Modeling Belief in Dynamic Systems. Part I: Foundations

Nir Friedman

Computer Science Division, 387 Soda Hall, University of California, Berkeley, CA 94720, nir@cs.berkeley.edu, http://www.cs.berkeley.edu/ nir

Joseph Y. Halpern

Computer Science Department, Cornell University, Ithaca, NY 14853, halpern@cs.cornell.edu, http://www.cs.cornell.edu/home/halpern

---

Abstract

Belief change is a fundamental problem in AI: Agents constantly have to update their beliefs to accommodate new observations. In recent years, there has been much work on axiomatic characterizations of belief change. We claim that a better understanding of belief change can be gained from examining appropriate *semantic* models. In this paper we propose a general framework in which to model belief change. We begin by defining belief in terms of knowledge and plausibility: an agent believes ϕ if he knows that ϕ is more plausible than $\neg\phi$. We then consider some properties defining the interaction between knowledge and plausibility, and show how these properties affect the properties of belief. In particular, we show that by assuming two of the most natural properties, belief becomes a KD45 operator. Finally, we add time to the picture. This gives us a framework in which we can talk about knowledge, plausibility (and hence belief), and time, which extends the framework of Halpern and Fagin for modeling knowledge in multi-agent systems. We then examine the problem of "minimal change". This notion can be captured by using *prior plausibilities*, an analogue to prior probabilities, which can be updated by "conditioning". We show by example that conditioning on a plausibility measure can capture many scenarios of interest. In a companion paper, we show how the two best-studied scenarios of belief change, *belief revision* and *belief update*, fit into our framework.

[⋆] Some of this work was done while both authors were at the IBM Almaden Research Center. The first author was also at Stanford while much of the work was done. IBM and Stanford's support are gratefully acknowledged. The work was also supported in part by the Air Force Office of Scientific Research (AFSC), under Contract F49620-91-C-0080 and grant F94620-96-1-0323 and by NSF under grants IRI-95-03109 and IRI-96-25901. A preliminary version of this paper appears in *Proceedings of the 5th Conference on Theoretical Aspects of Reasoning About Knowledge*, 1994, pp. 44-64, under the title "A knowledge-based framework for belief change, Part I: Foundations".

1 Introduction

In order to act in the world we must make assumptions, such as "the corridor is clear" or "my car is parked where I left it". These assumptions, however, are defeasible. We can easily imagine situations where the corridor is blocked, or where the car is stolen. We call the logical consequences of such defeasible assumptions *beliefs*. As time passes, we constantly obtain new information that might cause us to make additional assumptions or withdraw some of our previous assumptions. The problem of *belief change* is to understand how beliefs should change.

The study of belief change has been an active area in philosophy and in artificial intelligence [27,38]. In the literature, two instances of this general phenomenon have been studied in detail: *Belief revision* [1,27] attempts to describe how an agent should accommodate a new belief (possibly inconsistent with his other beliefs) about a static world. *Belief update* [38], on the other hand, attempts to describe how an agent should change his beliefs as a result of learning about a change in the world. Belief revision and belief update describe only two of the many ways in which beliefs can change. Our goal is to construct a framework to reason about belief change in general. This paper describes the details of that framework. In a companion paper [23] we consider the special cases of belief revision and update in more detail.

Perhaps the most straightforward approach to belief change is to simply represent an agent's beliefs as a closed set of formulas in some language and then put constraints on how these beliefs can change. This is essentially the approach taken in [1,27,38]; as their results show, much can be done with this framework. The main problem with this approach is that it does not provide a good semantics for belief. As we hope to show in this paper and in [23], such a semantics can give us a much deeper understanding of how and why beliefs change. Moreover, this semantics provides the tools to deal with complicating factors such actions, external events, and multiple agents.

One standard approach to giving semantics to beliefs is to put a *preference ordering* on the set of worlds that the agent considers possible. Intuitively, such an ordering captures the relative likelihood of worlds. Various authors [3,31,38,57] have then interpreted "the agent believes ϕ" as "ϕ is true in the most plausible worlds that the agent considers possible". An alternative approach is to put a probability measure over the set of possible worlds. Then we can interpret "the agent believes ϕ" as "the probability of ϕ is close to 1" [50]. We examine a new approach to modeling uncertainty based on *plausibility measures*, introduced in [20,24], where a plausibility measure just associates with an event (i.e., a set of possible worlds) its *plausibility*, an element in some partially ordered set. This approach is easily seen to generalize other approaches to modeling uncertainty, such as probability measures, belief functions, and preference orderings. We interpret the "agent believes ϕ" as "the plausibility of ϕ is greater than that of $\neg\phi$". As we show, this is often (but not always) equivalent to "ϕ is true in the most plausible worlds".

By modeling beliefs in this way, there is an assumption that the plausibility measure is part of the agent's epistemic state. (This assumption is actually made explicitly in [3,41].) This implies that the plausibility measure is *subjective*, that is, it describes the agent's estimate of the plausibility of each event. But actually, an even stronger assumption is being made: namely, that the agent's epistemic state is characterized by a *single* plausibility measure. We feel that this latter assumption makes the models less expressive than they ought to be. In particular, they cannot represent a situation where the agent is not sure about what is plausible, such as "Alice does not know that it typically does not rain in San Francisco in the summer". To capture this, we need to allow Alice to consider several plausibility measures possible; in some it typically does not rain and in others it typically does. [1] As we shall see, this extra expressive power is necessary to capture some interesting scenarios of belief change.

To deal with this, in addition to plausibility measures, we add a standard accessibility relation to represent *knowledge*. Once we have knowledge in the picture, we define belief by saying that an agent *believes* ϕ if she *knows* that ϕ is typically true. That is, according to all the plausibility measures she considers possible, ϕ is more plausible than $\neg\phi$.

The properties of belief depend on how the plausibility measure interacts with the accessibility relation that defines knowledge. We study these interactions, keeping in mind that plausibility generalizes probability. In view of this, it is perhaps not surprising that many of the issues studied by Fagin and Halpern [14] when considering the interaction of knowledge and probability also arise in our framework. There are, however, a number of new issues that arise in our framework due to the interaction between knowledge and belief. As we shall see, if we take what are perhaps the most natural restrictions on this interaction, our notion of belief is characterized by the axioms of the modal logic KD45 (where an agent has complete introspective knowledge about her beliefs, but may have false beliefs). Moreover, the interaction between knowledge and belief satisfies the standard properties considered by Kraus and Lehmann [40]. Although our major goal is not an abstract study of the properties of knowledge and belief, we view the fact that we have a concrete interpretation under which these properties can be studied to be an important side-benefit of our approach.

Having a notion of belief is not enough in order to study belief change. We want a framework that captures the beliefs of the agent before and after the change. This is achieved by introducing *time* explicitly into the framework. The resulting framework is an extension of the framework of Halpern and Fagin [33] for modeling knowledge in multi-agent systems, and allows to talk about knowledge, plausibility (and hence belief), and time. This framework is analogous to combination of knowledge, probability and time studied in [35]. As we show by example, having knowledge, plausibility, and time represented explicitly gives us a powerful and expressive framework for capturing belief change.

This framework is particularly suited to studying how plausibility changes over time. One

[1] In fact, this issue is discussed by Boutilier [3], although his framework does not allow him to represent such a situation.

important intuition we would like to capture is that of *minimal change.* Suppose an agent gets new information at time t. Certainly we would expect his plausibility assessment (and his beliefs) at time $t+1$ to incorporate this new information; otherwise, we would expect his assessment at time $t+1$ to have changed minimally from his assessment at time t. In probabilistic reasoning, it can be argued that *conditioning* captures this intuition. Conditioning incorporates the new information by giving it probability 1. Moreover, the relative probability of all events consistent with the new information is the same before and after conditioning, so, in this sense, conditioning changes things minimally. We focus here on a plausibilistic analogue of conditioning and argue that it captures the intuition of minimal change in plausibilities. We can then proceed much in the spirit of the Bayesian approach, but starting with a *prior plausibility* and conditioning. As we show, many situations previously studied in the literature, such as *diagnostic reasoning* [54], can be easily captured by using such prior plausibilities. Moreover, as we show in a companion paper [23], belief revision and belief update—which both attempt to capture intuitions involving minimal change in beliefs—can be captured in our framework by conditioning on an appropriate prior plausibility measure. Thinking in terms of priors also gives us insight into other representations of belief change, such as those of of [5,31,43].

The rest of this paper is organized as follows. In the next section, we review the syntax and semantics of the standard approach to modeling knowledge using Kripke structures and show how plausibility can be added to the framework. Much of our technical discussion of axiomatizations and decision procedures is closely related to that of [14]. In Section 3.1, we present our full framework which adds plausibility to the framework of [33] for modeling knowledge (and time) in multi-agent systems. In Section 4 we introduce prior plausibilities and show how they can be used. We conclude in Section 5 with some discussion of the general approach. Proofs of theorems are given in Appendix A.

2 Knowledge and Plausibility

In this section, we briefly review the standard models for knowledge and beliefs (see [34] for further motivation and details), describe a notion of plausibility, and then show how to combine the two notions. Finally, we compare the derived notion of belief with previous work on the subject.

2.1 The Logic of Knowledge

We start by examining the standard models for knowledge and belief. The syntax for the logic of knowledge is simple: we start with primitive propositions and close off under conjunction, negation, and the modal operators $K_1, \ldots, K_n$. A formula such as $K_i\phi$ is read "agent i knows ϕ". The logic of belief is the result of replacing the K_i operator by B_i. The formula, $B_i\phi$ is read "agent i believes ϕ". The resulting languages are denoted $\mathcal{L}^K$ and $\mathcal{L}^B$, respectively.

The semantics for these languages is given by means of *Kripke structures*. A *Kripke structure for knowledge* (or *belief*) is a tuple $(W, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, where W is a set of possible worlds, $\pi(w)$ is a truth assignment to the primitive propositions at world $w \in W$, and the $\mathcal{K}_i$'s are accessibility relations on the worlds in W. For convenience, we define $\mathcal{K}_i(w) = \{w' : (w, w') \in \mathcal{K}_i\}$. Intuitively, $\mathcal{K}_i(w)$ describes the set of worlds that agent i considers possible in w. We say that agent i knows (or believes) ϕ at world w, if all the worlds $\mathcal{K}_i(w)$ satisfy ϕ.

We assign truth values to formulas at each world in the structure. We write $(M, w) \models \phi$ if the formula ϕ is true at a world w in the Kripke structure M.

- $(M, w) \models p$ for a primitive proposition p if $\pi(w)(p) = \mathit{true}$,
- $(M, w) \models \neg\phi$ if $(M, w) \not\models \phi$,
- $(M, w) \models \phi \wedge \psi$ if $(M, w) \models \phi$ and $(M, w) \models \psi$,
- $(M, w) \models K_i\phi$ if $(M, w') \models \phi$ for all $w' \in \mathcal{K}_i(w)$.

The last clause captures the intuition that ϕ is known exactly when it is true in all possible worlds. When considering the language of beliefs $\mathcal{L}^B$, we typically use $\mathcal{B}_i$ rather than $\mathcal{K}_i$ to denote the accessibility relations. The truth condition for $B_i\phi$ is exactly the same as for $K_i\phi$.

Let $\mathcal{M}_K$ be the class of Kripke structures described above. We say that $\phi \in \mathcal{L}^K$ is valid in some $M \in \mathcal{M}_K$ if $(M, w) \models \phi$ for all w in M. We say that $\phi \in \mathcal{L}^K$ is *valid* in $\mathcal{M}_K$ if it is valid in all models $M \in \mathcal{M}_K$. We say that ϕ is satisfiable in $\mathcal{M}_K$ if there is a model $M \in \mathcal{M}_K$ and world w such that $(M, w) \models \phi$.

The definition of Kripke structure does not put any restriction on the $\mathcal{K}_i$ relations. By imposing conditions on the $\mathcal{K}_i$ relations we get additional properties of knowledge (or belief). These properties are captured by systems of axioms that describe the valid formulas in classes of structures that satisfy various constraints of interest. We briefly describe these systems and the corresponding constraints on the accessibility relations. Consider the following axioms and rules:

K1. All substitution instances of propositional tautologies

K2. $K_i\phi \wedge K_i(\phi \Rightarrow \psi) \Rightarrow K_i\psi$

K3. $K_i\phi \Rightarrow \phi$

K4. $K_i\phi \Rightarrow K_iK_i\phi$

K5. $\neg K_i\phi \Rightarrow K_i\neg K_i\phi$

K6. $\neg K_i \mathit{false}$

RK1. From ϕ and $\phi \Rightarrow \psi$ infer ψ

RK2. From ϕ infer $K_i\phi$

The system K contains the axioms K1 and K2 and the rules of inference RK1 and RK2. By adding axioms K4 and K5 we get system K45; if in addition we add axiom K6 we get system KD45; if instead we add axiom K3 to K45 we get the axiom system known as S5.

We now relate these axiom systems with restrictions on the accessibility relations. We start with some definitions. A relation $\mathcal{R}$ on W is *Euclidean* if $(x,y),(x,z) \in \mathcal{R}$ implies that $(y,z) \in \mathcal{R}$, for all x, y and z in W; it is *reflexive* if $(x,x) \in \mathcal{R}$ for all $x \in W$; it is *serial* if for all $x \in W$ there is a y such that $(x,y) \in R$; and it is *transitive* if $(x,y),(y,z) \in \mathcal{R}$ implies that $(x,z) \in \mathcal{R}$, for x, y and z in W. Let $\mathcal{M}_K^{et}$ be the set of Kripke structures with Euclidean and transitive accessibility relations, $\mathcal{M}_K^{est}$ be the subset of $\mathcal{M}_K^{et}$ where the accessibility relations are also serial, and $\mathcal{M}_K^{ert}$ be the subset of $\mathcal{M}_K^{et}$ where the accessibility relations are also transitive.

