_k}(u) = ord(u)$, i.e. the assigned ranking remains the same, for worlds that were less plausible than w and are not contained in the choice.

(b) Next consider $u \in P^{R_k}(w)$. This means that there is at least one $v \in P_{\geq}(w)$ such that $u \in v^{R_k}$ for the particular choice c that gave rise to $P^{R_k}(w)$. Denote the set of such v's by T. Then $ord^{R_k}(u) = ord(z)$ for $z \in min(T)$. Therefore, if a world is in $P^{R_k}(w)$, then it takes the position of the most plausible of the worlds from which it originated.

Step 4 Finally, for worlds $u, v \in W^{R_k}$: $u \geq^{R_k} v$ if and only if $ord^{R_k}(u) \geq ord^{R_k}(v)$, therefore again all the required properties are preserved.

Step 5 The other components of the model remain unchanged, expect from V which is simply restricted to the worlds in W^{R_k} and $cp^{R_k} := cp - C(R_k)$.

This kind of model-transformation in terms of $\geq$ becomes clearer below.

[16] A possible world will always survive following applications of inference rules, as its radius amounts to itself.

Example 4.

Let model M as in 4.3, with $ord(w_4) = 4, ord(w_3) = 3, ord(w_2) = 2, ord(w_1) = 1, ord(w_0) = 0$. In search of all the ways the pointed model (M, w_2) can change following an application of MP, we follow the procedure sketched above.

Step 1 First, we compute $\{v^{MP} \mid v \in P_{\geq}(w_2)\}$. According to 4.3, it amounts to $\{\{w_2\}, \{w_3, w_2\}, \{w_0\}\}$.

- As a result, $C(\{\{w_2\}, \{w_3, w_2\}, \{w_0\}\}) = \{\{w_2, w_3, w_0\}, \{w_2, w_0\}\}$. So $P^{MP}(w) = \{w_2, w_3, w_0\}$ or $P^{MP}(w) = \{w_2, w_0\}$.

- 1. In case $P^{MP}(w_2) = \{w_2, w_3, w_0\}$:

  Step 2 $W^{MP} = W \setminus \{u \in W \mid u \in \{w_2, w_1, w_0\} \setminus \{w_2, w_3, w_0\}\} = \{w_4, w_3, w_2, w_0\}$.

  Step 3 Since $w_4 \notin P^{MP}(w_2) \cup P_{\geq}(w_2)$, $ord^{MP}(w_4) = ord(w_4) = 4$. Then $w_3 \in P^{MP}(w_2)$ and, checking from which world(s) it originated in the particular choice, we get that $w_3 \in w_1^{MP}$, therefore $ord^{MP}(w_3) = ord(w_1) = 1$. Likewise, $w_2 \in P^{MP}(w_2)$ and $w_2 \in w_2^{MP}$, so $ord^{MP}(w_2) = ord(w_2) = 2$. Finally, $w_0 \in P^{MP}(w_2)$ and $w_0 \in w_0^{MP}$, so $ord^{MP}(w_0) = ord(w_0) = 0$.

  2. In case $P^{MP}(w_2) = \{w_2, w_0\}$:

  Step 2 $W^{MP} = W \setminus \{u \in W \mid u \in \{w_2, w_1, w_0\} \setminus \{w_2, w_0\}\} = \{w_4, w_3, w_2, w_0\}$.

  Step 3 As above, $ord^{MP}(w_4) = ord(w_4) = 4$. For the same reasons, $ord^{MP}(w_3) = ord(w_3) = 3$. Then, $w_2 \in P^{R_k}(w_2)$ and, checking from which world(s) it originated in the particular choice, we find $w_2 \in w_2^{MP}$, $w_2 \in w_1^{MP}$. But $ord(w_2) \geq ord(w_1)$ so $ord^{MP}(w_2) = ord(w_1) = 1$. Again, $w_0 \in P^{MP}(w_2)$ and $w_0 \in w_0^{MP}$, so $ord^{MP}(w_0) = ord(w_0) = 0$.

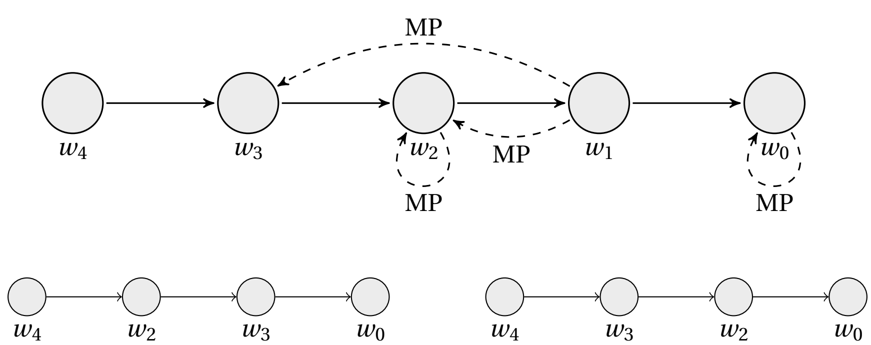


Figure 4.3: The first figure depicts the model M, with an MP-dashed arrow from w to w' denoting that w' is an MP-expansion of w. Then, we obtain two potential transformations of the pointed model (M, w_2), i.e. two MP-accessible pointed models, based on the two ways the set of w_2's more (or equally) plausible worlds can change due to MP.

