Some Remarks on the Model Theory of Epistemic Plausibility Models

Lorenz Demey*

Center for Logic and Analytic Philosophy
University of Leuven, Belgium
lorenz.demey@hiw.kuleuven.be

1 Introduction

Traditional epistemic logic can be seen as a particular branch of modal logic. Its semantics is defined in terms of Kripke models, and philosophical principles about knowledge (e.g. factivity: $K\varphi \to \varphi$) are shown to correspond to properties of the epistemic accessibility relation (e.g. reflexivity). By adding another (doxastic) accessibility relation, also *belief* can be treated in this framework. Belief is not assumed to be factive, but at least consistent ($\neg B\bot$), which corresponds to requiring the doxastic accessibility relation to be serial instead of reflexive. In this extended framework, one can study the interaction between knowledge and belief, such as the validity of $K\varphi \to B\varphi$ [11], [12]. Furthermore, since this framework is still 'just' a (multi-)modal logic, it inherits the mathematically well-developed model theory of modal logic.

This framework can also be used to model the interaction of (factive) *knowledge* with public announcements [10],[13] and other dynamic epistemic phenomena [1], [2]. The dynamics of *belief* (and other non-factive attitudes), however, cannot be modeled in this framework: if an agent receives a true piece of information φ while previously believing that $\neg\varphi$, then this agent is predicted to go insane and start believing *everything* (rather than performing a realistic process of *belief revision*) — thus contradicting the consistency requirement about belief. For more details we refer the reader to section 3.1 of [6]. To remedy this problem, epistemic plausibility models have been introduced (technical details will be presented later). In these models, one can again study knowledge, belief (and even other cognitive propositional attitudes), and their various interactions. Furthermore, this framework provides a realistic model of various dynamic phenomena, and thus solves the main problem of the previous approach. A prime example of the use of these models in game theory is the treatment in [4] of the backward induction paradox. Because epistemic plausibility models are much richer structures than Kripke models, however, they do not straightforwardly inherit the model-theoretical results of modal logic. Therefore, while epistemic plausibility structures are well-suited for modeling purposes, an extensive investigation of their model theory has been lacking so far.

The aim of the present paper is to fill exactly this gap, by initiating a systematic exploration of the model theory of epistemic plausibility models. Like in 'ordinary' modal logic, the focus will be on the notion of *bisimulation* — it turns out that finding the right generalization of this notion is not a trivial task. In Section 2, we introduce epistemic plausibility models and discuss some important operators which can be interpreted on such models, and their dynamic behaviour. In Section 3, we define various notions of bisimulations (parametrized by a language $\mathcal{L}$) and show that $\mathcal{L}$-bisimilarity implies $\mathcal{L}$-equivalence. We establish a Hennesy-Milner type theorem, and prove two undefinability results — thus shedding some light on the formal relationships between the various operators that can be interpreted on epistemic plausibility models. The notion of bisimulation for conditional belief, however, turns out to be unsatisfactory for several reasons. In Section 4, we discuss these reasons and explore two possible solutions: adding a modality to the language,

* I wish to thank Johan van Benthem, Davide Grossi and two anonymous LOFT-referees for their valuable comments on earlier versions of this paper. This research was funded by the University of Leuven's *Formal Epistemology Project*.

and putting extra constraints on the models. In Section 5, we establish some results about the interaction between bisimulation and dynamic model changes.

From a broader perspective, this paper can be seen as a reaction against a widespread trend in the technical modal logic literature (already since the 1960's), viz. the inclination to focus almost exclusively on the model theory of 'simple' local modalities that are interpreted by means of universal/existential quantification over some unordered set of accessible states. Despite their central importance in game theory, AI, philosophy, and linguistics, more complex modalities (which are interpreted by means of 'jumping' to the minimal states according to some plausibility ordering) have mainly been neglected from the model-theoretical perspective, without any good reason. This paper aims to show also these more complex (application-oriented) modalities can give rise to a relatively well-behaved and mathematically elegant metatheory.[1]

2 Epistemic plausibility models

We now introduce epistemic plausibility models. Let G be a non-empty set, whose elements will be called *agents*. Throughout this paper, we will keep the set of agents fixed, so that it can almost always be left implicit. Likewise, we assume that *Prop* is a (countably infinite) set of proposition letters, which will also be kept fixed throughout the paper.

Definition 1. *An* epistemic plausibility model *is a structure* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$, *where* W *is a non-empty set of* states, $\sim_i \subseteq W \times W$ *is the* epistemic accessibility relation *for agent* i, $\leq_{i,w} \subseteq W \times W$ *is the* plausibility order *for agent* i *at state* w, *and* $V\colon Prop \to \wp(W)$ *is a* valuation.