Theorem 1 [34] *The axiom system K (resp. K45, KD45, S5) is a sound and complete axiomatization of $\mathcal{L}^K$ with respect to $\mathcal{M}_K$ (resp. $\mathcal{M}_K^{et}$, $\mathcal{M}_K^{est}$, $\mathcal{M}_K^{ert}$).*

In this paper, we use the multi-agent systems formalism of [15] to model knowledge; this means that knowledge satisfies the axioms of S5. (We provide some motivation for this choice below; see [15] for further discussion.)

This implies that if an agent knows ϕ, then ϕ is true (K3) and that the agent is introspective—he knows what he knows and does not know (K4 and K5). Belief, on the other hand, is typically viewed as defeasible. Thus, it does not necessarily satisfy K3. It may satisfy a weaker property, such as K6, which says that the agent does not believe inconsistent formulas. Like knowledge, belief is taken to be introspective, as it satisfies K4 and K5. Thus, in the literature, belief has typically been take to satisfy K45 or KD45; we do the same here. According to Theorem 1, this means that the notion of knowledge we use is characterized by $\mathcal{M}_K^{ert}$ while belief is characterized by $\mathcal{M}_K^{et}$ or $\mathcal{M}_K^{est}$. [2]

2.2 Plausibility Measures

Most non-probabilistic approaches to belief change require (explicitly or implicitly) that the agent has some ordering over possible alternatives. For example, the agent might have a preference ordering over possible worlds [5,32,39] or an entrenchment ordering over formulas [28]. This ordering dictates how the agent's beliefs change. For example, in [32], the new beliefs are characterized by the most preferred worlds that are consistent with the new observation, while in [28] beliefs are discarded according to their degree of entrenchment until it is consistent to add the new observation to the resulting set of beliefs.

Keeping this insight in mind, we now describe *plausibility measures* [20,24]. This is a notion for handling uncertainty that generalizes previous approaches, including various notions of preference ordering. We briefly review the relevant definitions and results here.

Recall that a probability space is a tuple $(W, \mathcal{F}, \mathrm{Pr})$, where W is a set of worlds, $\mathcal{F}$ is

[2] As is well known, a relation is reflexive, Euclidean and transitive if and only if it is an equivalence relation (i.e., reflexive, symmetric and transitive). Thus, $\mathcal{M}_K^{ert}$ consists of these structures where the $\mathcal{K}_i$'s are equivalence relations.

an algebra of *measurable* subsets of W (that is, a set of subsets closed under union and complementation to which we assign probability), and Pr is a *probability measure*, that is, a function mapping each set in $\mathcal{F}$ to a number in $[0,1]$ satisfying the well-known probability axioms ($\Pr(\emptyset)=0$, $\Pr(W)=1$, and $\Pr(A\cup B)=\Pr(A)+\Pr(B)$, if A and B are disjoint).

A plausibility space is a direct generalization of a probability space. We simply replace the probability measure Pr by a *plausibility measure* Pl, which, rather than mapping sets in $\mathcal{F}$ to numbers in $[0,1]$, maps them to elements in some arbitrary partially ordered set. We read $\mathrm{Pl}(A)$ as "the plausibility of set A". If $\mathrm{Pl}(A)\le \mathrm{Pl}(B)$, then B is at least as plausible as A. Formally, a *plausibility space* is a tuple $S=(W,\mathcal{F},\mathrm{Pl})$, where W is a set of worlds, $\mathcal{F}$ is an algebra of subsets of W, and Pl maps sets in $\mathcal{F}$ to some domain D of *plausibility values* partially ordered by a relation $\le_D$ (so that $\le_D$ is reflexive, transitive, and anti-symmetric). We assume that D is *pointed*: that is, it contains two special elements $\top_D$ and $\bot_D$ such that $\bot_D \le_D d \le_D \top_D$ for all $d\in D$; we further assume that $\mathrm{Pl}(W)=\top_D$ and $\mathrm{Pl}(\emptyset)=\bot_D$. As usual, we define the ordering $<_D$ by taking $d_1 <_D d_2$ if $d_1\le_D d_2$ and $d_1\ne d_2$. We omit the subscript D from $\le_D$, $<_D$, $\top_D$ and $\bot_D$ whenever it is clear from context.

Since we want a set to be at least as plausible as any of its subsets, we require

A1 If $A\subseteq B$, then $\mathrm{Pl}(A)\le \mathrm{Pl}(B)$.

Some brief remarks on this definition: We have deliberately suppressed the domain D of plausibility values from the tuple S, since for the purposes of this paper, only the ordering induced by $\le$ on the subsets in $\mathcal{F}$ is relevant. The algebra $\mathcal{F}$ also does not play a significant role in this paper. Unless we say otherwise, we assume $\mathcal{F}$ contains all subsets of interest and suppress mention of $\mathcal{F}$, denoting a plausibility space as a pair (W,Pl).

Clearly plausibility spaces generalize probability spaces. We now briefly discuss a few other notions of uncertainty that they generalize:

- A *belief function* B on W is a function $B: 2^W\to[0,1]$ satisfying certain axioms [55]. These axioms certainly imply property A1, so a belief function is a plausibility measure.
- A *fuzzy measure* (or a *Sugeno measure*) f on W [59] is a function $f: 2^W\mapsto[0,1]$, that satisfies A1 and some continuity constraints. A *possibility measure* [12] Poss is a fuzzy measure such that $\mathrm{Poss}(W)=1$, $\mathrm{Poss}(\emptyset)=0$, and $\mathrm{Poss}(A)=\sup_{w\in A}(\mathrm{Poss}(\{w\})$.
- An *ordinal ranking* (or *κ-ranking*) on W (as defined by [31], based on ideas that go back to [57]) is a function $\kappa: 2^W\to I\!N^*$, where $I\!N^*=I\!N\cup\{\infty\}$, such that $\kappa(W)=0$, $\kappa(\emptyset)=\infty$, and $\kappa(A)=\min_{w\in A}(\kappa(\{w\}))$. Intuitively, an ordinal ranking assigns a degree of surprise to each subset of worlds in W, where 0 means unsurprising and higher numbers denote greater surprise. It is easy to see that if κ is a ranking on W, then (W,κ) is a plausibility space, where $x\le_{I\!N^*} y$ if and only if $y\le x$ under the usual ordering on the ordinals.
- A *preference ordering* on W is a partial order $\prec$ over W [41,56]. Intuitively, $w\prec w'$ holds if w is *preferred* to w'. Preference orders have been used to provide semantics for *default* (i.e., conditional) statements. In [24] we show how to map preference orders on W to plausibility measures on W in a way that preserves the ordering of events of the form $\{w\}$

as well as the truth values of defaults. We review these results below.

- A *parametrized probability distribution* (PPD) on W is a sequence $\{\Pr_i : i \geq 0\}$ of probability measures over W. Such structures provide semantics for defaults in ϵ-*semantics* [50,30]. In [24] we show how to map PPDs into plausibility structures in a way that preserves the truth-values of conditionals (again, see discussion below).

2.3 The Logic of Conditionals

Our goal is to describe the agent's beliefs in terms of plausibility. To do this, we describe how to evaluate statements of the form $B\phi$ given a plausibility space. In fact, we examine a richer logical language that also allows us to describe how the agent compares different alternatives. This is the logic of conditionals. Conditionals are statements of the form $\phi \rightarrow \psi$, read "given ϕ, ψ is plausible" or "given ϕ, then by default ψ". The syntax of the logic of conditionals is simple: we start with primitive propositions and close off under conjunction, negation and the modal operator $\rightarrow$. The resulting language is denoted $\mathcal{L}^C$.

Many semantics have been proposed in the literature for conditionals. Most of them involve structures of the form (W, X, π), where W is a set of possible worlds, $\pi(w)$ is a truth assignment to primitive propositions, and X is some "measure" on W such as a preference ordering, a κ-ranking, or a possibility measure. We now describe some of the proposals in the literature, and then show how they can be viewed as using plausibility measures. Given a structure (W, X, π), let $[\![\phi]\!] \subseteq W$ be the set of worlds satisfying ϕ.

- A *possibility structure* is a tuple (W, Poss, π), where Poss is a possibility measure on W. It satisfies a conditional $\phi \rightarrow \psi$ if either $\mathrm{Poss}([\![\phi]\!]) = 0$ or $\mathrm{Poss}([\![\phi \wedge \psi]\!]) > \mathrm{Poss}([\![\phi \wedge \neg\psi]\!])$ [13]. That is, either ϕ is impossible, in which case the conditional holds vacuously, or $\phi \wedge \psi$ is more possible than $\phi \wedge \neg\psi$.
- A κ-*structure* is a tuple (W, κ, π), where κ is an ordinal ranking on W. It satisfies a conditional $\phi \rightarrow \psi$ if either $\kappa([\![\phi]\!]) = \infty$ or $\kappa([\![\phi \wedge \psi]\!]) < \kappa([\![\phi \wedge \neg\psi]\!])$ [31].
- A *preferential structure* is a tuple $(W, \prec, \pi)$, where $\prec$ is a partial order on W. The intuition [56] is that a preferential structure satisfies a conditional $\phi \rightarrow \psi$ if all the most preferred worlds (i.e., the minimal worlds according to $\prec$) in $[\![\phi]\!]$ satisfy ψ. However, there may be no minimal worlds in $[\![\phi]\!]$. This can happen if $[\![\phi]\!]$ contains an infinite descending sequence $\ldots \prec w_2 \prec w_1$. What do we do in these structures? There are a number of options: the first is to assume that, for each formula ϕ, there are minimal worlds in $[\![\phi]\!]$; this is the assumption actually made in [41], where it is called the *smoothness* assumption. A yet more general definition—one that works even if $\prec$ is not smooth—is given in [45,4]. Roughly speaking, $\phi \rightarrow \psi$ is true if, from a certain point on, whenever ϕ is true, so is ψ. More formally,

  $(W, \prec, \pi)$ satisfies $\phi \rightarrow \psi$, if for every world $w_1 \in [\![\phi]\!]$, there is a world w_2 such that (a) $w_2 \preceq w_1$ (so that w_2 is at least as normal as w_1), (b) $w_2 \in [\![\phi \wedge \psi]\!]$, and (c) for all worlds $w_3 \prec w_2$, we have $w_3 \in [\![\phi \Rightarrow \psi]\!]$ (so any world more normal than w_2 that satisfies ϕ also satisfies ψ).

It is easy to verify that this definition is equivalent to the earlier one if $\prec$ is smooth.

- A *PPD structure* is a tuple $(W, \{\Pr_i : i \geq 0\}, \pi)$, where $\{\Pr_i\}$ is PPD over W. Intuitively, it satisfies a conditional $\phi \rightarrow \psi$ if the conditional probability ψ given ϕ goes to 1 in the limit. Formally, $\phi \rightarrow \psi$ is satisfied if $\lim_{i\rightarrow\infty} \Pr_i([\![\psi]\!]|[\![\psi]\!]) = 1$ [30] (where $\Pr_i([\![\psi]\!]|[\![\phi]\!])$ is taken to be 1 if $\Pr_i([\![\phi]\!]) = 0$).

In [24] we use plausibility to provide semantics for conditionals and show that our definition generalizes the definition in the various approaches we just described. We briefly review the definitions and results here.

A *plausibility structure* is a tuple $PL = (W, \mathrm{Pl}, \pi)$, where Pl is a plausibility measure on W. Conditionals are evaluated according to a rule that is essentially that used in possibility structures:

- $PL \models \phi \rightarrow \psi$ if either $\mathrm{Pl}([\![\phi]\!]) = \perp$ or $\mathrm{Pl}([\![\phi \wedge \psi]\!]) > \mathrm{Pl}([\![\phi \wedge \neg\psi]\!])$.

Intuitively, $\phi \rightarrow \psi$ holds vacuously if ϕ is impossible; otherwise, it holds if $\phi \wedge \psi$ is more plausible than $\phi \wedge \neg\psi$. It is easy to see that this semantics for conditionals generalizes the semantics of conditionals in possibility structures and κ-structures. The following result shows that it also generalizes the semantics of conditionals in preferential structures and PPD structures.

Proposition 2 [24]

(a) If $\prec$ is a preference ordering on W, then there is a plausibility measure $\mathrm{Pl}_\prec$ on W such that $(W, \prec, \pi) \models \phi \rightarrow \psi$ if and only if $(W, \mathrm{Pl}_\prec, \pi) \models \phi \rightarrow \psi$.

(b) If $PP = \{\Pr_i\}$ is a PPD on W, then there is a plausibility measure Pl_{PP} such that $(W, \{\Pr_i\}, \pi) \models \phi \rightarrow \psi$ if and only if $(W, \mathrm{Pl}_{PP}, \pi) \models \phi \rightarrow \psi$.

We briefly describe the construction of $\mathrm{Pl}_\prec$ and Pl_{PP} here, since we use them in the sequel. Given a preference order $\prec$ on W, let D_0 be the domain of plausibility values consisting of one element d_w for every element $w \in W$. We define a partial order on D_0 using $\prec$: $d_v < d_w$ if $w \prec v$. (Recall that $w \prec w'$ denotes that w is preferred to w'.) We then take D to be the smallest set containing D_0 that is closed under least upper bounds (so that every set of elements in D has a least upper bound in D). For a subset A of W, we can then define $\mathrm{Pl}_\prec(A)$ to be the least upper bound of $\{d_w : w \in A\}$. Since D is closed under least upper bounds, $\mathrm{Pl}(A)$ is well defined. As shown in [24], this choice of $\mathrm{Pl}_\prec$ satisfies Proposition 2.

The construction in the case of PPD's is even more straightforward. Given a PPD $PP = \{Pr_i\}$ on W, we define Pl_{PP} as follows:

$\mathrm{Pl}_{PP}(A) \leq \mathrm{Pl}_{PP}(B)$ if and only if $\lim_{i\rightarrow\infty} \Pr_i(B|A \cup B) = 1$.

A straightforward argument shows that this choice of Pl_{PP} satisfies Proposition 2.