Overall, we have developed the ways an initial model changes after public announcements, radical upgrades and applications of inference rules. The semantics is finally given by:

Definition 4.3.5 (IWp-semantics)**.** For $w \in W^P$:

- $M, w \models p$ if and only if $p \in V(w)$, where $p \in \Phi$.
- $M, w \models \neg\phi$ if and only if $M, w \not\models \phi$.
- $M, w \models \phi \wedge \psi$ if and only if $M, w \models \phi$ and $M, w \models \psi$.
- $M, w \models K\phi$ if and only if $M, w' \models \phi$ for all $w' \in C_{\geq}(w)$.
- $M, w \models \Box\phi$ if and only if $M, w' \models \phi$ for all $w' \in P_{\geq}(w)$.
- $M, w \models B\phi$ if and only if $M, w' \models \phi$ for all $w' \in min((C_{\geq}(w))$.
- $M, w \models \langle R_k \rangle\phi$ if and only if $M', w' \models \phi$ for some (M', w') which is R_k-accessible from (M, w).
- $M, w \models [\psi!]\phi$ if and only if $M, w \not\models \psi$ or $M^{\psi}, w \models \phi$.
- $M, w \models [\psi \Uparrow]\phi$ if and only if $M^{\Uparrow\psi}, w \models \phi$.
- $M, w \dashv \phi$ if and only if $M, w \not\models \phi$.

For $w \in W^I$:

- $M, w \models \phi$ if and only if $\phi \in V(w)$
- $M, w \dashv \phi$ if and only if $\neg\phi \in V(w)$

Given the distinction between K and $\Box$ in Section 3.2.4, hard attitudes (like K) change only in the face of hard information and of those applications of inference rules that result in world-elimination. Softer attitudes (like $\Box$ and B) might additionally change following radical upgrades.

As a result, we have provided the means to extend the step-wise impossible worlds account, that escapes logical omniscience, whilst illuminating internal mental processes and external information of various sorts. Besides, IWp models and semantics allow us to combine results on epistemic and doxastic notions. We only note that, in accordance with the rich literature of plausibility models, this account can be extended to incorporate more actions and attitudes; consider for example that conditional beliefs of Chapter 2 can be utilized in IWp, too. In this way, we can provide more and more detailed investigations on belief change under omniscience-free, rule-sensitive systems. Finally, a quantitative flavour on the shortage of resources may be added simply by referring back to Section 4.1.2.

4.4 Reducing frameworks with impossible worlds

We have seen how IW and its extensions refined existing frameworks and overcame barriers to a realistic modelling of knowledge and belief. The key factor was the adoption of a *dynamic* framework *with* impossible worlds. Prompted by the work of Wansing (1990), and in search of a (complete) logic for such frameworks, we aim at a reduction to static frameworks involving solely possible worlds. In particular, Wansing showed how various models for knowledge and belief, for instance structures for awareness, can be viewed as Rantala ones that validate precisely the same formulas (given, of course, a fixed background language). This section's goal is to, roughly speaking, explore the other direction. More concretely, we want to show that impossible-worlds frameworks, but now in the sense of *Rasmussen*&*Bjerring*, IW, etc., can be reduced to *awareness-like* ones, thereby opening the way for a wider exploitation of the fruitful techniques and properties of the latter.

This investigation poses no threats for the status of the settings discussed earlier. We view its contribution as largely technical. That is, there is no reason to assume that the introduction of dynamic settings with impossible worlds is redundant in the face of a reduction that resembles awareness structures. As we will see, the components of the reduced models have no (obvious) intuitive reading nor can they possibly explain all the phenomena and ways out of the problem as the previous settings did. In other words, *Rasmussen*&*Bjerring*, IW, IWe, IWPA, IWp, explain properties of real reasoning in a balanced and intuitive way, while any attempt of a reduction merely constitutes a detour to obtain useful results, such as completeness. In addition, it is precisely because of this balance and superiority in terms of explanatory power that criticisms directed against awareness structures cannot be hereby replicated.

In what follows, we present a reduction and its consequences under an epistemic setting for the *Rasmussen*&*Bjerring* view. We also describe how this general method can be used for IW too. We expect that this first attempt provides the building blocks for complete logics under IWe, IWPA and IWp too. For example, once we build a suitable static, possible-worlds framework and a complete logic for it, then it suffices to provide reduction axioms for the additional actions.

1. The first challenge is, as implied above, the appropriate choice of a language that establishes the common ground[17] on which we will show that the reduction is indeed successful – i.e. that the same formulas are valid under the original and the reduced models. Of course, we seek a language that facilitates the crossing from a dynamic and impossible-worlds framework to a static one that only involves possible worlds, while *still* preserving the core of former. To that end:

   Definition 4.4.1 (Language for the reduction)**.** The language $\mathcal{L}_r$ is defined inductively from a countable set of propositional atoms Φ as:

   $$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi\wedge\phi \quad | \quad K\chi \quad | \quad \langle n\rangle\psi \quad | \quad L\phi \quad | \quad I\phi \quad | \quad J_n$$

   with $p \in \Phi$, $\chi \in \mathcal{L}_P$, $n = 0, 1, \ldots$, and ψ a knowledge assertion of the form $K\chi$.