As usual, $w \sim_i v$ is to be read as: "agent i cannot epistemically distinguish between states w and v". We assume this relation to be an equivalence relation. Furthermore, $w \leq_{i,s} v$ is to be read as: "at state s, agent i considers w at least as plausible as v". We take this relation to be a well-founded pre-order. For each $X \subseteq W$, we define the set of $\leq_{i,s}$-minimal elements as $\text{Min}_{\leq_{i,s}}(X) := \{x \in X \mid \forall y \in X : y \leq_{i,s} x \Rightarrow x \leq_{i,s} y\}$. That $\leq_{i,s}$ is a well-founded pre-order means that it is reflexive and transitive, and that for each nonempty $X \subseteq W$ also $\text{Min}_{\leq_{i,s}}(X)$ is nonempty. Note that the relation $\leq_{i,s}$ is not only dependent on agents, but also on states: it is possible for agent i to have different plausibility orderings at different states (from Section 4 onwards, more constraints will be placed on this state-dependency).

Various epistemic and doxastic notions can be interpreted on epistemic plausibility models. The three most important ones are: (i) $K_i\varphi$ (*i knows that φ*), (ii) $B_i^\alpha\varphi$ (*i believes that φ, conditional on α*), and (iii) $B_i^+\varphi$ (*i safely believes that φ*). 'Normal' belief can be defined in terms of conditional belief, by putting $B_i\varphi := B_i^\top\varphi$. 'Safe belief' is the name given in [3] to a doxastic attitude between belief and 'full' knowledge. This non-introspective attitude is sometimes called 'defeasible knowledge'; Stalnaker [14] even takes this operator to be a more faithful representation of our 'everyday notion' of knowledge than the full-fledged $\mathsf{S5}$-type K_i-operator.

We abbreviate $[w]_{\sim_i} := \{v \in W \mid w \sim_i v\}$ (the $\sim_i$-equivalence class of state $w \in W$). The semantics for the notions above can now be stated as follows:

Definition 2. *Consider an epistemic plausibility model* $\mathbb{M}$ *and state* w*; then*

- $\mathbb{M}, w \models K_i\varphi$ *iff* $\forall v \in [w]_{\sim_i} : \mathbb{M}, v \models \varphi$
- $\mathbb{M}, w \models B_i^\alpha\varphi$ *iff* $\forall v \in W : v \in \text{Min}_{\leq_{i,w}}([\![\alpha]\!]^{\mathbb{M}} \cap [w]_{\sim_i}) \Rightarrow \mathbb{M}, v \models \varphi$
- $\mathbb{M}, w \models B_i^+\varphi$ *iff* $\forall v \in [w]_{\sim_i} : v \leq_{i,w} w \Rightarrow \mathbb{M}, v \models \varphi$

We now turn to the dynamics. In this paper, we will focus on two specific dynamic phenomena: *public announcement* (hard information) and *radical upgrade* (soft information). Public announcement of a formula φ in an epistemic plausibility model $\mathbb{M}$ simply removes all $\neg\varphi$-states from

[1] In this extended abstract, all theorems, propositions and facts are stated without proof. The full version of this paper, which will soon be available online, contains detailed proofs of all results mentioned in this extended abstract.

the model. Radical upgrade with φ, on the other hand, makes all φ-states more plausible than all $\neg\varphi$-states, and leaves everything within these two zones untouched. Formally, this looks as follows:

Definition 3. *Consider an epistemic plausibility model* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and a formula* φ. *We now define the following epistemic plausibility models:*

- $\mathbb{M}!\varphi = \langle W^{!\varphi}, \{\sim_i^{!\varphi}\}_{i\in G}, \{\leq_{i,w}^{!\varphi}\}_{i\in G}^{w\in W^{!\varphi}}, V^{!\varphi}\rangle$, *where*
  - $W^{!\varphi} = [\![\varphi]\!]^{\mathbb{M}}$, *and* $V^{!\varphi}(p) := V(p) \cap [\![\varphi]\!]^{\mathbb{M}}$ *for any* $p \in Prop$
  - $\sim_i^{!\varphi} := \sim_i \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$ *for any* $i \in G$
  - $\leq_{i,w}^{!\varphi} := \leq_{i,w} \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})$ *for any* $i \in G$ *and* $w \in W^{!\varphi}$
- $\mathbb{M} \Uparrow \varphi = \langle W^{\Uparrow\varphi}, \{\sim_i^{\Uparrow\varphi}\}_{i\in G}, \{\leq_{i,w}^{\Uparrow\varphi}\}_{i\in G}^{w\in W^{\Uparrow\varphi}}, V^{\Uparrow\varphi}\rangle$, *where*
  - $W^{\Uparrow\varphi} := W$, *and* $V^{\Uparrow\varphi}(p) := V(p)$ *for any* $p \in Prop$
  - $\sim_i^{\Uparrow\varphi} := \sim_i$ *for any* $i \in G$
  - $\leq_{i,w}^{\Uparrow\varphi} := \big(\leq_{i,w} \cap ([\![\varphi]\!]^{\mathbb{M}} \times [\![\varphi]\!]^{\mathbb{M}})\big) \cup \big(\leq_{i,w} \cap ([\![\neg\varphi]\!]^{\mathbb{M}} \times [\![\neg\varphi]\!]^{\mathbb{M}})\big) \cup \big([\![\varphi]\!]^{\mathbb{M}} \times [\![\neg\varphi]\!]^{\mathbb{M}}\big)$ *for any* $i \in G$ *and* $w \in W^{\Uparrow\varphi}$

In order to be able to talk about these new models in the object language, we add operators $[!\varphi]$ and $[\Uparrow \varphi]$, thus obtaining the language $\mathcal{L}(K, B^c, B^+, !, \Uparrow)$. We now link up the models and the language by defining the semantics for the two dynamic modalities. Note that since public announcement is assumed to be *truthful*, it works with a precondition; this is not the case for radical upgrade.