These results show that our semantics for conditionals in plausibility structures generalizes the various approaches examined in the literature. Does it capture our intuitions about conditionals? In the AI literature, there has been discussion of the right properties of default statements (which are essentially conditionals). While there has been little consensus on what the "right" properties for defaults should be, there has been some consensus on a reasonable "core" of inference rules for default reasoning. This core, known as the KLM properties [41], consists of the following axiom and rules of inference:

LLE. From $\phi \Leftrightarrow \phi'$ and $\phi \rightarrow \psi$ infer $\phi' \rightarrow \psi$ (left logical equivalence)

RW. From $\psi \Rightarrow \psi'$ and $\phi \rightarrow \psi$ infer $\phi \rightarrow \psi'$ (right weakening)

REF. $\phi \rightarrow \phi$ (reflexivity)

AND. From $\phi \rightarrow \psi_1$ and $\phi \rightarrow \psi_2$ infer $\phi \rightarrow \psi_1 \wedge \psi_2$

OR. From $\phi_1 \rightarrow \psi$ and $\phi_2 \rightarrow \psi$ infer $\phi_1 \vee \phi_2 \rightarrow \psi$

CM. From $\phi \rightarrow \psi_1$ and $\phi \rightarrow \psi_2$ infer $\phi \wedge \psi_1 \rightarrow \psi_2$ (cautious monotonicity)

LLE states that the syntactic form of the antecedent is irrelevant. Thus, if ϕ_1 and ϕ_2 are equivalent, we can deduce $\phi_2 \rightarrow \psi$ from $\phi_1 \rightarrow \psi$. RW describes a similar property of the consequent: If ψ (logically) entails ψ', then we can deduce $\phi \rightarrow \psi'$ from $\phi \rightarrow \psi$. This allows us to can combine default and logical reasoning. REF states that ϕ is always a default conclusion of ϕ. AND states that we can combine two default conclusions: If we can conclude by default both ψ_1 and ψ_2 from ϕ, we can also conclude $\psi_1 \wedge \psi_2$ from ϕ. OR states that we are allowed to reason by cases: If the same default conclusion follows from each of two antecedents, then it also follows from their disjunction. CM states that if ψ_1 and ψ_2 are two default conclusions of ϕ, then discovering that ψ_1 holds when ϕ holds (as would be expected, given the default) should not cause us to retract the default conclusion ψ_2.

Do conditionals in plausibility structures satisfy the KLM properties? In general, the answer is no. It is almost immediate from the definition that a probability measure Pr is also a plausibility measure. Notice that $\Pr([\![\phi\wedge\psi]\!]) > \Pr([\![\phi\wedge\neg\psi]\!])$ if and only if $\Pr([\![\psi]\!] \mid [\![\phi]\!]) > 1/2$. Expanding the semantics of conditionals, we get that $\phi \rightarrow \psi$ holds in Pr exactly if $\Pr([\![\phi]\!]) = 0$ or $\Pr([\![\psi]\!] \mid [\![\phi]\!]) > 1/2$. It is easy to see that this definition does not satisfy the AND rule: it is not in general the case that $\phi \rightarrow \psi_1$ and $\phi \rightarrow \psi_2$ together imply $\phi \rightarrow (\psi_1 \wedge \psi_2)$, since $\Pr(A_1 \mid B) > 1/2$ and $\Pr(A_2 \mid B) > 1/2$ do not imply $\Pr(A_1 \cap A_2|B) > 1/2$. Since the AND rule is a fundamental feature of qualitative reasoning, we would like to restrict to plausibility structures where it holds. In [24] we show that the following condition is necessary and sufficient to guarantee that the And rule holds:

> **A2** If A, B, and C are pairwise disjoint sets, $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(C)$, and $\mathrm{Pl}(A \cup C) > \mathrm{Pl}(B)$, then $\mathrm{Pl}(A) > \mathrm{Pl}(B \cup C)$.

It turns out that conditionals in plausibility structures that satisfy A2 also satisfy LLE, RW, and CM. They also satisfy OR when one of the conditionals $\phi_1 \rightarrow \psi$ and $\phi_2 \rightarrow \psi$ is satisfied non-vacuously (that is, in a plausibility measure Pl such that either $\mathrm{Pl}([\![\phi_1]\!]) > \perp$ or $\mathrm{Pl}([\![\phi_2]\!]) > \perp$). To satisfy OR in general we need another condition:

A3 If $\mathrm{Pl}(A) = \mathrm{Pl}(B) = \perp$, then $\mathrm{Pl}(A \cup B) = \perp$.

A3 also has a nice axiomatic characterization. Let $N\phi$ be an abbreviation for $\neg\phi \rightarrow \mathit{false}$. (This operator is called the "outer modality" in [45].) Expanding the definition of $\rightarrow$, we get that $N\phi$ holds at w if and only if $\mathrm{Pl}([\![\neg\phi]\!]) = \perp$. Thus, $N\phi$ holds if $\neg\phi$ is considered completely implausible. We can think of the N modality as the plausibilistic version of necessity. It is easy to show that A3 corresponds to an AND rule for N. It holds exactly if $(N\phi \wedge N\psi) \Rightarrow N(\phi \wedge \psi)$.

A plausibility space (W, Pl) is *qualitative* if it satisfies A2 and A3. A plausibility structure (W, Pl, π) is qualitative if (W, Pl) is a qualitative plausibility space. In [24] we show that, in a very general sense, qualitative plausibility structures capture default reasoning. More precisely, we show that the KLM properties are sound with respect to a class of plausibility structures if and only if the class consists of qualitative plausibility structures. We also show that a very weak condition is necessary and sufficient in order for the KLM properties to be complete axiomatization of the language of default entailment considered in [41]. These results help explain why so many different approaches to giving semantics to conditionals are characterized by the KLM properties. In addition, as we shall see, it also shows that if we want belief to have some reasonable properties, then we need to restrict to qualitative plausibility measures.

2.4 Combining Knowledge and Plausibility

We now define a logic that combines knowledge and plausibility. Let $\mathcal{L}^{KC}$ be the language obtained by starting with primitive propositions, and closing off under conjunction, negation, and the operators K_i and $\rightarrow_i$, $i = 1, \ldots, n$. Note that we have a different conditional operator for each agent. We read $\phi \rightarrow_i \psi$ as "according to agent i's plausibility measure, ϕ typically implies ψ".

A *(Kripke) structure (for knowledge and plausibility)* is a tuple $(W, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{P}_1, \ldots, \mathcal{P}_n)$ where W, π and $\mathcal{K}_i$ are just as in Kripke structures for knowledge, while $\mathcal{P}_i$ is a *plausibility assignment*, a function that assigns a plausibility space to agent i at each world. Intuitively, the structure $\mathcal{P}_i(w) = (W_{(w,i)}, \mathrm{Pl}_{(w,i)})$ captures agent i's plausibility measure in the world w. For now we allow $W_{(w,i)}$ to be an arbitrary subset of W. We discuss some possible restrictions on $W_{(w,i)}$ below. It is reasonable to ask at this point where the plausibility spaces $\mathcal{P}_i(w)$ are coming from, and why we need a different one for each agent at each world. The answer to this question depends very much on the intended application. We defer further discussion of this issue until later.

We can now give semantics to formulas in $\mathcal{L}^{KC}$ in Kripke structures for knowledge and plausibility. This is done in a recursive way using the rules specified above for $\mathcal{L}^{K}$ and $\mathcal{L}^{C}$. Statements of the form $K_i\phi$ are evaluated according to $\mathcal{K}_i$:

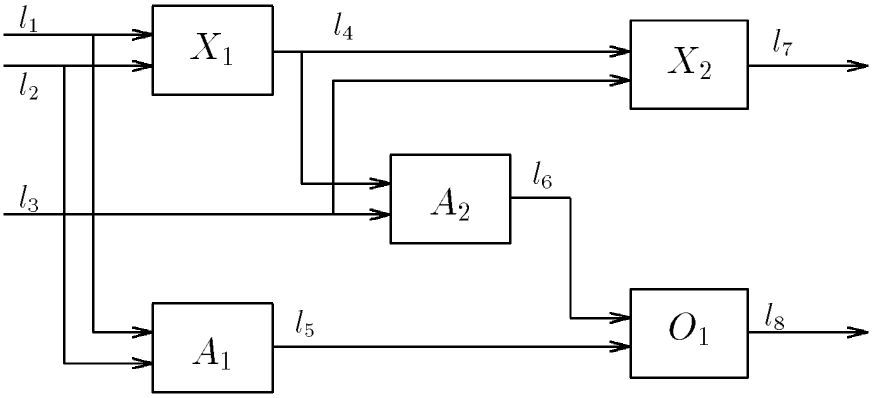

Fig. 1. A full adder. X_1 and X_2 are XOR gates, A_1 and A_2 are AND gates, and O_1 is an OR gate.

– $(M, w) \models K_i\phi$ if $(M, w') \models \phi$ for all $w' \in \mathcal{K}_i(w)$.

Statements of the form $\phi \rightarrow_i \psi$ are evaluated according to $\mathcal{P}_i$. Let $[\![\phi]\!]_{(w,i)} = \{w' \in W_{(w,i)} : (M, w') \models \phi\}$.

– $(M, w) \models \phi \rightarrow_i \psi$ if either $\mathrm{Pl}_{(w,i)}([\![\phi]\!]_{(w,i)}) = \perp$ or $\mathrm{Pl}_{(w,i)}([\![\phi \wedge \psi]\!]_{(w,i)}) > \mathrm{Pl}_{(w,i)}([\![\phi \wedge \neg\psi]\!]_{(w,i)})$.

We now define beliefs. Recall that $true \rightarrow_i \phi$ means that ϕ is more plausible than $\neg\phi$ according to agent's i plausibility measure. We might say that in this case the agent believes ϕ. However, recall that the agent can have different plausibility assessments at different worlds. Thus, there can be a model M, and worlds w, w' such that $(w, w') \in \mathcal{K}_i$, but $(M, w) \models true \rightarrow_i \phi$ while $(M, w') \models \neg(true \rightarrow_i \phi)$. (In Example 5, we show why this extra expressive power is necessary.) That is, ϕ is more plausible than $\neg\phi$ in one of the worlds the agent considers possible, but not in another. Since our intention is that the agent should not distinguish between accessible worlds, we would like the agent to have the same beliefs in all the worlds he considers possible. We say that an agent *believes* ϕ if he knows that ϕ is more plausible than $\neg\phi$ in all the worlds he considers possible. Thus, we define $B_i\phi$, read "agent i believes ϕ", as an abbreviation for $K_i(true \rightarrow_i \phi)$.

2.5 Example: Circuit Diagnosis

The following example illustrates some of the expressive power of this language. Although it only involves one agent and only one plausibility measure in any given structure, it can easily be extended to allow for many agents with different plausibility measures.

The circuit diagnosis problem has been well studied in the literature (see [8] for an overview). Consider a circuit that contains n logical components $c_1, \ldots, c_n$ and k lines $l_1, \ldots, l_k$. As a concrete example, consider the circuit of Figure 1.[3] The diagnosis task is to identify which

[3] The "full adder" example is often used in the diagnosis literature. In our discussion here we

components are faulty. The agent can set the values of input lines of the circuit and observe the output values. The agent then compares the actual output values to the expected output values and attempts to locate faulty components.

We model this situation using the tools we presented in the previous sections. We start by describing the agent's knowledge using a Kripke structure. We then construct two possible plausibility measures over worlds in this Kripke structures, and examine the resulting knowledge and belief.

Knowledge We model the agent's knowledge about the circuit using the Kripke structure $M^K_{diag} = (W_{diag}, \pi_{diag}, \mathcal{K}_{diag})$. Each possible world $w \in W_{diag}$ is composed of two parts: $\mathit{fault}(w)$, the *failure set*—that is, the set of faulty components in w, and $\mathit{value}(w)$, the value of all the lines in the circuit. We consider only worlds where the components that are not in the failure sets perform as expected. For example, in the circuit of Figure 1, if the AND gate A_1 is not faulty, then we require that l_5 has value "high" if and only if both l_1 and l_2 have the value "high". Most accounts of diagnosis assume that there is a logical theory Δ that describes the properties of the device. To capture our intuition, it must be the case that w is a possible world in M if and only if $\mathit{fault}(w)$ and $\mathit{value}(w)$ are together consistent with Δ.

The most straightforward language for reasoning about faults is the following: let $\Phi_{diag} = \{\mathit{faulty}(c_1), \ldots, \mathit{faulty}(c_n), \mathit{hi}(l_1), \ldots, \mathit{hi}(l_k)\}$ be the set of propositions, where each $\mathit{faulty}(c_i)$ denotes that component i is faulty and $\mathit{hi}(l_i)$ denotes that line i in a "high" state. We then define the interpretation π_{diag} in the obvious way: $\pi_{diag}(w)(\mathit{faulty}(c_i)) = \mathbf{true}$ if $c_i \in \mathit{fault}(w)$, and $\pi_{diag}(w)(\mathit{hi}(l_i)) = \mathbf{true}$ if $\langle l_i, 1\rangle \in \mathit{value}(w)$.

Next, we need to define the agent's knowledge. We define $o_w \subseteq \mathit{value}(w)$ to be the values of those lines the agent sets or observes. The agent knows which tests he has performed and the results he observed. Therefore, we have $(w, w') \in \mathcal{K}_{diag}$ if $o_w = o_{w'}$. For example, suppose the agent observes $\mathit{hi}(l_1) \wedge \mathit{hi}(l_2) \wedge \mathit{hi}(l_3) \wedge \mathit{hi}(l_7) \wedge \mathit{hi}(l_8)$. The agent then considers possible all worlds where the same observations hold. Since these observations are consistent with the correct behavior of the circuit, one of these worlds has an empty failure set. However, other worlds are possible. For example, it might be that the AND gate A_2 is faulty. This would not affect the outputs in this case, since if A_1 is non-faulty, then its output is "high", and thus, O_1's output is "high" regardless of A_2's output.

Now suppose that the agent observes $\mathit{hi}(l_1) \wedge \neg \mathit{hi}(l_2) \wedge \mathit{hi}(l_3) \wedge \mathit{hi}(l_7) \wedge \neg \mathit{hi}(l_8)$. These observations imply that the circuit is faulty. (If l_1 and l_3 are "high" and l_2 is "low", then the correct values for l_7 and l_8 should be "low" and "high", respectively.) In this case there are several possible failure sets, including $\{X_1\}$, $\{X_2, O_1\}$, and $\{X_2, A_2\}$.