   It is evident that part of the language is essentially an epistemic version of the one found in Section 3.3.3. We only note that we have restricted our attention to operators $\langle n\rangle$ prefixing only knowledge assertions, as the change they induce (on the accessibility function) essentially affects the acquisition of knowledge. Besides what the agent comes to know is the core of our interest, when studying the outcome of a reasoning process. Then, given a model of $\langle W^P, W^I, f, V\rangle$ as in *Rasmussen*&*Bjerring*, all the clauses, other

[17] This was straightforward for the direction of Wansing's work (simply using the language of Section 3.2.1). Since the set of impossible worlds in the induced Rantala model was associated with the awareness function of the original model, the truth clauses could be given without further machinery. In our case this is inevitable due to the dynamic effect of $\langle n\rangle$.

than those for $L\phi$, $I\phi$, J_n with regard to a possible world w[18], are given along the lines of Section 3.3.3). In fact, we only have to account for the additional formulas. Giving their truth clauses also shows the utility of the new operators:

- $M, w \vDash L\phi$ if and only if for all $u \in f(w) \cap W^P$: $M, u \vDash \phi$.
- $M, w \vDash I\phi$ if and only if for all $u \in f(w) \cap W^I$: $M, u \vDash \phi$.
- $M, w \vDash J_n\phi$ if and only if for all $u \in f(w) \cap W^I$ such that $u^n \neq \varnothing$, there is some $v \in u^n$: $M, v \vDash \phi$.

That is, $L\phi$ provides the standard quantification over the possible, epistemically accessible worlds. The point of $I\phi$ is similar, only now isolating the impossible, epistemically accessible words. The distinction is necessary to transform a setting involving both sorts of worlds to one that merely comprises possible worlds. Finally, the truth clause for $J_n\phi$ says: for every world in $f(w) \cap W^I$, if its n-radius is not empty, then there is at least one n-expansion of it validating ϕ. This will be instrumental for the shift from a dynamic to a static interpretation for $\langle n \rangle K\phi$.

2. We can now show that there is an equivalent, static formulation for the truth clause of $\langle n \rangle K\phi$, by utilizing this richer language.

   Lemma 3 (Reducing $\langle n \rangle K\phi$)**.** $\langle n \rangle K\phi$ is logically equivalent to $L\phi \wedge J_n\phi$.

   Proof. Let a model $M = \langle W^P, W^I, f, V \rangle$ and world $w \in W^P$. Suppose $M, w \vDash \langle n \rangle K\phi$. Then there is model $(M', w) \sim^n (M, w)$ such that $M', w \vDash K\phi$. That is, $M', w \vDash K\phi$ for a model M' with $f'(w) = c'$ for some $c \in C(\{v^n \mid v \in f(w)\})$ and, in turn, $M', u \vDash \phi$ for all $u \in c'$ [1]. By definition of choice, c' is just one way in which only one element can be picked from each v^n, for $v \in f(w)$. Because of this, [1] precisely means that for every $v \in f(w)$ whose n-radius v^n is not empty, there must be some $u \in v^n$ such that $M, u \vDash \phi$. Let any such v. If $v \in W^P$, then $v \in f(w) \cap W^P$ and since $v^n = \{v\}$, the previous argument boils down to $M, v \vDash \phi$. Therefore, $M, w \vDash L\phi$. In the case that $v \in W^I$, the clause boils down to $M, w \vDash J_n\phi$, by the way the clause for J_n was constructed. Finally, $M, w \vDash L\phi \wedge J_n\phi$.

   For the other direction, suppose that $M, w \vDash L\phi \wedge J_n\phi$, i.e. $M, w \vDash L\phi$ and $M, w \vDash J_n\phi$. The former gives us that for all $u \in f(w) \cap W^P$: $M, u \vDash \phi$ [1]. The latter gives us that for all $z \in f(w) \cap W^I$, with $z^n \neq \varnothing$ there is some $v \in z^n$: $M, v \vDash \phi$ [2]. Of course, each world is either in W^P or W^I. Then, by [1], [2] and the fact that $u^n = \{u\}$ for $u \in W^P$, we get that there is a choice $c' \in C(\{u^n \mid u \in f(w)\})$ such that if M' has $f'(w) = c'$ then $M', v \vDash \phi$ for all $v \in c'$. Therefore, $M', w \vDash K\phi$. So finally, $M, w \vDash \langle n \rangle K\phi$. □

3. Now we construct the candidate for a reduced model $\mathrm{M} = \langle \mathrm{W}, \mathrm{f}, \mathrm{V}, \mathrm{I}, \mathrm{J}_n \rangle$, given the original model $M = \langle W^P, W^I, f, V \rangle$, where:

   - $\mathrm{W} = W^P$
   - $\mathrm{f}(w) = f(w) \cap \mathrm{W}$, for $w \in \mathrm{W}$

[18] For impossible worlds, the clauses remain the same. Keep in mind that validity is defined in terms of possible worlds only.

- $\mathrm{V}(w) = V(w)$ for $w \in \mathrm{W}$.
- $\mathrm{I} : \mathrm{W} \to \mathcal{P}(\mathcal{L}_r)$ such that $\mathrm{I}(w) = \bigcap_{v \in f(w) \cap W^I} V(v)$. Intuitively, I takes a possible world w and yields the set of those formulas that are true at all impossible worlds accessible from w.
- $\mathrm{J}_n : \mathrm{W} \to \mathcal{P}(\mathcal{L}_r)$ such that $\mathrm{J}_n(w) = \bigcap_{\{v \in f(w) \cap W^I \mid v^n \neq \varnothing\}} \bigcup_{u \in v^n} V^*(u)$. Note that $V^*(u)$ denotes the set of *all* formulas true at u[19]. Intuitively, J_n takes a possible world w and yields the set of formulas that are true at some n-expansion of a world v, for every impossible world v accessible from w, with non-empty n-radius.