Definition 4. *Consider an epistemic plausibility model* $\mathbb{M}$ *and state* w; *then*

- $\mathbb{M}, w \models [!\varphi]\psi$ *iff (if* $\mathbb{M}, w \models \varphi$ *then* $\mathbb{M}!\varphi, w \models \psi$*)*
- $\mathbb{M}, w \models [\Uparrow \varphi]\psi$ *iff* $\mathbb{M} \Uparrow \varphi, w \models \psi$

Finally, dynamic epistemic/doxastic logics are constructed using the well-known *modular approach*: (i) one starts by taking (an axiomatization of) some static base logic, (ii) then one adds dynamic operators to this logic and (iii) finally, one provides a sound set of reduction axioms, which allow each formula in the dynamic language to be rewritten as an equivalent formula in the static language. Because of this final step, completeness of the dynamified logic is reduced to completeness of the static base logic. It also shows that the dynamic language $\mathcal{L}(K, B^c, B^+, !, \Uparrow)$ is equally expressive as the static language $\mathcal{L}(K, B^c, B^+)$.

3 Bisimulation for epistemic plausibility models

We now start our investigation of the model theory of epistemic plausibility models. The focus will be on the notion of *bisimulation*, which is also central in the model theory of Kripke models. Since we want to explore bisimulation for various languages, we make it into a parametrized notion, so that each language has its own notion of bisimulation, which 'does what it needs to do, and nothing more'.

Below are the definitions of K-bisimulation, B^+-bisimulation and B^c-bisimulation. Since K_i is just the universal modality for $\sim_i$, the notion of K-bisimulation is that of regular bisimulation from modal logic. The notion of B^+-bisimulation is a straightforward generalization. The notion of B^c-bisimulation, however, is much more intricate, since it involves universally quantifying over all formulas of the language $\mathcal{L}(B^+)$. We will return to this issue in later sections.

Definition 5. *Given epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim'_i\}_{i\in G}, \{\leq'_{i,w'}\}_{i\in G}^{w'\in W'}, V'\rangle$ *; a relation* $Z \subseteq W \times W'$ *is a* K-bisimulation *iff*

- *if* $(w, w') \in Z$, *then for all atoms* p: $w \in V(p) \Leftrightarrow w' \in V'(p)$
- *if* $(w, w') \in Z$ *and* $w \sim_i v$, *then there is a* $v' \in W'$ *such that* $(v, v') \in Z$ *and* $w' \sim'_i v'$
- *if* $(w, w') \in Z$ *and* $w' \sim'_i v'$, *then there is a* $v \in W$ *such that* $(v, v') \in Z$ *and* $w \sim_i v$

Definition 6. *Given epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim_i'\}_{i\in G}, \{\leq_{i,w'}'\}_{i\in G}^{w'\in W'}, V'\rangle$ *; a relation* $Z \subseteq W \times W'$ *is a* B^+-bisimulation *iff*

- *if* $(w, w') \in Z$*, then for all atoms* p*:* $w \in V(p) \Leftrightarrow w' \in V'(p)$
- *if* $(w, w') \in Z$ *and* $w \sim_i v$ *and* $v \leq_{i,w} w$*, then there is a* $v' \in W'$ *such that* $(v, v') \in Z$ *and* $w' \sim_i' v'$ *and* $v' \leq_{i,w'}' w'$
- *if* $(w, w') \in Z$ *and* $w' \sim_i' v'$ *and* $v' \leq_{i,w'}' w'$*, then there is a* $v \in W$ *such that* $(v, v') \in Z$ *and* $w \sim_i v$ *and* $v \leq_{i,w} w$

Definition 7. *Given epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim_i'\}_{i\in G}, \{\leq_{i,w'}'\}_{i\in G}^{w'\in W'}, V'\rangle$ *; a relation* $Z \subseteq W \times W'$ *is a* B^c-bisimulation *iff*

- *if* $(w, w') \in Z$*, then for all atoms* p*:* $w \in V(p) \Leftrightarrow w' \in V'(p)$
- *for all formulas* $\alpha \in \mathcal{L}(B^c)$*: if* $(w, w') \in Z$ *and* $v \in \mathrm{Min}_{\leq_{i,w}}([\![\alpha]\!]^{\mathbb{M}} \cap [w]_{\sim_i})$*, then there is a* $v' \in W'$ *such that* $(v, v') \in Z$ *and* $v' \in \mathrm{Min}_{\leq_{i,w'}'}([\![\alpha]\!]^{\mathbb{M}'} \cap [w']_{\sim_i'})$
- *for all formulas* $\alpha \in \mathcal{L}(B^c)$*: if* $(w, w') \in Z$ *and* $v' \in \mathrm{Min}_{\leq_{i,w'}'}([\![\alpha]\!]^{\mathbb{M}'} \cap [w']_{\sim_i'})$*, then there is a* $v \in W$ *such that* $(v, v') \in Z$ *and* $v \in \mathrm{Min}_{\leq_{i,w}}([\![\alpha]\!]^{\mathbb{M}} \cap [w]_{\sim_i})$

The following theorem shows that these are the 'right' notions, since they allow us to establish a characteristic feature of bisimulation: bisimilarity implies modal equivalence.