In general, there is more than one explanation for the observed faulty behavior. Thus, the agent can not *know* exactly which components are faulty, but he may have *beliefs* on that

loosely follow the examples of Reiter [54].

score.

Plausibility To model the agent's beliefs, we need to decide on the plausibility measure the agent has at any world. We assume that only failure sets are relevant for determining a world's plausibility. Thus, we start by constructing a plausibility measure over possible failures of the circuit. We assume that failures of individual components are independent of one another. If we also assume that the likelihood of each component failing is the same, we can construct a preference ordering on failure set as follows: If f_1 and f_2 are two failure sets, we say that f_1 is preferred to f_2 if $|f_1| < |f_2|$, that is, if f_1 consists of fewer faulty components than f_2. This preference ordering induces a plausibility measure using the construction of Proposition 2. In this measure $\mathrm{Pl}(F_1) < \mathrm{Pl}(F_2)$ if $\min_{f \in F_1}(|f|) < \min_{f \in F_2}(|f|)$.

We can construct the same plausibility measure based on probabilistic arguments using PPDs. Suppose that the probability of a single component failing is ϵ. Since we have assumed that failures are independent, it follows that the probability of a failure set f is $\epsilon^{|f|}(1-\epsilon)^{(n-|f|)}$, since there are $|f|$ components that fail, and $n-|f|$ components that do not fail. To model the behavior of small but unknown failure probability, we can consider the PPD $(\mathrm{Pr}_0, \mathrm{Pr}_1, \ldots)$, where in Pr_m the probability of a single failure is $1/(m+1)$. It is not hard to check that $\lim_{m\to\infty} \mathrm{Pr}_m(F_2)/\mathrm{Pr}_m(F_1) = 0$ if and only if $\mathrm{Pl}(F_2) < \mathrm{Pl}(F_1)$ in the plausibility measure described above. Interestingly, this plausibility measure is almost identical to the κ-ranking in which $\kappa(\{f\}) = |f|$. The only difference is that if $|f_1| = |f_2|$, $\mathrm{Pl}(\{f_1\})$ is incomparable to $\mathrm{Pl}(\{f_2\})$ in the plausibility measure we constructed, while they are equal according to the κ-ranking.

In some situations it might be unreasonable to assume that all components have equal failure probability. Thus, we might assume that for each component c_i there is a probability ϵ_i of failure. If we assume independence, then given $\vec{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$, the probability of a failure set f is $\Pi_{c_i \in f}\epsilon_i \Pi_{c_i \notin f}(1-\epsilon_i)$. We can construct a PPD that captures the effect of the ϵ_i's getting smaller, but at possibly different rates: Suppose g is a bijection from $\mathbb{N}^m$ to $\mathbb{N}$. If $\vec{m} = (m_1, \ldots, m_n)$, let $\mathrm{Pr}_{g(\vec{m})}$ be the distribution where the probability of c_i failing is $1/(m_i+1)$, for $i = 1, \ldots, n$. In this case, we get that $\lim_{m\to\infty} \mathrm{Pr}_m(f_2)/\mathrm{Pr}_m(f_1) = 0$ if and only if f_2 is a strict subset of f_1, i.e., if f_1 contains all the components in f_2 and more. Since we do not assume any relations among the failure probabilities of different components, it is not possible to compare failure sets unless one is a subset of the other. Thus, we can define $f \prec f'$ if $f \subset f'$. Using the construction of Proposition 2, we can again consider the plausibility measure Pl induced by $\prec$. It is not hard to see that $\mathrm{Pl}(F_1) \le \mathrm{Pl}(F_2)$ if for every failure set $f_1 \in F_1 - F_2$ there is some $f_2 \in F_2$ such that $f_2 \prec f_1$. As our construction shows, this plausibility measure can be induced by either a preference ordering or a PPD; however, it cannot be captured by a κ-ranking or a possibility measure, since the ordering on failure sets is partial.

Beliefs We now have the required components to examine the agent's beliefs. Using the two plausibility measures we just described, we can construct two possible structures $M_{diag,1}$

and $M_{diag,2}$. In both structures we set $W_{(w,1)} = \mathcal{K}_{diag}(w)$, and in both $M_{diag,1}$ and $M_{diag,2}$ the plausibility measure is induced from a preference ordering on failures (using the construction of Proposition 2). In $M_{diag,1}$, we take the plausibility measure to be such that $\mathrm{Pl}_{(w,1)}(\{w\}) \geq \mathrm{Pl}_{(w,1)}(\{w'\})$ if and only if $|\mathit{fault}(w)| \leq |\mathit{fault}(w')|$, and in $M_{diag,2}$ so that $\mathrm{Pl}_{(w,1)}(\{w\}) \geq \mathrm{Pl}_{(w,1)}(\{w'\})$ if and only if $\mathit{fault}(w) \subseteq \mathit{fault}(w')$. It is easy to see that, in both structures, if there is a world w in which these observations occur and where $\mathit{fault}(w) = \emptyset$, then the agent believes that the circuit is faultless. If the agent detects an error, he believes that it is caused by one of the *minimal explanations* of his observations, where the notion of minimality differs in the two structures. We now make this statement more precise. Let f be a failure set. Let D_f be the formula that denotes that f is the failure set, so that $(M, w) \models D_f$ if and only if $\mathit{fault}(w) = f$. The agent believes that f is a possible *diagnosis* (i.e., an explanation of his observations) if $\neg B_1 \neg D_f$. The set of diagnoses the agent considers possible is $\mathrm{Bel}(M, w) = \{f : (M, w) \models \neg B_1 \neg D_f\}$. We say that a failure set f is *consistent* with an observation o if it is possible to observe o when f occurs, i.e., if there is a world w in W such that $\mathit{fault}(w) = f$ and $o_w = o$.[4]

Proposition 3 *(a) $Bel(M_{diag,1}, w)$ contains all failure sets f that are consistent with o_w such that there is no failure set f' with $|f'| < |f|$ which is consistent with o_w.*

(b) $Bel(M_{diag,2}, w)$ contains all failure sets f that are consistent with o_w such that there is no failure set f' with $f' \subset f$ which is consistent with o_w.

Proof. Straightforward; left to the reader. $\square$

Thus, both $\mathrm{Bel}(M_{diag,1}, w)$ and $\mathrm{Bel}(M_{diag,2}, w)$ consist of minimal sets of failure sets consistent with o_w, for different notions of minimality. In the case of $M_{diag,1}$, "minimality" means "of minimal cardinality", while in the case of $M_{diag,2}$, it means "minimal in terms of set containment". This proposition shows that $M_{diag,1}$ and $M_{diag,2}$ capture standard assumptions made in model-based diagnosis; $M_{diag,1}$ captures the assumptions made in [10], while $M_{diag,2}$ captures the assumptions made in [54]. More concretely, in our example, if the agent observes $hi(l_1) \wedge \neg hi(l_2) \wedge hi(l_3) \wedge hi(l_7) \wedge \neg hi(l_8)$, then in $M_{diag,1}$ she would believe that X_1 is faulty, since $\{X_1\}$ is the only diagnosis with cardinality one. On the other hand, in $M_{diag,2}$ she would believe that one of the three minimal diagnoses occurred: $\{X_1\}$, $\{X_2, O_1\}$ or $\{X_2, A_2\}$.

2.6 Properties of Knowledge and Plausibility

Kripke structures for knowledge and plausibility are quite similar to the Kripke structures for knowledge and probability introduced by Fagin and Halpern [14]. The only difference is that in Kripke structures for knowledge and probability, $\mathcal{P}_i(w)$ is a probability space rather than a

[4] Note that if Δ is a theory that describes the properties of circuit, then a failure f is consistent with observation o, if and only if f and o are consistent according to Δ.

plausibility space. Fagin and Halpern explore various natural restrictions on the interactions between the probability spaces $\mathcal{P}_i(w)$ and the accessibility relations $\mathcal{K}_i$. Here we investigate restrictions on the interaction between the plausibility spaces and the accessibility relations. Not surprisingly, some of these conditions are exact analogues to conditions investigated by Fagin and Halpern.

Given our interest in the KLM properties, we will be interested in structures that satisfy the following condition:

QUAL $\mathcal{P}_i(w)$ is qualitative for all worlds w and agents i.

The same arguments that show that A2 gives us the AND rule also show that it gives us property K2 for beliefs. More precisely, we have the following result.

Theorem 4 *If M satisfies QUAL, then for all worlds w in M, we have*

(a) $(M,w) \models ((\sigma \rightarrow_i \phi) \wedge (\sigma \rightarrow_i \psi)) \Rightarrow (\sigma \rightarrow_i (\phi \wedge \psi))$
(b) $(M,w) \models B_i\phi \wedge B_i\psi \Rightarrow B_i(\phi \wedge \psi)$
(c) $(M,w) \models B_i\phi \wedge B_i(\phi \wedge \psi) \Rightarrow B_i\psi$.

Proof. Straightforward; left to the reader. $\square$

In view of this result, we typically assume that QUAL holds whenever we want to reason about belief.

The set $W_{(w,i)}$ consists of all worlds to which agent i assigns some degree of plausibility in world w. We would not expect the agent to place a positive probability on worlds that he considers impossible. Similarly, he would not want to consider as plausible (even remotely) a world he knows to be impossible. This intuition leads us to the following condition, called CONS for *consistency* (following [14]):

CONS $W_{(w,i)} \subseteq \mathcal{K}_i(w)$ for all worlds w and all agents i. [5]

A consequence of assuming CONS is a stronger connection between knowledge and belief. Since CONS implies that the most plausible worlds are in $\mathcal{K}_i(w)$, it follows that if the agent knows ϕ he also believes ϕ. (Indeed, as we shall see, this condition characterizes CONS.)

[5] We remark that CONS is inappropriate if we use $\rightarrow$ to model, not plausibility, but counterfactual conditions, as is done by Lewis [45]. If CONS holds, then it is easy to see that $K_i\phi \Rightarrow K_i(\neg\phi \rightarrow_i \psi)$ is valid, for all ψ. That is, if agent i knows ϕ, then he knows that in the most plausible worlds where $\neg\phi$ is true, ψ is vacuously true, because there are no plausible worlds where $\neg\phi$ is true. On the other hand, under the counterfactual reading, it makes perfect sense to say "I know the match is dry, but it is not the case that if it were wet, then it would light if it were struck."

In probability theory, the agent assigns probability 1 to the set of all worlds. Since $1 > 0$, this means the agent assigns non-zero probability to some sets of worlds. It is possible to have $\top = \perp$ in plausibility spaces. If this happens, the agent considers all sets to be completely implausible. The following condition, called NORM for *normality* (following [45]), says this does not happen:

NORM $\mathcal{P}(w, i)$ is normal, that is, $\top_{(w,i)} >\perp_{(w,i)}$, for all worlds w and all agents i.

We can strengthen this condition somewhat to one that says that the agent never considers the real world implausible. This suggests the following condition: $\mathrm{Pl}_{(w,i)}(\{w\}) >\perp$. Stating this condition, however, leads to a technical problem. Recall that $\mathrm{Pl}_{(w,i)}$ is defined over the set of measurable subsets of $W_{(w,i)}$. In general, however, singletons may not be measurable. Thus, we examine a slightly weaker condition which we call REF for *reflexive* (following [45]):

REF For all worlds w and all agents i,
- $w \in W_{(w,i)}$, and
- $\mathrm{Pl}_{(w,i)}(A) >\perp$ for all $A \in \mathcal{F}_{(w,i)}$ such that $w \in A$.

As we said in the introduction, much of the previous work using conditionals assumed (implicitly or explicitly) that the agent considers only one plausibility measure possible. This amounts to assuming that the plausibility measure is a function of the agent's epistemic state. This is captured by an assumption called SDP (following [14]) for *state determined plausibilities*:

SDP For all worlds w and w' and all agents i, if $(w, w') \in \mathcal{K}_i$ then $\mathcal{P}_i(w) = \mathcal{P}_i(w')$.

It is easy to see that SDP implies that an agent knows his plausibility measure. In particular, as we shall see, with SDP we have that $\phi \rightarrow_i \psi$ implies $K_i(\phi \rightarrow_i \psi)$.

It is easy to verify that the structures described in the diagnosis example of Section 2.5 satisfy CONS, REF, and SDP. As mentioned in the introduction, SDP is not appropriate in all situations; at times we may want to allow the agent to consider possible several plausibility measures. To capture this, we need to generalize SDP. The following example might help motivate the formal definition.

Example 5 *This is a variation of the Liar's Paradox. On a small Pacific island there are two tribes, the Rightfeet and the Leftfeet. The Rightfeet are known to usually tell the truth, while the Leftfeet are known to usually lie. Alice is a visitor to the island. She encounters a native, Bob, and discusses with him various aspects of life on the island. Now, Alice does not know to what tribe Bob belongs. Thus, she considers it possible both that Bob is a Rightfoot and that he is a Leftfoot. In the first case, she should believe what he tells her and in the second she should be skeptical.*

One possible way of capturing this situation is by partitioning the worlds Alice considers possible into two sets, according to Bob's tribe. Let W_R (resp. W_L) be the set of worlds *that Alice considers possible where Bob is a Rightfoot (resp. Leftfoot). As the discussion above suggests, Alice's plausibility measure at the worlds of W_R gives greater plausibility to worlds where Bob is telling the truth than to worlds where Bob is lying; the opposite situation holds at worlds of W_L. In such a structure, the formula $\neg K_{Alice}\neg(tell(\phi) \rightarrow_{Alice} \neg\phi) \wedge \neg K_{Alice}\neg(tell(\phi) \rightarrow_{Alice} \phi)$ is satisfiable, where $tell(\phi)$ is the formula that holds when Bob tells Alice ϕ. On the other hand, in structures satisfying SDP, this formula is satisfiable only when $tell(\phi)$ has plausibility $\perp$ in all the worlds that Alice considers possible.*

While this example may seem contrived, in many situations it is possible to extract parameters such as Leftfoot and Rightfoot that determine which conditional statements are true. For example, when we introduce time into the picture (in Section 3.1), these parameters might be the agent's own actions in the future. Such a partition allows us to make statements such as "I do not know whether ϕ is plausible or not, but I know that if I do a, then ϕ is plausible", where ϕ is some statement about the future. If the agent does not know the value of these parameters, she will not necessarily know which conditionals are true at a given world (as was the case in the example above).