The semantics based on M is given as follows:

- $\mathrm{M}, w \vDash p$ if and only if $p \in \mathrm{V}(w)$.
- $\mathrm{M}, w \vDash \neg\phi$ if and only if $\mathrm{M}, w \nvDash \phi$.
- $\mathrm{M}, w \vDash \phi \wedge \psi$ if and only if $\mathrm{M}, w \vDash \phi$ and $\mathrm{M}, w \vDash \psi$.
- $\mathrm{M}, w \vDash L\phi$ if and only if for all $u \in \mathrm{f}(w)$: $\mathrm{M}, u \vDash \phi$.
- $\mathrm{M}, w \vDash I\phi$ if and only if $\phi \in \mathrm{I}(w)$.
- $\mathrm{M}, w \vDash J_n\phi$ if and only if $\phi \in \mathrm{J}_n(w)$.
- $\mathrm{M}, w \vDash K\phi$ if and only if $\mathrm{M}, w \vDash L\phi$ and $\mathrm{M}, w \vDash I\phi$.
- $\mathrm{M}, w \vDash \langle n \rangle K\phi$ if and only if $\mathrm{M}, w \vDash L\phi$ and $\mathrm{M}, w \vDash J_n\phi$.

Let's now show that this candidate M indeed reduces M.

Theorem 4.4.1 (Reduction)**.** Given a model $M = \langle W^P, W^I, f, V \rangle$, construct its (candidate) reduced model M according to the previous definition. Then M is indeed a reduction of M, i.e. for any $w \in W^P$ and formula $\phi \in \mathcal{L}_r$: $M, w \vDash \phi$ if and only if $\mathrm{M}, w \vDash \phi$.

Proof. The proof goes by induction on the complexity of ϕ.

- Base case: for $\phi := p$ we get $M, w \vDash p$ if and only if $p \in V(w)$ if and only if $p \in \mathrm{V}(w)$ if and only if $\mathrm{M}, w \vDash p$.
- For $\phi := \neg\psi$ and Induction Hypothesis that the result holds for ψ. Then $M, w \vDash \neg\psi$ if and only if $M, w \nvDash \psi$ if and only if (by I.H.) $\mathrm{M}, w \nvDash \psi$ if and only if $\mathrm{M}, w \vDash \neg\psi$.
- For $\phi := \chi \wedge \psi$ with Induction Hypothesis that the result holds for χ and ψ. Then $M, w \vDash \chi \wedge \psi$ if and only if $M, w \vDash \chi$ and $M, w \vDash \psi$ if and only if (by I.H.) $\mathrm{M}, w \vDash \chi$ and $\mathrm{M}, w \vDash \psi$ if and only if $\mathrm{M}, w \vDash \chi \wedge \psi$.
- For $\phi := L\psi$ with Induction Hypothesis that the result holds for ψ. Then $M, w \vDash L\psi$ if and only if for all $u \in f(w) \cap W^P$: $M, u \vDash \psi$ if and only if (by I.H.) for all $u \in \mathrm{f}(w)$: $\mathrm{M}, u \vDash \psi$ if and only if $\mathrm{M}, w \vDash L\psi$.
- For $\phi := I\psi$ with Induction Hypothesis that the result holds for ψ. Then $M, w \vDash I\psi$ if and only if for all $u \in f(w) \cap W^I$: $M, u \vDash \psi$ if and only if for all $u \in f(w) \cap W^I$: $\psi \in V(u)$ if and only if $\psi \in \mathrm{I}(w)$ if and only if $\mathrm{M}, w \vDash I\psi$.

[19] While $V(v)$, $V^*(v)$ coincide for impossible worlds, this is not the case for the possible ones, hence the need for new notation.

- For $\phi := J_n\psi$ with Induction Hypothesis that the result holds for ψ. Then $M, w \vDash J_n\psi$ if and only if for all $u \in f(w) \cap W^I$ such that $u^n \neq \varnothing$ there is some $v \in u^n$: $M, v \vDash \psi$. This is the case if and only if for all $u \in f(w) \cap W^I$ such that $u^n \neq \varnothing$ there is some $v \in u^n$: $\psi \in V^*(v)$. By construction, this is the case if and only if $\psi \in \mathrm{J}_n(w)$ if and only if M, $w \vDash J_n\psi$.
- For $\phi := K\psi$ with Induction Hypothesis that the result holds for ψ. Then $M, w \vDash K\psi$ if and only if for all $u \in f(w)$: $M, u \vDash \psi$. Since $u \in W^P \cup W^I$, this is the case if and only if $M, w \vDash L\psi$ and $M, w \vDash I\psi$. Given the previous steps of the proof, this is the case if and only if M, $w \vDash L\psi$ and M, $w \vDash I\psi$, if and only if M, $w \vDash K\psi$.
- For $\phi := \langle n \rangle K\psi$ with Induction Hypothesis that the result holds for $K\psi$. Then $M, w \vDash \langle n \rangle K\psi$ if and only if (by Lemma 3), $M, w \vDash L\psi \wedge J_n\psi$ if and only if $M, w \vDash L\psi$ and $M, w \vDash J_n\psi$ if and only if (by previous steps) M, $w \vDash L\psi$ and M, $w \vDash J_n\psi$, if and only if M, $w \vDash \langle n \rangle K\psi$.

□

4. Next, we turn to why this reduction resembles awareness structures and exploit this observation to construct a sound and complete axiomatic system. Recall that awareness structures have the form $\langle W, R, V, \mathcal{A} \rangle$. Therefore, an awareness structure is composed by a set of possible worlds, an epistemic accessibility relation, a valuation on the set of possible worlds and a function assigning a set of formulas to each world. Abstracting away from the conceptual reading of this last component under awareness, we can draw the analogy to M. Again, we have a set of possible worlds W, an accessibility function to denote which worlds are considered epistemically accessible, a valuation on the set of possible worlds and two functions that assign a set of formulas to each possible world. For simplicity, these functions are called *awareness-like*. Now, recall, Theorem 3.2.1. The key axiom is the one that reduces explicit knowledge to implicit knowledge and awareness (whose semantic clause is determined by the awareness function). We can obtain similar axioms for $K\phi$ and $\langle n \rangle K\phi$ reducing these formulas to formulas involving $L\phi$ and $I\phi$, $J_n\phi$ – with $L\phi$ essentially capturing the possible-worlds interpretation of knowledge and the other two formulas being determined by the awareness-like functions. With these comments in mind, we proceed to our claim for a sound and complete logic Λ_r.