Theorem 1. *Consider two epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim_i'\}_{i\in G}, \{\leq_{i,w'}'\}_{i\in G}^{w'\in W'}, V'\rangle$*, and a relation* $Z \subseteq W \times W'$*.*

1. *If* Z *is a* K*-bisimulation, then for all* $\varphi \in \mathcal{L}(K)$ *and for all* $(w, w') \in Z$*, it holds that* $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$*.*
2. *If* Z *is a* B^+*-bisimulation, then for all* $\varphi \in \mathcal{L}(B^+)$ *and for all* $(w, w') \in Z$*, it holds that* $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$*.*
3. *If* Z *is a* B^c*-bisimulation, then for all* $\varphi \in \mathcal{L}(B^c)$ *and for all* $(w, w') \in Z$*, it holds that* $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$*.*

Using these separate notions of bisimulations, we can now introduce bisimulations for languages which have more than just one of the operators $K/B^+/B^c$ in a modular way (although conditional belief complicates matters a little bit). Obviously, these combined notions lead to results analogous to Theorem 1; we state just two of these as Theorem 2, for future reference.

Definition 8. *Consider epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim_i'\}_{i\in G}, \{\leq_{i,w'}'\}_{i\in G}^{w'\in W'}, V'\rangle$ *and a relation* $Z \subseteq W \times W'$*.*

- Z *is a* $\{K, B^+\}$*-bisimulation iff* Z *is a* K*-bisimulation and a* B^+*-bisimulation*
- Z *is a* $\{K, B^c\}$*-bisimulation iff* Z *is a* K*-bisimulation and a* B^c*-bisimulation, with the universal quantifiers in Definition 7 ranging over* $\mathcal{L}(K, B^c)$ *(instead of just over* $\mathcal{L}(B^c)$*)*
- Z *is a* $\{K, B^+, B^c\}$*-bisimulation iff* Z *is a* K*-bisimulation, a* B^+*-bisimulation, and a* B^c*-bisimulation, with the universal quantifiers in Definition 7 ranging over* $\mathcal{L}(K, B^+, B^c)$ *(instead of just over* $\mathcal{L}(B^c)$*)*

Theorem 2. *Consider two epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim_i'\}_{i\in G}, \{\leq_{i,w'}'\}_{i\in G}^{w'\in W'}, V'\rangle$*, and a relation* $Z \subseteq W \times W'$*.*

1. *If* Z *is a* $\{K, B^c\}$*-bisimulation, then for all* $\varphi \in \mathcal{L}(K, B^c)$ *and for all* $(w, w') \in Z$*, it holds that* $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$*.*
2. *If* Z *is a* $\{K, B^+\}$*-bisimulation, then for all* $\varphi \in \mathcal{L}(K, B^+)$ *and for all* $(w, w') \in Z$*, it holds that* $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$*.*

One can also wonder about the converse direction of theorems such as Theorem 2: if $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$ for all $\varphi \in \mathcal{L}(K, B^c)$, then is there always a $\{K, B^c\}$-bisimulation $Z \subseteq W \times W'$ such that $(w, w') \in Z$? One of the main results from the model theory of basic modal logic, viz. the Hennesy-Milner theorem (cf. [7], Theorem 2.24) says that this question can be answered positively, at least when the models are assumed to be image-finite. This theorem can easily be generalized to epistemic plausibility models:

Definition 9. *Consider an epistemic plausibility model* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$*. We say that* $\mathbb{M}$ *is* image-finite *if for all* $i \in G$ *and all* $w \in W$*, the set* $[w]_{\sim_i}$ *is finite.*

Theorem 3. *Consider two image-finite models* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim'_i\}_{i\in G}, \{\leq'_{i,w'}\}_{i\in G}^{w'\in W'}, V'\rangle$*. Then for all states* $w \in W$ *and* $w' \in W'$*, if* $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$ *for all* $\varphi \in \mathcal{L}(K, B^c)$*, then* w *and* w' *are* $\{K, B^c\}$*-bisimilar.*

Bisimulations are often used to establish $\mathcal{L}$-equivalence of two models (for some language $\mathcal{L}$). Any aspect in which these two models *do* differ is then immediately seen to be undefinable in $\mathcal{L}$. The notions of bisimulation which have been introduced thus far allow us to prove the following two undefinability results (among others). These results can be seen as tying up some loose ends, in the sense that the results were expected, but not yet explicitly proved in the existing literature.