Example 5 motivates the condition called *uniformity*.

UNIF For all worlds w and agents i, if $w' \in W_{(w,i)}$ then $\mathcal{P}_i(w) = \mathcal{P}_i(w')$. [6]

It is not hard to show that UNIF holds if and only if, for each agent i, we can partition the set of possible worlds in such a way that for each cell C in the partition, there is a plausibility space (W_C, Pl_C) such that $W_C \subseteq C$ and $\mathcal{P}_i(w) = (W_C, \mathrm{Pl}_C)$ for all worlds $w \in C$. Moreover, if CONS also holds, then this partition refines the partition induced by the agent's knowledge, i.e., if C is a cell in the partition and w is some world C, then $C \subseteq \mathcal{K}_i(w)$. It easily follows that SDP and CONS together imply UNIF.

When we model uncertainty about the relative plausibility of different worlds this way it is reasonable to demand that the plausibility measure totally orders all events; i.e., it is a *ranking*. The RANK assumption is:

RANK For all worlds w and agents i, $\mathcal{P}_i(w)$ is a ranking, that is, for all sets $A, B \subseteq W_w$ either $\mathrm{Pl}_w(A) \leq \mathrm{Pl}_w(B)$ or $\mathrm{Pl}_w(B) \leq \mathrm{Pl}_w(A)$, and $\mathrm{Pl}_w(A \cup B) = \max(\mathrm{Pl}_w(A), \mathrm{Pl}_w(B))$.

Note that κ-rankings and possibility measures are two examples of rankings. Additionally, *rational* preference orderings of [41] are essentially rankings in the sense that for each rational preference ordering we can construct a ranking that satisfies exactly the same conditional statements [17,24].

While rankings are quite natural, they have often been rejected as being too inexpressive [29]. In a ranking there is a total order on events. The standard argument for partial orders is as follows: In general, an agent may not be able to determine the relative plausibility of a and b.

[6] This condition is not the same as uniformity as defined in [45]; rather, it corresponds in the Lewis terminology to absoluteness.

If the plausibility measure is a ranking, the agent is forced to make this determination; with a partial order, he is not. This argument loses much of its force in our framework, once we combine knowledge and plausibility. As we said above, the agent's ignorance can be modeled by allowing him to consider (at least) two rankings possible, one in which a is more plausible than b, and one in which b is more plausible that a. The agent then believes neither that a is more plausible than b nor that b is more plausible than a.

2.7 Knowledge and Belief

How reasonable is the notion of belief we have defined? In this section, we compare it to other notions considered in the literature.

Recall that $\mathcal{L}^B$ be the language where the only modal operators are $B_1, \ldots, B_n$. Let $\mathcal{L}^{KB}$ be the language where we have $K_1, \ldots, K_n$ and $B_1, \ldots, B_n$ (but no $\rightarrow_i$ operators). It is not hard to see (and will follow from our proofs below) that to get belief to satisfy even minimal such as K2, we need the AND rule to hold. Thus, in this section, we restrict attention to Kripke structures for knowledge and plausibility that satisfy QUAL. We then want to investigate the impact of adding additional assumptions. Let $\mathcal{M}$ be the set of all Kripke structures for knowledge and plausibility that satisfy QUAL, and let $\mathcal{M}^{\mathrm{CONS}}$ (resp. $\mathcal{M}^{\mathrm{CONS,NORM}}$) be the structures satisfying QUAL and CONS (resp. QUAL, CONS and NORM).

Work on belief and knowledge in the literature [34,37,44] has focused on the modal systems S5, KD45, K45, and K with semantics based on Kripke structures as described in Section 2.1. Before we examine the properties of belief in our approach, we relate our semantics of belief (in terms of plausibility) to the more standard Kripke approach, which presumes that belief is defined in terms of a binary relation $\mathcal{B}_i$. Can we define a relation $\mathcal{B}_i$ in terms of $\mathcal{K}_i$ and $\mathcal{P}_i$ such that $(M, w) \models B_i\phi$ if and only if $(M, v) \models \phi$ for all $v \in \mathcal{B}_i(w)$? We show that this is possible in some structures, but not in general.

Let $S = (W, \mathrm{Pl})$ be a qualitative plausibility space. We say that $A \subseteq W$ is a set of *most plausible* worlds if $\mathrm{Pl}(A) > \mathrm{Pl}(\overline{A})$ (where $\overline{A}$ is the complement of A, i.e., $W - A$) and for all $B \subset A$, $\mathrm{Pl}(B) \not> \mathrm{Pl}(\overline{B})$. That is, A is a minimal set of worlds that is more plausible than its complement. It is easy to verify that if such a set exists, then it must be unique. To see this, suppose that A and A' are both sets of most plausible worlds. We now show that $\mathrm{Pl}(A \cap A') > \mathrm{Pl}(\overline{A \cap A'})$. Since A and A' are both most plausible sets of worlds, this will show that we must have $A = A'$. To see that $\mathrm{Pl}(A \cap A') > \mathrm{Pl}(\overline{A \cap A'})$, first note that $A \cap A'$, $A - A'$ and $\overline{A}$ are pairwise disjoint. Since A and A' are most plausible sets of worlds, we have that $\mathrm{Pl}((A \cap A') \cup (A - A')) = \mathrm{Pl}(A) > \mathrm{Pl}(\overline{A})$ and $\mathrm{Pl}((A \cap A') \cup \overline{A}) \geq \mathrm{Pl}((A \cap A') \cup (A' - A)) = \mathrm{Pl}(A') > \mathrm{Pl}(\overline{A'}) \geq \mathrm{Pl}(A - A')$. We can apply A2 to get that $\mathrm{Pl}(A \cap A') > \mathrm{Pl}((A - A') \cup \overline{A}) = \mathrm{Pl}(\overline{A \cap A'})$.

In finite plausibility structures (that is, ones with only finitely many worlds), it is easy to see that there is always a (unique) set of most plausible worlds. In general, however, a set of most plausible worlds does not necessarily exist. For example, consider the space $S_0 = (W, \mathrm{Pl})$, where $W = \{w_i : i \geq 0\}$ and Pl is defined as follows: $\mathrm{Pl}(A) = \infty$ if A contains an infinite number of worlds, and $\mathrm{Pl}(A) = \max_{w_i \in A}(i)$ otherwise. Suppose that $\mathrm{Pl}(A) > \mathrm{Pl}(\overline{A})$. $\overline{A}$ must be finite, for otherwise $\mathrm{Pl}(\overline{A}) = \infty$. Thus, A must be infinite. Suppose $w_i \in A$. It is easy to see that $A - \{w_i\}$ is infinite and $\overline{A - \{w_i\}}$ is finite. Thus, $\mathrm{Pl}(A - \{w_i\}) > \mathrm{Pl}(\overline{A - \{w_i\}})$. This shows that there does not exist a set of most plausible worlds in S.

If there is no set of most plausible worlds, then we may not be able to find a relation $\mathcal{B}_i$ that characterizes agent i's beliefs. For example, consider the structure $M = (W, \pi, \mathcal{K}_1, \mathcal{P}_1)$, where $W = \{w_i : i \geq 0\}$ is the set of worlds described in S_0 above; π assigns truth values to primitive propositions $p_1, p_2, \ldots$ in such a way that $\pi(w_i)(p_j) = \mathbf{true}$ if and only if $j \geq i$; $\mathcal{K}_1$ is the complete accessibility relation $\mathcal{K}_1 = W \times W$; and $\mathcal{P}_1(w_i)$ is the space S_0 described above. It is not hard to verify that $(M, w_0) \models B_1\phi$ if and only if $[\![\neg\phi]\!]_{(w_0,i)}$ is a finite set, i.e., there is an index i such that for all $j \geq i$, we have $(M, w_j) \models \phi$. Thus, $(M, w_0) \models B_1 p_j$ for all $j \geq 0$. Yet there are no worlds in the model that satisfy all the propositions p_j at once. Thus, there is no accessibility relation $\mathcal{B}_1$ that characterizes agent 1's beliefs in w_0.

On the other hand, we can show that if there is always a set of most plausible worlds, then we can characterize the agents' beliefs by an accessibility relation. Let $S = (W, \mathrm{Pl})$ be a plausibility space. Define $\mathrm{MP}(S)$ to be the set of most plausible worlds in S if it exists, and $\emptyset$ if $\mathrm{Pl}(W) = \bot$. Otherwise $\mathrm{MP}(S)$ is not defined.

Proposition 6 *Let M be a Kripke structure for knowledge and plausibility. If $MP(\mathcal{P}_i(w'))$ is defined for all $w' \in \mathcal{K}_i(w)$, then $(M, w) \models B_i\phi$ if and only if $(M, w'') \models \phi$ for all $w'' \in \cup_{w' \in \mathcal{K}_i(w)} MP(\mathcal{P}_i(w'))$.*

Proof. Straightforward; left to the reader. □

This proposition implies that, if most plausible sets of worlds always exist in M, then we can set $\mathcal{B}_i(w) = \cup_{w' \in \mathcal{K}_i(w)} \mathrm{MP}(\mathcal{P}_i(w'))$ and recover the usual Kripke-style semantics for belief.

This discussion shows that our model of belief is more general than the classical Kripke-structure account of beliefs, since there are models where the agent's beliefs are not determined by a set of accessible worlds. However, as we shall see, this does not lead to new properties of beliefs in $\mathcal{L}^B$. Roughly speaking, this is because we have a finite model property: a formula in $\mathcal{L}^B$ is satisfiable if and only if it is satisfiable in a finite model (see Theorem 13 below). It is easy to verify that in a finite model $\mathrm{MP}(\mathcal{P}_i(w))$ is always defined. We note, however, that this finite model property is no longer true when we consider the interaction of beliefs with other modalities, such as time, or when we examine the first-order case. In these situations, the two models of beliefs are not equivalent. Plausibility is strictly more expressive; see [25].

We now examine the formal properties of belief and knowledge in structures of knowledge and plausibility. We start by restricting our attention to $\mathcal{L}^B$. As we show below, the modal system K precisely characterizes the valid formulas of $\mathcal{L}^B$ in the class $\mathcal{M}$. However, in the literature, belief has typically been taken to be characterized by the modal system K45 or KD45, not K. We get K45 by restricting to models that satisfy CONS, and KD45 by further restricting to models that satisfy NORM. Thus, the two requirements that are most natural, at least if we have a probabilistic intuition for plausibility, are already enough to make B_i a KD45 operator.

Theorem 7 *K (resp., K45, KD45) is a sound and complete axiomatization for $\mathcal{L}^B$ with respect to $\mathcal{M}$ (resp., $\mathcal{M}^{\mathrm{CONS}}$, $\mathcal{M}^{\mathrm{CONS,NORM}}$).*

Proof. See Appendix A.1. □

We now consider knowledge and belief together. This combination has been investigated in the literature [40,58]. In particular, Kraus and Lehmann [40] define Kripke structures for knowledge and belief that have two accessibility relations, one characterizing the worlds that are knowledge-accessible and one characterizing worlds that are belief-accessible. K_i and B_i are defined, as usual, in terms of these relations. They argue that the two accessibility relations must be coherent in the sense that the agent knows what she believes and believes what she knows to be true. Kraus and Lehmann describe restrictions on the interaction between the two relations that force this coherence. They show that in the resulting structures, the interactions between knowledge and belief are characterized by the following axioms.

KB1. $B_i\phi \Rightarrow K_iB_i\phi$

KB2. $K_i\phi \Rightarrow B_i\phi$

It turns out that KB1 holds in $\mathcal{M}$ and KB2 is a consequence of CONS. To see this, recall that $B_i\phi \equiv K_i(\mathit{true} \to \phi)$. Using positive introspection for knowledge (axiom K4), we derive that $B_i\phi \Rightarrow K_iK_i(\mathit{true} \to \phi)$. This is equivalent to axiom KB1. When M satisfies CONS, we have that $W_{(w,i)} \subseteq \mathcal{K}_i(w)$. If $(M,w) \models K_i\phi$, then all worlds in $\mathcal{K}_i(w)$ satisfy ϕ. This implies that there are no worlds satisfying $\neg\phi$ in $W_{(w,i)}$, and thus $B_i\phi$ must hold. Thus, KB2 must hold.

We now state this formally. Let $\mathrm{AX}^{\mathrm{KB}}$ consist of the S5 axioms for the operators K_i, the K axioms for the operators B_i, together with KB1; let $\mathrm{AX}^{\mathrm{KB,CONS}}$ consist of $\mathrm{AX}^{\mathrm{KB}}$ together with the K4 and K5 axioms for B_i and KB2; and let $\mathrm{AX}^{\mathrm{KB,CONS,NORM}}$ consist of $\mathrm{AX}^{\mathrm{KB,CONS}}$ together with the K6 axiom for B_i.

Theorem 8 *AX^{KB} (resp., $AX^{\mathrm{KB,CONS}}$, $AX^{\mathrm{KB,CONS,NORM}}$) is a sound and complete axiomatization of $\mathcal{L}^{KB}$ with respect to $\mathcal{M}$ (resp., $\mathcal{M}^{\mathrm{CONS}}$, $\mathcal{M}^{\mathrm{CONS,NORM}}$).*

Proof. See Appendix A.1. □

As an immediate corollary, we get that there is a close relationship between our framework and that of [40]. Let KL be the logic of Kraus and Lehmann:

Corollary 9 *For any $\phi \in \mathcal{L}^{KB}$, $KL \models \phi$ if and only if $\mathcal{M}^{\mathrm{CONS,NORM}} \models \phi$.*

We now relate to three other notions of beliefs in the literature—those of Moses and Shoham [49], Voorbraak [58], and Lamarre and Shoham [43].