   Definition 4.4.2 (Axiomatization of Λ_r). Λ_r is axiomatized by:

   (PC) All instances of propositional tautologies.

   (L) $L(\phi \to \psi) \to (L\phi \to L\psi)$

   (AX_1) $K\phi \leftrightarrow (L\phi \wedge I\psi)$

   (AX_2) $\langle n \rangle K\phi \leftrightarrow (L\phi \wedge J_n\phi)$

   and the rules:

   (MP) From ϕ and $\phi \to \psi$, infer ψ.

   (L-N) From ϕ infer $L\phi$.

5. We are now going to show that Λ_r is indeed sound and complete with respect to the awareness-like structures.

Theorem 4.4.2 (Soundness for Λ_r)**.** Λ_r is sound with respect to awareness-like structures.

Proof. The usual arguments suffice regarding (PC), (L), (MP), (L-N). The validity of (AX_1) and (AX_2) is a direct consequence of their construction. □

Towards completeness, we need the following preliminaries. We stick to the procedure employing a canonical model. Taking (maximal) Λ_r-consistent sets and showing Lindebaum's lemma follow the standard paradigm. Then the canonical model in our case is given by:

Definition 4.4.3 (Canonical Model for the awareness-like structures)**.** The canonical model for the logic Λ_{r} is the tuple $\mathcal{M} = \langle \mathcal{W}, \mathcal{F}, \mathcal{V}, \mathcal{I}, \mathcal{J}_n \rangle$ where:

- $\mathcal{W}$ is the set of all Λ_{r}-maximal consistent sets.
- $\mathcal{F}$ is a function from $\mathcal{W}$ to $\mathcal{P}(\mathcal{W})$ such that $u \in \mathcal{F}(w)$ if and only if $\{\phi \mid L\phi \in w\} \subseteq u$.
- $\mathcal{V}(w) = \{p \mid p \in w\}$, with $w \in \mathcal{W}$.
- $\mathcal{I}(w) = \{\phi \mid I\phi \in w\}$, with $w \in \mathcal{W}$.
- $\mathcal{J}_n(w) = \{\phi \mid J_n\phi \in w\}$ with $w \in \mathcal{W}$.

Unsurprisingly, there is an alternative but equivalent definition of $\mathcal{F}$ in terms of the dual $\hat{L}$, i.e. $\hat{L}\phi := \neg L \neg \phi$. Then $u \in \mathcal{F}(w)$ if and only if $\{\hat{L}\phi \mid \phi \in u\} \subseteq w$. The existence lemma is then obtained by the traditional routine. That is, for any world $w \in \mathcal{W}$, if $\hat{L}\phi \in w$ then there is some $v \in \mathcal{W}$ such that $v \in \mathcal{F}(w)$ and $\phi \in v$. We will also use the dual modality in showing the truth lemma that follows. Finally:

Theorem 4.4.3 (Completeness for Λ_{r})**.** Λ_{r} is (strongly) complete with respect to awareness-like structures.

As we have seen earlier, and as one can verify in Blackburn et al. (2001), it is enough to show the truth lemma. We perform induction on the complexity of ϕ to show that $\mathcal{M}, w \vDash \phi$ if and only if $\phi \in w$.

Proof.

- The claim for propositional atoms and the boolean cases clearly holds, due to the construction of the canonical model and the properties of maximal consistent sets.
- Let $\phi := \hat{L}\psi$ with Induction Hypothesis that the claim holds for ψ. Then $\mathcal{M}, w \vDash \hat{L}\psi$ if and only if for some $u \in \mathcal{F}(w)$: $\mathcal{M}, u \vDash \psi$. From I.H. $\psi \in u$. From the definition of $\mathcal{F}$: $\hat{L}\psi \in w$. For the other direction, suppose that $\hat{L}\psi \in w$. Then by the existence lemma, there is some $u \in \mathcal{W}$ such that $u \in \mathcal{F}(w)$ and $\psi \in u$, therefore using the I.H. we get, $\mathcal{M}, w \vDash \hat{L}\psi$.
- Let $\phi := I\psi$ with Induction Hypothesis that the claim holds for ψ. Then $\mathcal{M}, w \vDash I\psi$ if and only if $\psi \in \mathcal{I}(w)$ if and only if $I\psi \in w$.
- $\phi := J_n\psi$ with Induction Hypothesis that the claim holds for ψ. Then $\mathcal{M}, w \vDash J_n\psi$ if and only if $\psi \in \mathcal{J}_n(w)$ if and only if $J_n\psi \in w$.

- $\phi := K\psi$ with Induction Hypothesis that the claim holds for ψ. Then $\mathcal{M}, w \vDash K\psi$ if and only if $\mathcal{M}, w \vDash L\psi$ and $\mathcal{M}, w \vDash I\phi$ if and only if (from previous steps) $L\psi \in w$ and $I\psi \in w$ if and only if $L\psi \wedge I\psi \in w$ if and only if (from AX_1) $K\psi \in w$.
- $\phi := \langle n \rangle K\psi$ with Induction Hypothesis that the claim holds for $K\psi$. Then $\mathcal{M}, w \vDash \langle n \rangle K\psi$ if and only if $\mathcal{M}, w \vDash L\psi$ and $\mathcal{M}, w \vDash J_n\psi$ if and only if (as above) $L\psi \in w$ and $J_n\psi \in w$ if and only if $L\psi \wedge J_n\psi \in w$ if and only if (from AX_2) $\langle n \rangle K\psi \in w$.