Proposition 1. *Conditional belief cannot be defined in terms of knowledge and safe belief.*[2]

Proposition 2. *Safe belief cannot be defined in terms of knowledge and conditional belief.*

4 Structural bisimulations

We already noted in the previous section that the notion of B^c-bisimulation introduced in Definition 7 is much more intricate than the other notions. We will now argue that this definition is unsatisfactory for both theoretical and practical reasons.

On the *theoretical* level, since Definition 7 involves universal quantification over $\mathcal{L}(B^c)$, it is not strictly structural. Rather than stating conditions on $\sim_i$ and $\leq_{i,w}$ (as is done in Definitions 5 and 6 of bisimulations for knowledge and safe belief), it essentially involves truth sets of (arbitrary) formulas. A related issue is that this definition of bisimilarity for *models* cannot be turned into a definition of bisimilarity for *frames* by simply dropping the 'atoms' clause (as can be done with Definitions 5 and 6): it depends on truth sets of formulas ($[\![\alpha]\!]^{\mathbb{M}}$ and $[\![\alpha]\!]^{\mathbb{M}'}$), and thus also on the concrete valuations of the models $\mathbb{M}$ and $\mathbb{M}'$.

Practically speaking, Definition 7 makes it often very difficult to prove that two given epistemic plausibility models are actually B^c-bisimilar. In the appendix of [9], induction on the complexity of α (with a cleverly strengthened induction hypothesis) is used to establish that the zig- and zag-conditions of Definition 7 hold for all formulas α. However, this approach is geared towards proving one particular B^c-bisimilarity result (about two artificially crafted models), and cannot easily be generalized to the general case (proving B^c-bisimilarity of arbitrary models). Similar remarks apply to the proof of our Proposition 2.

We will now propose two different solutions to this problem, and explore and compare their advantages and disadvantages. Both solutions involve reducing conditional belief to other modalities which have more standard notions of bisimulation. The first approach involves both extending the language and putting some mild constraints on the epistemic plausibility models. The second approach puts more heavy constraints on the models, but does not need to extend the language. Both solutions have in common that we end up only needing fully structural notions of bisimulation, without any universal quantification over formulas.

[2] This theorem does not contradict Proposition 5, since that definability theorem holds for a *restricted* class of epistemic plausibility models, whereas this undefinability theorem holds for the *entire* class of epistemic plausibility models.

4.1 Adding a new modality

The first approach[3] combines language engineering and putting some mild constraints on the models. These constraints are captured by the following definition:

Definition 10. *An epistemic plausibility model* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *is called* uniform *iff the plausibility relations are uniform within epistemic equivalence classes, i.e. iff for any* $i \in G$ *and* $w, v \in W$*: if* $w \sim_i v$ *then* $\leq_{i,w} = \leq_{i,v}$.

This is a natural condition to impose on epistemic plausibility models: it leads to the (intuitively plausible) epistemic/doxastic introspection principle that agents know their (conditional) beliefs. Furthermore, uniformity is a *dynamically robust* notion, in the sense that if an epistemic plausibility model is uniform, then after it has undergone some dynamics, it is still uniform.

Fact 4. *If an epistemic plausibility model* $\mathbb{M}$ *is uniform, then* $\mathbb{M} \models B_i^\alpha\varphi \to K_iB_i^\alpha\varphi$.

Fact 5. *If an epistemic plausibility model* $\mathbb{M}$ *is uniform, then so are* $\mathbb{M}!\varphi$ *and* $\mathbb{M} \Uparrow \varphi$.

Uniform epistemic plausibility models will become very important later on. First, however, we need to set up some other things. For any agent $i \in G$ and state w in a plausibility model, let us abbreviate $<_{i,w} := \leq_{i,w} - \geq_{i,w}$ and $\cong_{i,w} := \leq_{i,w} \cap \geq_{i,w}$ (so $x <_{i,w} y$ iff $x \leq_{i,w} y$ and not $y \leq_{i,w} x$; and $x \cong_{i,w} y$ iff $x \leq_{i,w} y$ and $y \leq_{i,w} x$). Note that since $\leq_{i,w}$ is a pre-order and thus not necessarily antisymmetric, it is possible that $x \cong_{i,w} y$ and yet $x \neq y$.

We now extend our language with a modality $[>_i]$ to talk about this strict version of the plausibility order. As in Definition 2, the semantics for this modality is relativized to the epistemic equivalence classes:

Definition 11. *Consider an epistemic plausibility model* $\mathbb{M}$ *and state* w*; then*

$$\mathbb{M}, w \models [>_i]\varphi \text{ iff } \forall v \in [w]_{\sim_i} : v <_{i,w} w \Rightarrow \mathbb{M}, v \models \varphi$$

Adding this new modality $[>_i]$ as a primitive operator is justified, in the sense that it cannot be defined in even the richest language of the previous section:

Proposition 3. *The modality* $[>_i]$ *cannot be defined in* $\mathcal{L}(K, B^c, B^+)$.