Moses and Shoham [49] also view belief as being derived from knowledge. The intuition that they try to capture is that once the agent makes a defeasible assumption, the rest of his beliefs should follow from his knowledge. In this sense, Moses and Shoham can be viewed as focusing on the implications of an assumption and not on how it was obtained. We can understand their notion as saying that ϕ is believed if it is known to be true in the most plausible worlds. But for them, plausibility is not defined by an ordering. Rather, it is defined in terms of a formula, which can be thought of as characterizing the most plausible worlds. More formally, for a fixed formula α, they define $B_i^\alpha \phi$ to be an abbreviation for $K_i(\alpha \Rightarrow \phi)$.[7] The following result relates our notion of belief to that of Moses and Shoham.

Lemma 10 *Let M be a propositional Kripke structure of knowledge and plausibility satisfying CONS and SDP. Suppose that w, i, and α are such that the most plausible worlds in $\mathcal{P}_i(w)$ are exactly those worlds in $\mathcal{K}_i(w)$ that satisfy α, i.e., $MP(\mathcal{P}_i(w)) = \{w' \in \mathcal{K}_i(w) : (M, w') \models \alpha\}$. Then for any formula $\phi \in \mathcal{L}^{KB}$ that includes only the modalities K_i and B_i, $(M, w) \models \phi$ if and only if $(M, w) \models \phi^*$, where ϕ^* is the result of recursively replacing each subformula of the form $B_i\psi$ in ϕ by $K_i(\alpha \Rightarrow \psi^*)$.*

Proof. See Appendix A.1. $\square$

Voorbraak [58] distinguishes two notions of knowledge: *objective knowledge* and *true justified belief*. He then studies the interaction of both notions of knowledge with beliefs. The intuition we assign to knowledge is similar to Voorbraak's intuition for objective knowledge. However, Voorbraak objects to the axiom $K_i\phi \Rightarrow B_i\phi$, and suggests $B_i\phi \Rightarrow B_iK_i\phi$. The difference lies in the interpretation of belief. Voorbraak's notion of belief is stronger than ours. His view is that the agent cannot distinguish what he believes from what he knows (indeed, he believes that what he believes is the same as what he knows). Our notion of belief is weaker, in that we allow agents to be aware of the defeasibility of their beliefs.

Lamarre and Shoham [43] investigate the notion of knowledge as justified true belief using a framework that is very similar to ours. They start with an explicit preference ordering over possible worlds, and then define $B^\alpha\phi$ to read "given evidence α, ϕ holds in the most plausible α-worlds". Their formal account of $B^\alpha\phi$ is exactly $\alpha \rightarrow_i \phi$ in our notation. Unlike

[7] Shoham and Moses also examine two variants of this definition. These mainly deal with the cases where α is inconsistent with the agent's knowledge. For simplicity, we assume here that α is consistent with the agent's knowledge.

us, they examine a notion of knowledge as "belief stable under incorporation of correct facts", which is rather different then our notion of objective knowledge. Thus, while the technical construction is similar, the resulting framework is substantially different. Lamarre and Shoham take plausibility to be the only primitive, and use it to determine both knowledge and belief. We take both knowledge and plausibility to be primitive, and use them to define belief.

2.8 Axiomatizing the Language of Knowledge and Plausibility

Up to now, we have considered just the restricted language $\mathcal{L}^{KB}$. We now present sound and complete axiomatizations for the full language $\mathcal{L}^{KC}$. The technical details are much in the spirit of the axiomatizations presented in [14] for knowledge and probability. Our complete axiomatization for $\mathcal{M}$ consists of two "modules": a complete axiomatization for knowledge (i.e., S5) and a complete axiomatization for conditionals. In the general case, there are no axioms connecting knowledge and plausibility. For each of the conditions we consider, we provide an axiom that characterizes it. The axioms characterizing NORM, REF, RANK, and UNIF are taken from [45] and [7] (see also [17,24]), while the axioms for CONS and SDP (and also UNIF) correspond directly to the axioms suggested in [14] for their probabilistic counterparts. We also provide complete characterizations of the complexity of the validity problem for all the logics considered, based on complexity results for knowledge [34] and for conditionals [21].

The axiom system can be modularized into three components: propositional reasoning, reasoning about knowledge, and reasoning about conditionals. The component for propositional reasoning consists of K1 and RK1 (from Section 2.1); the component for reasoning about knowledge consists of K2–K5 and RK2 (from Section 2.1); the component for reasoning about conditionals consists of the standard axioms and rules for conditional logic C1–C4, RC1, and RC2 described in [17,24] following [7,45]:

C1. $\phi \rightarrow \phi$

C2. $((\phi \rightarrow \psi_1) \wedge (\phi \rightarrow \psi_2)) \Rightarrow (\phi \rightarrow (\psi_1 \wedge \psi_2))$

C3. $((\phi_1 \rightarrow \psi) \wedge (\phi_2 \rightarrow \psi)) \Rightarrow ((\phi_1 \vee \phi_2) \rightarrow \psi)$

C4. $((\phi_1 \rightarrow \phi_2) \wedge (\phi_1 \rightarrow \psi)) \Rightarrow ((\phi_1 \wedge \phi_2) \rightarrow \psi)$

R1. From ϕ and $\phi \Rightarrow \psi$ infer ψ

RC1. From $\phi \Leftrightarrow \phi'$ infer $(\phi \rightarrow \psi) \Rightarrow (\phi' \rightarrow \psi)$

RC2. From $\psi \Rightarrow \psi'$ infer $(\phi \rightarrow \psi) \Rightarrow (\phi \rightarrow \psi')$

Let AX consist of K1–K5, C1–C4, RK1, RK2, RC1, and RC2.

Theorem 11 *AX is a sound and complete axiomatization for $\mathcal{L}^{KC}$ with respect to $\mathcal{M}$.*

Proof. See Appendix A.2. $\square$

We now capture the conditions described above—CONS, NORM, REF, SDP, UNIF, and RANK—axiomatically.

RANK, NORM, REF, and UNIF correspond the axioms C5–C8, respectively, from [17,24]:

C5. $\phi \rightarrow \psi \wedge \neg(\phi \rightarrow \neg\xi) \Rightarrow \phi \wedge \xi \rightarrow \psi$
C6. $\neg(\mathit{true} \rightarrow \mathit{false})$.
C7. $N\phi \rightarrow \phi$
C8. $[(\phi \rightarrow \psi) \Rightarrow N(\phi \rightarrow \psi)] \wedge [\neg(\phi \rightarrow \psi) \Rightarrow N\neg(\phi \rightarrow \psi)]$

CONS and SDP correspond to the following axioms, respectively;

C9. $K_i\phi \Rightarrow N_i\phi$
C10. $(\phi \rightarrow_i \psi) \Rightarrow K_i(\phi \rightarrow_i \psi)$

It is interesting to note that the axioms for CONS and UNIF are derived from the axioms defined in [14] by replacing $w(\phi) = 1$ (the probability of ϕ is 1) by $N_i\phi$, which has a similar reading. We show that adding the appropriate axioms to AX gives a sound and complete axiomatization of the logic with respect to the class of structures satisfying the corresponding conditions.

Theorem 12 *Let $\mathcal{A}$ be a subset of $\{RANK, NORM, REF, UNIF, CONS, SDP\}$ and let A be the corresponding subset of $\{C5, C6, C7, C8, C9, C10\}$. Then $AX \cup A$ is a sound and complete axiomatization with respect to the structures in $\mathcal{M}$ satisfying $\mathcal{A}$.*

Proof. See Appendix A.2. $\square$

We now consider the complexity of the validity problem. Our results are based on a combination of results for complexity of epistemic logics [34] and conditional logics [21]. Again, the technical details are much in the spirit of those in [14].

We start with few results that will be useful in our discussion of complexity. As is often the case in modal logics, we can prove a "small model property" for our logic: if a formula is satisfiable at all, it is satisfiable in a small model. Let $Sub(\phi)$ be the set of subformulas in ϕ. It is easy to see that an upper bound on $|Sub(\phi)|$ is the number of symbols in ϕ.

Theorem 13 *Let $\mathcal{A}$ be a subset of $\{CONS, NORM, REF, SDP, UNIF, RANK\}$. The formula ϕ is satisfiable in a Kripke structure satisfying $\mathcal{A}$ if and only if it is satisfiable in a Kripke structure with at most $2^{|Sub(\phi)|}$ worlds.*

Proof. See Appendix A.2. $\square$

This shows that if ϕ is satisfiable, then it is satisfiable in a model with at most exponential number of worlds. Such a "small model" result is useful when we consider upper bound on the complexity of checking satisfiable. Roughly speaking, if there is a small model, then we can construct this model in time, say, exponential in the size of the formula. However, there is one problem with the result we have just proved. This "small" number of worlds does not necessarily mean that we can compactly describe the Kripke structure. Recall that $\mathrm{Pl}_{(w,i)}$ describes an ordering over subsets of $W_{(w,i)}$. Thus, in the worst case, we need to describe an ordering on $2^{|W_{(w,i)}|}$ sets of worlds. Thus, the representation of a structure might be exponential in the number of worlds. Fortunately, we can show that a satisfiable formula is satisfiable in a small model with a compact representation.

We start with a definition. We say that $M = (W, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{P}_1, \ldots, \mathcal{P}_n)$ is a *preferential (Kripke) structure* if for each $\mathcal{P}_i(w)$, there is a preference ordering $\prec_{(w,i)}$ on $W_{(w,i)}$ that induces $\mathrm{Pl}_{(w,i)}$ using the construction of Proposition 2. Recall that a preference ordering is a binary relation on the set of possible worlds. Thus, if W is finite, we can describe the relations $\mathcal{K}_i$ and the preference orderings $\leq_{(w,i)}$ using tables of size at most $|W|^2$. So the representation of such structures is polynomial in $|W|$. Is it possible to find a small preferential Kripke structure satisfying ϕ? Indeed we can. Using results of [21], we immediately get the following lemma:

Lemma 14 *Let $\mathcal{A}$ be a subset of {CONS, NORM, REF, SDP, UNIF, RANK}. If a formula ϕ is satisfiable in a Kripke structure satisfying $\mathcal{A}$ with N worlds, then ϕ is satisfiable in a preferential Kripke structure with at most $|\mathrm{Sub}(\phi)|N$ worlds.*

Combining this with Theorem 13, we conclude that if ϕ is satisfiable, then it is satisfiable in a structure of exponential size with an exponential description. It can be shown that this result is essentially optimal (see [34,21]). However, if there is only one agent and we assume CONS and either UNIF or SDP, then we can get polynomial-sized models.

Theorem 15 *Let $\mathcal{A}$ be a subset of {CONS, NORM, REF, SDP, UNIF, RANK} containing CONS and either SDP or UNIF. If ϕ talks about the knowledge and plausibility of only one agent, then ϕ is satisfiable in a Kripke structure satisfying $\mathcal{A}$ if and only if it is satisfiable in a preferential Kripke structure satisfying $\mathcal{A}$ with at most $|\mathrm{Sub}(\phi)|^3$ worlds.*

Proof. See Appendix A.2. $\square$

We now consider the complexity of decision procedure for the validity problem. The difficulty of deciding whether ϕ is valid is a function of the length of ϕ, written $|\phi|$.

Theorem 16 *Let $\mathcal{A}$ be a subset of {CONS, NORM, REF, SDP, UNIF, RANK}. If CONS $\in \mathcal{A}$, but it is not the case that UNIF or SDP is in $\mathcal{A}$, then the validity problem with respect* *to structures satisfying $\mathcal{A}$ is complete for exponential time. Otherwise, the validity problem is complete for polynomial space.*

Proof. See Appendix A.2. $\square$

If we restrict attention to the case of one agent and structures satisfying CONS and either UNIF or SDP, then we can do better.

Theorem 17 *Let $\mathcal{A}$ be a subset of $\{CONS, NORM, REF, SDP, UNIF, RANK\}$ containing CONS and either UNIF or SDP. For the case of one agent, the validity problem in models satisfying $\mathcal{A}$ is co-NP-complete.*

Proof. See Appendix A.2. $\square$

3 Adding Time

In the previous section, we developed a model of knowledge and beliefs. Having a good model of knowledge and belief is not enough in order to study how beliefs change. Indeed, if we are mainly interested in agents' beliefs, the additional structure of plausibility spaces does not play a significant role in a static setting. However, if we introduce an explicit notion of time, we expect the plausibility measure to (partially) determine how agents change their beliefs. As we shall see, this gives a reasonable notion of belief change.

In this section, we introduce time into the framework. We then examine how time, knowledge, and plausibility interact. In particular, we suggest a notion of *conditioning* that captures the intuition that plausibility changes in the minimal way that is required by changes to the agent's knowledge.

3.1 Knowledge and Plausibility in Multi-Agent Systems

A straightforward approach to adding time is by introducing another accessibility relation on worlds, which characterizes their temporal relationship (see, for example, [40]). We introduce more structure into the description by adopting the framework of Halpern and Fagin [33] for modeling multi-agent systems. This structure gives a natural definition of knowledge and an intuitive way to describe agents' interactions with their environment. We start by describing the framework of Halpern and Fagin, and then add plausibility.

The key assumption in this framework is that we can characterize the system by describing it in terms of a *state* that changes over time. This is a powerful and natural way to model systems. Formally, we assume that at each point in time, each agent is in some *local state*. Intuitively, this local state encodes the information that is available to the agent at that time. In addition, there is an *environment*, whose state encodes relevant aspects of the system that are not part of the agents' local states. For example, if we are modeling a robot that navigates in some office building, we might encode the robot's sensor input as part of the robot's local state. If the robot is uncertain about his position, we would encode this position in the environment state.

A *global state* is a tuple $(s_e, s_1, \ldots, s_n)$ consisting of the environment state s_e and the local state s_i of each agent i. A *run* of the system is a function from time (which, for ease of exposition, we assume ranges over the natural numbers) to global states. Thus, if r is a run, then $r(0), r(1), \ldots$ is a sequence of global states that, roughly speaking, is a complete description of what happens over time in one possible execution of the system. We take a *system* to consist of a set of runs. Intuitively, these runs describe all the possible sequences of events that could occur in a system.