□

Overall, we have provided a sound and complete logic with respect to the awareness-like structures, which are nothing but reductions of our familiar impossible-world models.

This has been a "detour to completeness" for *Rasmussen*&*Bjerring*. At the beginning, we promised that with this first attempt available, analogous results can be obtained for IW. The procedure can be followed with some minor, but crucial, modifications. Most importantly, as far as the language for the reduction is concerned, it is better to employ sequences of rules, rather than individual rules, as primitives (with the semantic clause given by the analysis we performed in Section 4.1.2; then individual applications are simply special cases). This is because, for the evaluation of arrays of reasoning, both the special operator J and the awareness-like function J of the reduction need to "scan" for worlds that follow from epistemically accessible worlds *after* all reasoning steps are taken (this is why they should be now indexed by sequences). Resorting to sequences $\langle \ddagger \rangle^n$, as well as tailoring the truth clause of $J_{\langle \ddagger \rangle^n}$ and construction of $\mathrm{J}_{\langle \ddagger \rangle^n}$ – so that n is replaced by the rules whose application comprise the sequence – allows for similar results according to the foregoing method. Of course, one might simply follow another way and develop simple, indexed "step-logics". For instance, by allowing only one application of rules, i.e. $\langle R_k \rangle K\psi$ in the language[20], it becomes clear that R_k might simply replace n throughout the procedure. Then we get an axiomatization like the one in Definition 4.4.2, where again we have R_k instead of n.

On a final note and in connection with IWe, IWPA and IWp: irrespective of the route, a reduction for IW, and the appropriate axiomatic system it generates, seem to us essential for analogous results within these systems too. It certainly paves the way to a logic for resource depletion. Then, the task seems easier for IWPA, given the reduction axioms of PAL (Section 2.1). However, obtaining a complete system for IWp is trickier; it any case, though, we need reduction axioms, utilizing conditional beliefs, on top of the "basic" system supplied for IW. Apart from these potential applications, we hope that the detour provides a general method to reduce and further explore the properties of frameworks with impossible worlds.

[20] Noticing ,though, that this is a point where objections might be easily raised.

Chapter 5

Conclusions and further research

To sum up, we first presented the standard epistemic and doxastic logics and explained where the problem of logical omniscience lies. We continued by describing how tools from DEL enhance the potential of a realistic logical modelling. With these in hand, we discussed leading attempts to resolve the problem, noticing their useful insights while still critically evaluating their performance. The remarks on the advantages and the disadvantages of the existing approaches paved the way for our own proposals against the problem. In particular, we provided a range of frameworks that take up the highlighted challenges, attacking the problem by prioritizing the investigation of an agent's reasoning steps. In this spirit, we hope that we have taken a first step towards a realistic and detailed solution.

More specifically, Chapter 1 described how systems of modal logic facilitated the formal study of knowledge and belief, but also generated logical omniscience. We then argued on why resolving the problem is a worthwhile task and gleaned properties seen as necessary for an attractive solution. Chapter 2 then briefly discussed how the simple account can be brought closer to real-life phenomena by introducing machinery from DEL and Belief Revision.

The examination of Chapter 3 began with syntactically-oriented attempts; *Syntactic structures* involve a valuation divorcing the evaluation of sentences from the usual recursive computation. Yet they fall short in conveying any property of propositional attitudes and of real, moderately competent agents. Next, Rasmussen's setting, taking its cue from Duc's *dynamic epistemic logic*, argues for a dynamic logic focusing on the agent's effort to apply inference rules. However, the lack of semantics limits the strength of this approach and undermines the adequacy of the designed logic. We continued with a large cluster of logical frameworks that promise to escape omniscience by discerning *explicit* and *implicit* attitudes. *Awareness* and *algorithmic knowledge* respectively employ an awareness function and an algorithm to locate the distinguishing factor. They too suffer criticisms, though, due to the arbitrariness of these components and the threat of retaining forms of the problem. *Justification Logic* and *logics of justified knowledge and belief* integrate the much anticipated notion of justification into the discussion and better approximate real-life situations. Concerns are then raised as these approaches might yield rather contrived ways out of the problem, or simply inherit the remarks made for *awareness*. The third family of proposals suggests the inclusion of *impossible worlds*.

Although initial attempts, solely resorting to their anarchic nature, were unsuccessful in terms of realistic and balanced modelling, this line of research still produced more sophisticated settings. Jago, who uses impossible worlds and puts forward an argument of *Indeterminacy* to explain the fallibility of agents, while still respecting normative standards, faces opposition on the motivation and adequacy of this notion. Finally, Rasmussen&Bjerring build a dynamic framework that merges a step-wise analysis of reasoning and the benefits from the introduction of impossible worlds. Despite their balanced response to the problem, doubts arise due to modelling choices that subvert its explanatory power.

Both the virtues and the shortcomings of these attempts fueled the enterprise of Chapter 4. In order to face open challenges and give an accurate idea of how we attain knowledge and belief, we first supplied a step-wise proposal with semantics. Specifically, we built two families of refinements. First, *Rule-based worlds*, eventually leading to the sound and complete logic Λ_{RW}, capture reasoning processes by structuring the domain of worlds according to inference rules. This allowed for a closer examination of Rasmussen's axioms and subsequently for a fuller attack against logical omniscience. The second approach, dubbed *Impossible worlds*, worked towards the same direction, only now capturing the effect of inference rules in a dynamic impossible-worlds framework that incorporates cognitive costs too. A quantitative extension exemplified the importance of cognitive effort and formally introduced what mostly was the informal motivation of some approaches in Chapter 3. Afterwards, we furnished this approach with actions of hard and soft information (public announcements, radical upgrades) and gave a detailed analysis of reasoning, by merging impossible worlds and ideas from the rich DEL literature. Finally, we showed how models with impossible worlds can be reduced to awareness-like ones. This "detour" is important in constructing sound and complete logics, with Λ_r being our case in point.