The $[>]$-modality is actually so expressive that, together with the knowledge operator, it is able to define the notion of conditional belief — at least, when we restrict ourselves to the *uniform* epistemic plausibility models introduced at the beginning of this subsection.

Proposition 4. *For all uniform models* $\mathbb{M}$*, it holds that* $\mathbb{M} \models B_i^\alpha\varphi \leftrightarrow K_i((\alpha \wedge \neg\langle >_i\rangle\alpha) \to \varphi)$.

We now introduce the notion of $[>]$-bisimilarity, which — as desired — is fully structural:

Definition 12. *Given epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim_i'\}_{i\in G}, \{\leq_{i,w'}'\}_{i\in G}^{w'\in W'}, V'\rangle$ *; a relation* $Z \subseteq W \times W'$ *is a* $[>]$-bisimulation *iff*

- *if* $(w, w') \in Z$*, then for all atoms* p*:* $w \in V(p) \Leftrightarrow w' \in V'(p)$
- *if* $(w, w') \in Z$ *and* $w \sim_i v$ *and* $v <_{i,w} w$*, then there is a* $v' \in W'$ *such that* $(v, v') \in Z$ *and* $w' \sim_i' v'$ *and* $v' <_{i,w'}' w'$
- *if* $(w, w') \in Z$ *and* $w' \sim_i' v'$ *and* $v' <_{i,w'}' w'$*, then there is a* $v \in W$ *such that* $(v, v') \in Z$ *and* $w \sim_i v$ *and* $v <_{i,w} w$

[3] This approach is based on a suggestion by Johan van Benthem and Davide Grossi.

Part 1 of Theorem 6 shows that this is the right notion of bisimulation. Furthermore, we get combined notions of bisimulation in the obvious way. In particular, $\{K, [>]\}$-bisimulations are combined K- and $[>]$-bisimulations; since both of the latter notions are purely structural, also $\{K, [>]\}$-bisimulation is structural. Part 2 of Theorem 6 is the analogue of part 1 for this combined notion. Most importantly, part 3 states that when we restrict ourselves to the class of *uniform* models, we can get equivalence for conditional belief[4] by means of a structural notion of bisimulation. Finally, part 4 says that if we restrict to uniform image-finite models, then (structural) $\{K, [>]\}$-bisimilarity implies $\{K, B^c\}$-bisimilarity (which involves universal quantification over formulas).

Theorem 6. *Consider two epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim'_i\}_{i\in G}, \{\leq'_{i,w'}\}_{i\in G}^{w'\in W'}, V'\rangle$*, and a relation* $Z \subseteq W \times W'$.

1. *If Z is a $[>]$-bisimulation, then for all $\varphi \in \mathcal{L}([>])$ and for all $(w, w') \in Z$, it holds that $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$.*
2. *If Z is a $\{K, [>]\}$-bisimulation, then for all $\varphi \in \mathcal{L}(K, [>])$ and for all $(w, w') \in Z$, it holds that $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$.*
3. *If $\mathbb{M}$ and $\mathbb{M}'$ are uniform, and Z is a $\{K, [>]\}$-bisimulation, then for all $\varphi \in \mathcal{L}(K, B^c)$ and for all $(w, w') \in Z$, it holds that $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$.*
4. *If $\mathbb{M}$ and $\mathbb{M}'$ are uniform and image-finite, then for any states $w \in W$ and $w' \in W'$, we have that if w and w' are $\{K, [>]\}$-bisimilar, then they are $\{K, B^c\}$-bisimilar as well.*

We finish this subsection by providing an overview of the first strategy to solve the main issue of Section 3 (viz. finding a structural notion of bisimulation for conditional belief) and evaluating its advantages and disadvantages.

This strategy has two components. The first component is to impose an extra condition on epistemic plausibility models, viz. uniformity. We argued that this is relatively harmless, since it can be given an intuitive motivation in terms of doxastic/epistemic introspection, and because it is dynamically robust (cf. Facts 4 and 5). The second component involves what van Benthem calls "redesigning one's language to fit more standard bisimulations" [5, p. 310]. We introduced a new modality $[>]$ and showed that together with knowledge, it can define conditional belief (for uniform models) (cf. Propositions 3 and 4). We then used the structural notion of $\{K, [>]\}$-bisimilarity to establish $\mathcal{L}(K, B^c)$-equivalence and even $\{K, B^c\}$-bisimilarity itself (cf. Theorem 6).

The main disadvantage of this approach lies in its second component: the $[>]$-operator was introduced for the sole purpose of defining conditional belief (while maintaining a structural notion of bisimulation). In itself, however, it does not seem to have any intuitive epistemic/doxastic reading.[5] Therefore, this solution ends up looking a bit *ad hoc*.