Given a system $\mathcal{R}$, we refer to a pair (r, m) consisting of a run $r \in \mathcal{R}$ and a time m as a *point*. If $r(m) = (s_e, s_1, \ldots, s_n)$, we define $r_i(m) = s_i$; thus, $r_i(m)$ is agent i's local state at the point (r, m). Finally, to reason in a logical language about such a system, we need to assign truth values to primitive propositions. An *interpreted system* is a tuple $(\mathcal{R}, \pi)$ consisting of a system $\mathcal{R}$ together with a mapping π that associates a truth assignment with the primitive propositions at each state of the system.

An interpreted plausibility system can be viewed as a Kripke structure for knowledge. We say two points (r, m) and (r', m') are *indistinguishable* to agent i, and write $(r, m) \sim_i (r', m')$, if $r_i(m) = r'_i(m')$, i.e., if the agent has the same local state at both points. This is consistent with the intuition that an agent's local state encodes all the information available to the agent. Taking $\sim_i$ to define the $\mathcal{K}_i$ relation, we get a Kripke structure over points.[8]

This definition of knowledge has proved useful in many applications in distributed systems and AI (see [15] and the references therein). As argued above, we want to add the notion of plausibility so that we can model the agent's beliefs. It is straightforward to do so by adding a plausibility assessment for each agent at each point. Formally, an *interpreted plausibility system* is a tuple $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}_1, \ldots, \mathcal{P}_n)$, where $\mathcal{R}$ and π are as before, and the plausibility assignment $\mathcal{P}_i$ maps each point (r, m) to a plausibility space $\mathcal{P}_i(r, m) = (W_{(r,m,i)}, \mathrm{Pl}_{(r,m,i)})$.

In order to reason about the temporal aspects of the system, we add to the language temporal modalities in the standard fashion (see [26]). These include $\bigcirc\phi$ for "ϕ is true at the next time step" We call this language $\mathcal{L}^{KCT}$. Evaluation of temporal modalities at a point (r, m) is done by examining the future points on the run r: Given a point (r, m) in an interpreted system $\mathcal{I}$, we have that

[8] It is straightforward to extend these definitions to deal with continuous time. This is done, for example, in [6].

– $(\mathcal{I}, r, m) \models \bigcirc\phi$ if $(\mathcal{I}, r, m+1) \models \phi$.[9]

This framework is clearly a temporal extension of the logic of knowledge and plausibility described in the previous section.

3.2 Example: Circuit Diagnosis Revisited

We now show how the framework can be used to extend the example of Section 2.5 to incorporate time, allowing the agent to perform a sequence of tests.

We want to model the *process* of diagnosis. That is, we want to model the agent's beliefs about the circuit while it performs a sequence of tests, and how the observations at each step affects her beliefs. Thus, we want to model the agent and the circuit as part of a system. To do so, we need to describe the agent's local state and the state of the environment. The construction we used in Section 2.5 provides a natural division between the two: The agent's state is the sequence of input–output relations observed, while the environment's state describes the faulty components of the circuit and the values of all the lines. This corresponds to our intuitions, since the agent can observe only the input–output relations. Each run describes the results of a specific series of tests the agent performs and the results he observes. We make two additional assumptions: (1) the agent does not forget what tests were performed and their results, and (2) the faults are persistent and do not change over time. Formally, we define the agent's state $r_1(m)$ to be $\langle o_{(r,0)}, \ldots, o_{(r,m)} \rangle$, where $o_{(r,m)}$ describes the input–output relation observed at time m. We define the environment state $r_e(m) = (\mathit{fault}(r,m), \mathit{value}(r,m))$ to be the failure set at (r) and the values of all the lines. We capture the assumption that faults do not change by requiring that $\mathit{fault}(r,m) = \mathit{fault}(r,0)$. The system $\mathcal{R}_{diag}$ consists of all runs r satisfying these requirements in which $\mathit{value}(r,m)$ is consistent with $\mathit{fault}(r,m)$ and $o_{(r,m)}$ for all m.

Given the system $\mathcal{R}_{diag}$, we can define two interpreted plausibility systems corresponding to the two plausibility measures we considered in Section 2.5. In both systems, $W_{(r,m,1)} = \mathcal{K}_i(r,m)$. In $\mathcal{I}_{diag,1}$, we compare two points (r_1, m) and (r_2, m) by comparing the size of $\mathit{fault}(r_1,m)$ and $\mathit{fault}(r_2,m)$, while in $\mathcal{I}_{diag,2}$ we check whether one failure set is a subset of the other. At a point (r,m), the agent considers possible all the points where he performed the same tests up to time m and observed the same results. As before, the agent believes that the failure set is one of the minimal explanations of his observations. As the agent performs more tests, his knowledge increases and his beliefs might change.

We define $\mathrm{Bel}(\mathcal{I}, r, m)$ to be the set of failure sets (i.e., diagnoses) that the agent considers possible at (r,m). Belief change in $\mathcal{I}_{diag,1}$ is characterized by the following proposition.

Proposition 18 *If there is some $f \in \mathit{Bel}(\mathcal{I}_{diag,1}, r, m)$ that is consistent with the new ob-*

[9] It is easy to add other temporal modalities such as *until, eventually, since*, etc. These do not play a role in this work.

servation $o_{(r,m+1)}$, then Bel($\mathcal{I}_{diag,1}, r, m+1$) consists of all the failure sets in Bel($\mathcal{I}_{diag,1}, r, m$) that are consistent with $o_{(r,m+1)}$. If all $f \in B(\mathcal{I}_{diag,1}, r, m)$ are inconsistent with $o_{(r,m+1)}$, then $B(\mathcal{I}_{diag,1}, r, m+1)$ consists of all failure sets of cardinality j that are consistent with $o_{(r,m+1)}$, where j is the least cardinality for which there is at least one failure set consistent with $o_{(r,m+1)}$.

Proof. Straightforward; left to the reader. $\square$

Thus, in $\mathcal{I}_{diag,1}$, a new observation consistent with the current set of most likely explanations reduces this set (to those consistent with the new observation). On the other hand, a surprising observation (one inconsistent with the current set of most likely explanations) has a rather drastic effect. It easily follows from Proposition 18 that if $o_{(r,m+1)}$ is surprising, then $\mathrm{Bel}(\mathcal{I}_{diag,1}, r, m) \cap \mathrm{Bel}(\mathcal{I}_{diag,1}, r, m+1) = \emptyset$, so the agent discards all his current explanations in this case. Moreover, an easy induction on m shows that if $\mathrm{Bel}(\mathcal{I}_{diag,1}, r, m) \cap \mathrm{Bel}(\mathcal{I}_{diag,1}, r, m+1) = \emptyset$, then the cardinality of the failure sets in $\mathrm{Bel}(\mathcal{I}_{diag,1}, r, m+1)$ is greater than the cardinality of failure sets in $\mathrm{Bel}(\mathcal{I}_{diag,1}, r, m)$. Thus, in this case, the explanations in $\mathrm{Bel}(\mathcal{I}_{diag,1}, r, m+1)$ are more complicated than those in $B(\mathcal{I}_{diag,1}, r, m)$. Notice that if we can characterize the observation $o_{(r,m+1)}$ in our language—that is, if we have a formula ϕ such $(\mathcal{I}, r', m') \models \phi$ if and only if $o_{(r',m')} = o_{(r,m+1)}$—then we can also express the fact that agent i considers it surprising: This is true precisely if $(\mathcal{I}_{diag,1}, r, m) \models B_i \neg \bigcirc \phi$.

Belief change in $\mathcal{I}_{diag,2}$ is quite different, as the following proposition shows. Given a failure set f, we define $ext(f) = \{f' : f \subseteq f'\}$. Thus, $ext(f)$ consists of all the failure sets that extend f.

Proposition 19 *Bel($\mathcal{I}_{diag,2}, r, m+1$) consists of the minimal (according to $\subseteq$) failure sets in $\cup_{f \in Bel(\mathcal{I}_{diag,2}, r, m)} ext(f)$ that are consistent with $o_{(r,m+1)}$.*

Proof. Straightforward; left to the reader. $\square$

We see that, as with $\mathcal{I}_{diag,1}$, failure sets that are consistent with the new observation are retained. However, unlike $\mathcal{I}_{diag,1}$, failure sets that are discarded are replaced by more complicated failure sets even if some of the explanations considered most likely at (r, m) are consistent with the new observation. Moreover, while new failure sets in $\mathrm{Bel}(\mathcal{I}_{diag,1}, r, m+1)$ can be unrelated to failure sets in $\mathrm{Bel}(\mathcal{I}_{diag,1}, r, m)$, in $\mathcal{I}_{diag,2}$ the new failure sets must be extensions of some discarded failure sets. Thus, in $\mathcal{I}_{diag,1}$ the agent does not consider new failure sets as long as the observation is not surprising. On the other hand, in $\mathcal{I}_{diag,2}$ the agent has to examine new candidates after each test. The latter behavior is essentially that described by Reiter [54, Section 5].

3.3 Axiomatizing the Language of Knowledge, Plausibility and Time

We now present sound and complete axiomatization for the language $\mathcal{L}^{KCT}$. The technical details are much in the spirit of the results of Section 2.8, with two exceptions. First, we need to deal also with the temporal modality $\bigcirc$. Second, instead of dealing with worlds, we are dealing with systems that have some structure, i.e., the distinction between agents' local state and the environment's state. As we shall see, both issues can be dealt with in a straightforward manner.

The axiom system AX^T consists of the axioms and rule in the axiom system AX of Section 2.8 and the following axioms and rule the describe the properties of $\bigcirc$:.

T1. $\bigcirc\phi \wedge \bigcirc(\phi \Rightarrow \psi) \Rightarrow \bigcirc\psi$
T2. $\bigcirc\phi \equiv \neg\bigcirc\neg\phi$
RT1. From ϕ infer $\bigcirc\phi$.

Let $\mathcal{C}$ be the set of all plausibility interpreted systems.

Theorem 20 *The axiom system AX^T is a sound and complete axiomatization of $\mathcal{L}^{KCT}$ with respect to $\mathcal{C}$.*

Proof. See Appendix A.3. □

We can also prove a result analogous to Theorem 12 that describes a complete axiomatization for the classes of systems satisfying some of the assumptions we examined in Section 2.4.

Theorem 21 *Let $\mathcal{A}$ be a subset of $\{RANK, NORM, REF, UNIF, CONS, SDP\}$ and let A be the corresponding subset of $\{C5, C6, C7, C8, C9, C10\}$. Then $AX^T \cup A$ is a sound and complete axiomatization with respect to systems in $\mathcal{C}$ satisfying $\mathcal{A}$.*

Proof. See Appendix A.3. □

4 Prior Plausibilities

The formal framework of knowledge, plausibility and time described in the previous section raises a serious problem: While it is easy to see where the $\sim_i$ relations that define knowledge come from, the same cannot be said for the plausibility spaces $\mathcal{P}_i(r,m)$. We now present one possible answer to this question, inspired by probability theory.

Up to now, we have allowed the plausibility assessment at each point to be almost arbitrary. In particular, the plausibility space $\mathcal{P}_i(r, m)$ can be quite different from $\mathcal{P}_i(r, m + 1)$. Typically, we would expect there to be some relationship between these successive plausibility assessments. For example, it seems reasonable to expect that the new plausibility assessment should incorporate whatever was learned at $(r, m+1)$, but otherwise involve minimal changes from $\mathcal{P}_i(r, m)$.

One way of doing this in probability theory is by *conditioning.* If we start with a probability function Pr and observe E, where $\Pr(E) > 0$, then the conditional probability function $\Pr_E$ is defined so that $\Pr_E(A) = \Pr(A \cap E)/\Pr(E)$. Typically $\Pr_E(A)$ is denoted $\Pr(A|E)$. Notice that $\Pr_E$ incorporates the new information E by giving it probability 1. It also is a minimal change from Pr in the sense that if $A, B \subseteq E$, then $\Pr(A)/\Pr(B) = \Pr(A|E)/\Pr(B|E)$: the relative probability of events consistent with E is not changed by conditioning. [10]

Conditioning is a standard technique in probability theory, and can be justified in a number of ways, one of which is the notion of "minimal change" we have just described. Another justification is a "Dutch book" argument [9,53], which shows that if an agent uses some other method of updating probabilities, then it is possible to construct a betting game in which he will always lose. Probability measures are particular instances of plausibility measures. Can we generalize the notion of conditioning to plausibility measures?

It immediately follows from the definitions that the ordering of the likelihood of events induced by $\Pr_E$ is determined by the ordering induced by Pr:

$\Pr(A|E) \leq \Pr(B|E)$ if and only if $\Pr(A \cap E) \leq \Pr(B \cap E)$.

We want the analogous property for plausibility:

COND $\mathrm{Pl}(A|C) \leq \mathrm{Pl}(B|C)$ if and only if $\mathrm{Pl}(A \cap C) \leq \mathrm{Pl}(B \cap C)$.

This rule determines the order induced by posterior plausibilities. Since we are interested only in this aspect of plausibility, any method of conditioning that satisfies COND will do for our present purposes. (See [20] for an examination of other properties we might require of conditioning.) Notice that any two methods for conditioning are isomorphic in the following sense: Let $S_1 = (W_1, \mathrm{Pl}_1)$ and $S_2 = (W_2, \mathrm{Pl}_2)$ be two plausibility spaces. We say that S_1 and S_2 are *(order) isomorphic* if there is a bijection h from W_1 to W_2 such that, for $A, B \subseteq W_1$, we have $\mathrm{Pl}_1(A) \leq \mathrm{Pl}_1(B)$ if and only if $\mathrm{Pl}_2(h(A)) \leq \mathrm{Pl}_2(h(B))$. Any two definitions of conditioning that satisfy COND result in order-isomorphic plausibility spaces (see [20]).