Throughout the thesis, we have hinted at several topics suggestive of further investigation. There are, at least, three directions one could follow: (a) enrich/refine the frameworks of Chapter 4, (b) obtain complete axiomatizations for the DEL-inspired extensions of impossible-worlds settings, and (c) use empirical research towards a stronger view on the quantification of resources.

To begin with (a), we have already mentioned that systems accounting for external information are still compatible with the method followed towards a quantitative extension. Therefore, we can easily fit the formal portrayal of resource consumption under these too. It is less easy, albeit interesting, to study how other actions and, in general, insights from DEL and Belief Revision can be combined with our impossible-worlds settings; till now we have confined ourselves to inference rules, public announcements and radical upgrades. In addition, the multi-agent case, and therefore the study of the agents' interaction, common knowledge etc. poses another natural line of research. We also emphasize that while higher-order assertions (e.g. the agent's knowledge/beliefs about her own knowledge/beliefs or reasoning tracks) are prohibited in our settings, it would be interesting to allow for introspective agents – and still treat idealization in this respect. For now, we only suggest that under *Rule-based worlds*, introspective agents may be allowed, given that a new class of models is specified so that its transitions additionally reflect the outcome of introspection.

On the more technical side of (b), our first attempt towards a reduction is indicative of the path one might follow. In particular, we expect that, given the usual reduction axioms for public announcements, completeness results for the corresponding extension may be easily obtained. On the other hand, it is not as straightforward to speculate on a logic with respect to our plausibility models, due to the intricate effect of the actions.

The last direction requires an interdisciplinary point of view. Our own attempt to explain that the agent's insufficiency of resources perturbs reasoning offered the formal equipment to compare her cognitive capacity and the costs for each reasoning step. We did not delve into questions regarding the most appropriate system of units or the most precise way to quantify over these, with respect to memory, time etc. For this, we crucially rely on empirical indications and, in particular, on results from cognitive science and psychology of reasoning. The import of these disciplines, once assimilated by the logical system, may result in a more accurate modelling device.

Bibliography

Alchourrón, C. E., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2):510–530.

Alxatib, S. and Pelletier, F. J. (2011). The psychology of vagueness: Borderline cases and contradictions. *Mind and Language*, 26(3):287–326.

Artemov, S. (2008). The logic of justification. *The Review of Symbolic Logic*, 1(4):477–513.

Artemov, S. (2011). *Why Do We Need Justification Logic?*, pages 23–38. Springer Netherlands, Dordrecht.

Artemov, S. and Fitting, M. (2016). Justification logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition.

Artemov, S. and Kuznets, R. (2006). *Logical Omniscience Via Proof Complexity*, pages 135–149. Springer Berlin Heidelberg, Berlin, Heidelberg.

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*, TARK 09, pages 14–23, New York, NY, USA. ACM.

Artemov, S. and Kuznets, R. (2013). Logical omniscience as infeasibility.

Baltag, A., Moss, L. S., and Solecki, S. (1998). The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '98, pages 43–56, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

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., Renne, B., and Smets, S. (2012). *The Logic of Justified Belief Change, Soft Evidence and Defeasible Knowledge*, pages 168–190. Springer Berlin Heidelberg, Berlin, Heidelberg.

Baltag, A., Renne, B., and Smets, S. (2014). The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic*, 165(1):49–81.

Baltag, A. and Smets, S. (2008). A qualitative theory of dynamic interactive belief revision. *Logic and the Foundations of Game and Decision Theory, Texts in Logic and Games*, (3):9–58.

Berto, F. (2013). Impossible worlds. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2013 edition.

Berto, F. and Plebani, M. (2015). *Ontology and Metaontology. A Contemporary Guide*. Bloomsbury Academic.

Bezhanishvili, N. and van der Hoek, W. (2014). Structures for epistemic logic. In *Johan van Benthem on Logic and Information Dynamics*.

Bjerring, J. C. (2013). Impossible worlds and logical omniscience: an impossibility result. *Synthese*, 190(13):2505–2524.

Blackburn, P., de Rijke, M., and Venema, Y. (2001). *Modal Logic*. Cambridge University Press, New York, NY, USA.

Buss, S. R. (1998). *Handbook of Proof Theory*. Elsevier.

Danto, A. C. (1967). On knowing that we know. In Stroll (1967), pages 32–53.

Dieussaert, K., Schaeken, W., Schroyens, W., and D'Ydewalle, G. (2000). Strategies during complex conditional inferences. *Thinking & Reasoning*, 6(2):125–160.

Duc, H. N. (1997). Reasoning about rational, but not logically omniscient, agents. *Journal of Logic and Computation*, 7(5):633.

Eberle, R. A. (1974). A logic of believing, knowing, and inferring. *Synthese*, 26(3/4):356–382.

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

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

Fagin, R., Halpern, J. Y., Moses, Y., and Y., V. M. (1995a). *Reasoning About Knowledge*. MIT press.

Fagin, R., Halpern, J. Y., and Vardi, M. Y. (1995b). A nonstandard approach to the logical omniscience problem. *Artificial Intelligence*, 79(2):203 – 240.