4.2 Assuming connectedness

The second approach tries to keep the advantages of the first one, while avoiding its major drawback, viz. the ad hoc introduction of new operators. The basic idea is that, with an extra condition on the epistemic plausibility models, conditional belief can be reduced to knowledge and safe belief. Hence, the B^+-operator plays the role of the $[>]$-operator in the previous approach, but unlike the $[>]$-operator, it *does* have an intuitive doxastic interpretation. The extra condition on the models that we need is local connectedness:

Definition 13. *An epistemic plausibility model* $\mathbb{M} = \langle W, \{\sim_i\}_{i\in G}, \{\leq_{i,w}\}_{i\in G}^{w\in W}, V\rangle$ *is called* locally connected *iff for all agents $i \in G$ and states $w, v \in W$ it holds that if $w \sim_i v$, then $w \leq_{i,w} v$ or $v \leq_{i,w} w$.*

[4] Actually for $\mathcal{L}(K, B^c)$ — but this is no heavy restriction, since it is natural to study both notions simultaneously anyway.

[5] Cf. "The intuitive meaning of these operators [such as $[>]$, LD] is not very clear, but they can be used to define other interesting modalities, capturing various 'doxastic attitudes'.", [3, p. 32].

Whether this is a natural condition is a bit more doubtful than in the case of uniformity. At least, local connectedness is dynamically robust:

Fact 7. *If an epistemic plausibility model* $\mathbb{M}$ *is locally connected, then so are* $\mathbb{M}!\varphi$ *and* $\mathbb{M} \Uparrow \varphi$*.*

When we require the models to be both uniform (cf. the previous subsection) and locally connected, then conditional belief can be defined in terms of knowledge and safe belief.[6]

Proposition 5. *For all uniform and locally connected models* $\mathbb{M}$*, it holds that* $\mathbb{M} \models B_i^\alpha \varphi \leftrightarrow \big(\hat{K}_i \alpha \to \hat{K}_i(\alpha \wedge B_i^+(\alpha \to \varphi))\big)$*.*

Using this definability result, we immediately obtain the analogon of Theorem 6:

Theorem 8. *Consider two epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i \in G}, \{\leq_{i,w}\}_{i \in G}^{w \in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim_i'\}_{i \in G}, \{\leq_{i,w'}'\}_{i \in G}^{w' \in W'}, V'\rangle$*, and a relation* $Z \subseteq W \times W'$*.*

1. *If* $\mathbb{M}$ *and* $\mathbb{M}'$ *are uniform and locally connected, and* Z *is a* $\{K, B^+\}$*-bisimulation, then for all* $\varphi \in \mathcal{L}(K, B^+, B^c)$ *and for all* $(w, w') \in Z$*, it holds that* $\mathbb{M}, w \models \varphi \Leftrightarrow \mathbb{M}', w' \models \varphi$*.*
2. *If* $\mathbb{M}$ *and* $\mathbb{M}'$ *are uniform, locally connected and image-finite, then for any states* $w \in W$ *and* $w' \in W'$*, we have that if* w *and* w' *are* $\{K, B^+\}$*-bisimilar, then they are* $\{K, B^c\}$*-bisimilar as well.*

Just as we did in the previous subsection, we now provide an overview of the second strategy to solve the main issue of Section 3. This approach reduced conditional belief to knowledge and safe belief, which are both intuitively clear epistemic/doxastic notions. Therefore, the main issue of the first approach, viz. the *ad hoc* character of its introduction of the $[>]$-operator, is avoided. In order to get the desired results about $\mathcal{L}(K, B^c)$-equivalence and $\{K, B^c\}$-bisimilarity (cf. Theorem 8), we required the epistemic plausibility models to be not only uniform, but also locally connected. The uniformity constraint inherits of course all of its justifications (intuitive epistemic/doxastic interpretation and dynamic robustness) from the previous subsection. However, the new constraint, local connectedness, seems to be less motivated: while it is also dynamically robust (cf. Fact 7), it might not have as intuitive an interpretation as the uniformity constraint.

5 Dynamics and bisimulation

In this section, we will make some remarks about the interaction between bisimulation and dynamic model changes. This requires deciding which approach to conditional belief is to be adopted. For the sake of concreteness, we will henceforth adopt the approach developed in Subsection 4.2. However, one should keep in mind that this section could easily be rewritten in terms of the approach developed in Subsection 4.1.

The main use of bisimulations is to prove $\mathcal{L}$-equivalence of two models (for some language $\mathcal{L}$). Using the well-known reduction axioms for $[!\varphi]$ and $[\Uparrow \varphi]$, every formula of the dynamic language $\mathcal{L}(K, B^+, B^c, !, \Uparrow)$ can be rewritten as an equivalent formula of the original static language $\mathcal{L}(K, B^+, B^c)$. Thus, information about what will be the case after some change has taken place can be *pre-encoded* in the static language. We will now combine this pre-encoding strategy with Theorem 8:

Theorem 9. *Consider two uniform and locally connected epistemic plausiblity models* $\mathbb{M} = \langle W, \{\sim_i\}_{i \in G}, \{\leq_{i,w}\}_{i \in G}^{w \in W}, V\rangle$ *and* $\mathbb{M}' = \langle W', \{\sim_i'\}_{i \in G}, \{\leq_{i,w'}'\}_{i \in G}^{w' \in W'}, V'\rangle$*, states* $w \in W$ *and* $w' \in W'$*, and a* $\{K, B^+\}$*-bisimulation* $Z \subseteq W \times W'$ *such that* $(w, w') \in Z$*. Furthermore, consider an arbitrary formula* $\varphi \in \mathcal{L}(K, B^+, B^c)$*; then:*

1. *If* $\mathbb{M}, w \models \varphi$ *and* $\mathbb{M}', w' \models \varphi$*, then* $\forall \psi \in \mathcal{L}(K, B^+, B^c) : \mathbb{M}!\varphi, w \models \psi \Leftrightarrow \mathbb{M}'!\varphi, w' \models \psi$*.*
2. $\forall \psi \in \mathcal{L}(K, B^+, B^c) : \mathbb{M} \Uparrow \varphi, w \models \psi \Leftrightarrow \mathbb{M}' \Uparrow \varphi, w' \models \psi$*.*

[6] A similar definition was already proposed by Boutilier[8, p. 104].

We finish by making a remark about the *strength* of bisimulation. Theorem 8 tells us that bisimulation 'now' implies modal equivalence 'now'. Theorem 9, however, tells us that bisimulation 'now' implies modal equivalence '*later*' (i.e. after the model has undergone some dynamic effects). Since both uniformity and local connectedness are dynamically robust (cf. Facts 5 and 7), Theorem 9 can be repeated to prove that the same holds for any sequence of epistemic dynamics (e.g. if $\mathbb{M}, w$ and $\mathbb{M}', w'$ are $\{K, B^+\}$-bisimilar, then $(((\mathbb{M}!\varphi_1) \Uparrow \varphi_2) \Uparrow \varphi_3)!\varphi_4, w$ and $(((\mathbb{M}'!\varphi_1) \Uparrow \varphi_2) \Uparrow \varphi_3)!\varphi_4, w'$ are $\mathcal{L}(K, B^+, B^c)$-equivalent — provided they survive the public announcements, of course). Hence, if two epistemic plausibility models are bisimilar at one point, then their entire epistemic-doxastic futures are indistinguishable.

6 Conclusion

The aim of this paper has been to explore the model theory of epistemic plausibility models, which has been largely ignored in the present literature. We focused on the notion of bisimulation, and presented several extensions, parametrized by a language $\mathcal{L}$. Using these notions, we proved various bisimulation-implies-equivalence type theorems, a Hennesy-Milner type theorem, and, perhaps most importantly, two undefinability results. We also presented and compared two alternative ways of getting bisimulations for conditional belief. Finally, we established some results about the interaction between bisimulation and dynamic model changes, and commented on the strength of bisimulation to establish equivalence 'now and in the future'.

References

1. Baltag, A., Moss, L.S.: Logics for epistemic programs. Synthese 139, 1 – 60 (2004)
2. Baltag, A., Moss, L.S., Solecki, S.: The logic of common knowledge, public announcements, and private suspicions. In: Gilboa, I. (ed.) Proceedings of the 7th conference on Theoretical Aspects of Rationality and Knowledge (TARK '98), pp. 43 –56 (1998)
3. Baltag, A., Smets, S.: A qualitative theory of dynamic interactive belief revision. In: Bonanno, G., van der Hoek, W., Woolridge, M. (eds.) Texts in Logic and Games, vol. 3, pp. 9 – 58. Amsterdam University Press (2008)
4. Baltag, A., Smets, S., Zvesper, J.A.: Keep 'hoping' for rationality. a solution to the backward induction paradox. Synthese 169, 301 – 333 (2009)
5. van Benthem, J.: Extensive games as process models. Journal of Logic, Language and Information 11, 289 – 313 (2002)
6. van Benthem, J.: Dynamic logic for belief revision. Journal for Applied Non-Classical Logics 17, 129 – 155 (2007)
7. Blackburn, P., de Rijke, M., Venema, Y.: Modal logic. Cambridge University Press (2001)
8. Boutilier, C.: Conditional logics of normality: a modal approach. Artificial Intelligence 68, 87 – 154 (1994)
9. Dégremont, C., Roy, O.: Agreement theorems in dynamic-epistemic logic (2010), gRIPh Working Paper 0902; available online at `http://www.rug.nl/filosofie/onderzoek/workingpapers/WorkingPaperRoy.pdf`
10. Gerbrandy, J., Groeneveld, W.: Reasoning about information change. Journal of Logic, Language and Information 6, 147 – 169 (1997)
11. Halpern, J.Y.: Should knowledge entail belief? Journal of Philosophical Logic 25, 483 – 494 (1996)
12. van der Hoek, W.: Systems for knowledge and belief. Journal of Logic and Computation 3, 173 – 195 (1993)
13. Plaza, J.: Logics of public communications. In: Emrich, M.L., Pfeifer, M.S., Hadzikadic, M., Ras, Z.W. (eds.) Proceedings of the 4th international symposium on methodologies for intelligent systems, pp. 201 – 216 (1989)
14. Stalnaker, R.: On logics of knowledge and belief. Philosophical Studies 128, 169 – 199 (2006)