Fitting, M. (2014). Tr-2014004: Justification logics and realization. Technical report, CUNY Academic Works.

Fiutek, V. (2013). *Playing with knowledge and belief*. PhD thesis, University of Amsterdam.

Griggs, R. A. and Cox, J. R. (1982). The elusive thematic-materials effect in wason's selection task. *British Journal of Psychology*, 73(3):407–420.

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic*, 17(2):157–170.

Halpern, J. Y., Moses, Y., and Vardi, M. Y. (1994). Algorithmic knowledge. In *Proceedings of the 5th Conference on Theoretical Aspects of Reasoning About Knowledge*, TARK '94, pages 255–266, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Halpern, J. Y. and Pucella, R. (2011). Dealing with Logical Omniscience. *Artificial Intelligence*, 175(1):220–235.

Hintikka, J. (1962). *Knowledge and Belief: an introduction to the logic of the two notions*. Ithaca, N.Y.,Cornell University Press.

Hintikka, J. (1975). Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4(4):475–484.

Jago, M. (2006). Hintikka and cresswell on logical omniscience. *Logic and Logical Philosophy*, 15(3):325–354.

Jago, M. (2009). Epistemic logic for rule-based agents. *Journal of Logic, Language and Information*, 18(1):131–158.

Jago, M. (2014). *The Impossible: An Essay on Hyperintensionality*. Oxford University Press.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5):1449–1475.

Konolige, K. (1988). What awareness isn't: A sentential view of implicit and explicit belief. *Journal of Symbolic Logic*, 53(2):667–668.

Lehrer, K. (2000). *Theory of Knowledge*. Westview Press.

Lehrer, K. and Paxson, T. (1969). Knowledge: Undefeated justified true belief. *Journal of Philosophy*, 66(8):225–237.

Lemmon, E. J. (1967). If I know, do I know that I know? In Stroll (1967), pages 54–83.

Levesque, H. J. (1984). A logic of implicit and explicit belief. In *Proceedings of the Fourth AAAI Conference on Artificial Intelligence*, AAAI'84, pages 198–202. AAAI Press.

Moreno, A. (1998). Avoiding logical omniscience and perfect reasoning: a survey. *AI Communications*, 11(2):101–122.

Moses, Y. (1988). Resource-bounded knowledge. In *Proceedings of the 2Nd Conference on Theoretical Aspects of Reasoning About Knowledge*, TARK '88, pages 261–275, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Parikh, R. (2008). Sentences, belief and logical omniscience, or what does deduction tell us? *The Review of Symbolic Logic*, 1:459–476.

Plaza, J. (2007). Logics of public communications. *Synthese*, 158(2):165–179.

Priest, G. (2006). *In Contradiction*. Oxford University Press UK.

Rasmussen, M. S. (2015). Dynamic epistemic logic and logical omniscience. *Logic and Logical Philosophy*, 24:377–399.

Rasmussen, M. S. and Bjerring, J. C. (2015). A dynamic solution to the problem of logical omniscience. *(forthcoming)*.

Ripley, D. (2011). *Contradictions at the Borders*, pages 169–188. Springer Berlin Heidelberg, Berlin, Heidelberg.

Schacter, D. L. (1987). Implicit Memory: History and Current Status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3):501–518.

Schacter, D. L. and Tulving, E. (1994). *Memory Systems 1994*. MIT Press.

Simon, H. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118.

Spohn, W. (1988). Ordinal conditional functions. a dynamic theory of epistemic states. In Harper, W. L. and Skyrms, B., editors, *Causation in Decision, Belief Change, and Statistics, vol. II*. Kluwer Academic Publishers.

Stalnaker, R. (1991). The problem of logical omniscience, I. *Synthese*, 89(3):425–440.

Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies*, 128(1):169–199.

Stenning, K. and van Lambalgen, M. (2008). *Human Reasoning and Cognitive Science*. Boston, USA: MIT Press.

Stokhof, M. and van Lambalgen, M. (2011). Abstractions and idealisations: The construction of modern linguistics. *Theoretical Linguistics*, 37(1-2):1–26.

Stroll, A., editor (1967). *Epistemology*. Harper and Rowe, New York.

van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155.

van Benthem, J. (2008). Logic and reasoning: Do the facts matter? *Studia Logica: An International Journal for Symbolic Logic*, 88(1):67–84.

van Benthem, J., van Ditmarsch, H., van Eijck, J., and Jaspars, J. (2016). *Logic in Action*.

van Ditmarsch, H. (2005). Prolegomena to dynamic logic for belief revision. *Synthese*, 147(2):229–275.

van Ditmarsch, H., Halpern, J., van der Hoek, W., and Kooi, B. (2015). *Handbook of Epistemic Logic*. College Publications.

van Ditmarsch, H., van der Hoek, W., and Kooi, B. (2004). Public announcements and belief expansion. *Proceedings of AiML-2004 (Advances in Modal Logic), University of Manchester*, pages 62–73.

van Ditmarsch, H., van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*. Springer Publishing Company, Incorporated, 1st edition.

Vardi, M. Y. (1986). On epistemic logic and logical omniscience. In *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning About Knowledge*, TARK '86, pages 293–305, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Velázquez-Quesada, F. R. (2011). *Small Steps in Dynamics of Information*. PhD thesis, Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam (UvA), Amsterdam, The Netherlands. ILLC Dissertation series DS-2011-02.

Wansing, H. (1990). A general possible worlds framework for reasoning about knowledge and belief. *Studia Logica*, 49(4):523–539.

Yap, A. (2014). Idealization, epistemic logic, and epistemology. *Synthese*, 191(14):3351–3366.