This discussion suggests that we define $\mathrm{Pl}_{(r,m+1,i)}$ to be the result of conditioning $\mathrm{Pl}_{(r,m,i)}$

[10] There is another sense in which $\Pr_E$ represents the minimal change from Pr. If we measure the "distance" of a probability distribution $\Pr'$ from Pr in terms of the *cross-entropy* of $\Pr'$ relative to Pr, then it is well known that $\Pr_E$ is the distribution that minimizes the relative cross-entropy from Pr among all distributions $\Pr'$ such that $\Pr'(E) = 1$ [42]. Indeed, this holds true for other distance measures as well [11].

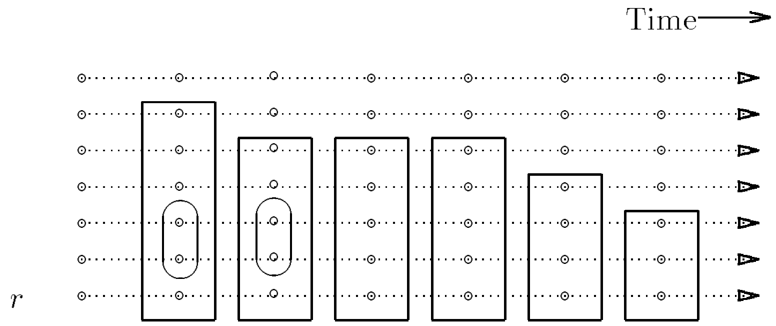

Fig. 2. Schematic description of how the agent's knowledge evolves in time in synchronous systems with perfect recall. The boxes represent the set of points in $\mathcal{K}_i(r,m)$. Since the system is synchronous, at each time point, the agent consider possible points at the same time. Since the agent has perfect recall, as time progresses, the agent considers smaller and smaller sets of runs possible. The ovals represent two disjoint events that correspond to the same set of runs.

on the new knowledge gained by agent i at $(r, m+1)$. This, however, leads to the following technical problem. If the agent gains new knowledge at $(r, m+1)$, then $r_i(m) \neq r_i(m+1)$. This implies that the sets of points the agent considers possible are disjoint, i.e., $\mathcal{K}_i(r,m) \cap \mathcal{K}_i(r,m+1) = \emptyset$. But then CONS implies that $\mathrm{Pl}_{(r,m,i)}$ and $\mathrm{Pl}_{(r,m+1,i)}$ are defined over disjoint spaces, so we cannot apply COND.

We circumvent this difficulty by working at the level of runs. The approach we propose resembles the Bayesian approach to probabilities. Bayesians assume that agents start with priors on all possible events. If we were thinking probabilistically, we could imagine the agents in a multi-agent system starting with priors on the runs in the system. Since a run describes a complete history over time, this means that the agents are putting a prior probability on the sequences of events that could happen. We would then expect the agent to modify his prior by conditioning on whatever information he has learned. This is essentially the approach taken in [35] to defining how the agents' probability distribution changes in a multi-agent system. We can do the analogous thing with plausibility.

We start by making the simplifying assumption that we are dealing with *synchronous* systems where agents have *perfect recall* [36]. Intuitively, this means that the agents know what the time is and do not forget the observations they have made. Formally, a system is synchronous if for any i, $(r,m) \sim_i (r',m')$ only if $m = m'$. Notice that by restricting to synchronous systems, if we further assume that the plausibility measure $\mathcal{P}_i(r,m)$ satisfies CONS, we never have to compare the plausibilities of two different points on the same run. In synchronous systems, agent i has perfect recall if $(r', m+1) \sim_i (r, m+1)$ implies $(r',m) \sim_i (r,m)$. Thus, agent i considers run r possible at the point $(r, m+1)$ only if he also considers it possible at (r,m). This means that any runs considered impossible at (r,m) are also considered impossible at $(r, m+1)$; an agent does not forget what he knew.

Just as with probability, we assume that an agent has a prior plausibility measure on runs, that describes his prior assessment on the possible executions of the system. As the agent gains knowledge, he updates his prior by conditioning. More precisely, at each point (r,m), the agent conditions his previous assessment on the set of runs considered possible at (r,m).

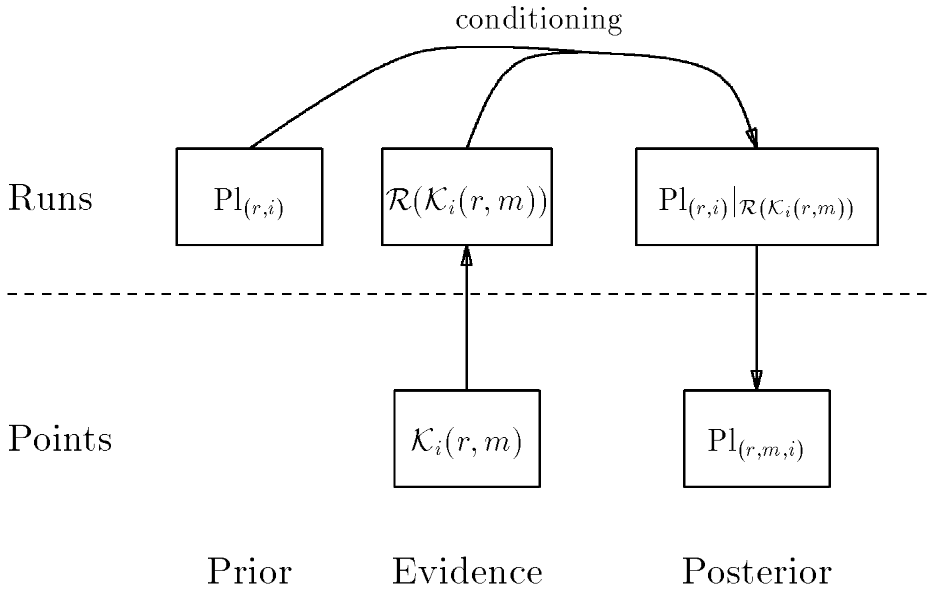

Fig. 3. Schematic description of the entities involved in the definition of priors. Note some are defined over runs and some over points.

This is process is shown in Figure 2. This results in an updated assessment (posterior) of the plausibility of runs. This posterior induces, via a projection from runs to points, a plausibility measure on points. We can think of agent i's posterior at time m as simply his prior conditioned on his knowledge at time m.

To make this precise, let $S = (W, \mathrm{Pl})$ be a plausibility space. Define the *projection* of S on E as $S|_E = (W|_E, \mathrm{Pl}|_E)$, where $W|_E = W \cap E$ and $\mathrm{Pl}|_E$ is the restriction of Pl to $W|_E$. Projection is similar to conditioning: for any definition of conditioning that satisfies COND if $A, B \subseteq E$, then $\mathrm{Pl}(A|E) \leq \mathrm{Pl}(B|E)$ if and only if $\mathrm{Pl}|_E(A) \leq \mathrm{Pl}|_E(B)$. Indeed, $S|_E$ is essentially isomorphic to any conditional plausibility measure that results from conditioning on E.[11]

We can now define what it means for a plausibility measure on points to be generated by a prior. Suppose that agent i's prior plausibility at run r is $\mathcal{P}_{(r,i)} = (\mathcal{R}_{(r,i)}, \mathrm{Pl}_{(r,i)})$, where

[11] To make this precise, we need a notion that is slightly more general than isomorphism. Let $P = (W, \mathrm{Pr})$ be a probability space. A set A is called a *support* of P if $\mathrm{Pr}(\overline{A}) = 0$. We can define a similar notion for plausibility spaces. Let $S = (W, \mathrm{Pl})$ be a plausibility space. We say that $A \subseteq W$ is a *support* of S, if for all $B \subseteq W$, $\mathrm{Pl}(B) = \mathrm{Pl}(B \cap A)$. Thus, only $B \cap A$ is relevant for determining the plausibility of B. This certainly implies that $\mathrm{Pl}(\overline{A}) = \bot$, since we must have $\mathrm{Pl}(\overline{A}) = \mathrm{Pl}(A \cap \overline{A}) = \mathrm{Pl}(\emptyset)$, but the converse does not hold in general. In probability spaces, $\mathrm{Pr}(\overline{A}) = 0$ implies that $\mathrm{Pr}(B) = \mathrm{Pr}(B \cap A)$ for all B, but the analogous condition does not hold for arbitrary plausibility spaces. We say that two plausibility spaces S_1 and S_2 are *essentially (order) isomorphic* if there are supports C_1 and C_2 of S_1 and S_2, respectively, such that $S_1|_{C_1}$ is isomorphic to $S_2|_{C_2}$. It is easy to see that, as expected, essential isomorphism defines an equivalence relation among plausibility spaces. Finally, it is easy to see that if $S = (W, \mathrm{Pl})$, then $(W, \mathrm{Pl}(\cdot|E))$ is essentially isomorphic to $S|_E$ when we use any conditioning method that satisfies COND.

$\mathcal{R}_{(r,i)} \subseteq \mathcal{R}$. Our intuition is that the agent conditions the prior by his knowledge at time (r, m). In our framework, the agent's knowledge at time m is the set of point $\mathcal{K}_i(r, m)$. We need to convert this set of points to an event in terms of runs. If A is a set of points, we define $\mathcal{R}(A) = \{r : \exists m((r, m) \in A)\}$ to be the set of runs on which the points in A lie. Using this notation, the set of runs agent i considers possible at (r, m) is simply $\mathcal{R}(\mathcal{K}_i(r, m))$. Thus, after conditioning on this set of runs, we get agent i's *posterior* at (r, m), which is simply the projection of the prior on the observation: $\mathrm{Pl}_{(r,i)}|_{\mathcal{R}(\mathcal{K}_i(r,m))}$. We now use this plausibility measure, which is a measure on a set of runs, to define $\mathcal{P}_i(r, m)$, which is a measure on a set of points. We do so in the most straightforward way: we project each run to a point that lies on it. Formally, we say that $\mathcal{P}_i(r, m)$ is the *time m projection* of $\mathcal{P}_{(r,i)}|_{\mathcal{R}(\mathcal{K}_i(r,m))}$ if $\mathcal{P}_i(r, m) = (W_{(r,m,i)}, \mathrm{Pl}_{(r,m,i)})$, where $W_{(r,m,i)} = \{(r', m) \in \mathcal{K}_i(r, m) : r' \in \mathcal{R}_{(r,i)}\}$ and for all $A \subseteq W_{(r,m,i)}$, we have that $\mathrm{Pl}_{(r,m,i)}(A) = \mathrm{Pl}_{(r,i)}|_{\mathcal{R}(\mathcal{K}_i(r,m))}(\mathcal{R}(A))$. $\mathrm{Pl}_{(r,m,i)}$ is the agent's plausibility measure at (r, m). This process is described in Figure 3. The main complications are due to the transition back and forth between entities defined over runs and ones defined over points.

We remark that if the system satisfies perfect recall as well as synchrony, our original intuition that $\mathcal{P}_i(r, m + 1)$ should be the result of conditioning $\mathcal{P}_i(r, m)$ on the knowledge that agent i acquires at $(r, m + 1)$ can be captured more directly. We can in fact construct $\mathcal{P}_i(r, m + 1)$ from $\mathcal{P}_i(r, m)$ by what can be viewed as conditioning on the agent's new information: We take $\mathcal{P}_i(r, m)$ and project it one time step forward by replacing each point (r', m) by $(r', m+1)$. We then condition on $\mathcal{K}_i(r, m+1)$ (i.e., the agent's knowledge at $(r, m+1)$) to get $\mathcal{P}_i(r, m, i+1)$.

Proposition 22 *Let $\mathcal{I}$ be a synchronous system satisfying perfect recall such that $\mathrm{Pl}_{(r,m,i)}$ is the time m projection of a prior $\mathrm{Pl}_{(r,i)}$ on runs for all runs r, times m, and agents i. Let $prev(A) = \{(r, m) : (r, m + 1) \in A\}$. Then $\mathrm{Pl}_{(r,m+1)}(A) \leq \mathrm{Pl}_{(r,m+1)}(B)$ if and only if $\mathrm{Pl}_{(r,m)}(prev(A)) \leq \mathrm{Pl}_{(r,m)}(prev(B))$, for all runs r, times m, and sets $A, B \in W_{(r,m+1)}$.*

Proof. Straightforward; left to the reader. □

We say that $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ satisfies PRIOR if $\mathcal{I}$ is synchronous and for each run r and agent i there is a prior plausibility $\mathcal{P}_{(r,i)}$ such that for all m, $\mathcal{P}_i(r, m)$ is the time m projection of $\mathcal{P}_{(r,i)}$.

Example 23 *It is easy to verify that the two systems we consider in Section 3.2 satisfy PRIOR. In both systems, the prior $\mathcal{P}_{(r,i)}$ is independent of the run r, and is determined by the failure set in each run.*

By using prior plausibility measures, we have reduced the question of where the plausibility measure at each point comes from to the simpler question of where the prior comes from. While this question is far from trivial, it is analogous to a question that needs to be addressed by anyone using a Bayesian approach. Just as with probability theory, in many applications there is a natural prior (or class of priors) that we can use.

By conditioning on plausibility rather than probability, we can deal with a standard problem in the Bayesian approach, that of conditioning on an event of measure 0: Notice that whenever a prior assigns an event a probability measure of 0 it is not possible to condition on that event. The standard solution in the Bayesian school is to give every event of interest, no matter how unlikely, a small positive probability.[12] We may well discover that a formula ϕ that we believed to be true, i.e., one that was true in all the most plausible worlds, is in fact false. Under the probabilistic interpretation of plausibility, this means that we are essentially conditioning on an event $(\neg\phi)$ of measure 0. The plausibility approach has no problem with this: the conditioning process described above still makes perfect sense.

4.1 Conditioning as Minimal Change of Belief

In this section we examine the properties of conditioning as an approach to minimal change of beliefs and relate our approach to others in the literature.

Recall that QUAL guarantees that belief is closed under logical implication and conjunction (Theorem 4). In a synchronous system where the prior satisfies QUAL, it is not hard to see that conditioning preserves QUAL. Thus, we get the following result.

Proposition 24 *Let $\mathcal{I}$ be a synchronous system satisfying perfect recall and PRIOR. If the prior $\mathrm{Pl}_{(r,i)}$ satisfies A2 for all runs r and agents i, then axiom K2 is valid in $\mathcal{I}$ for B_i.*

Proof. Straightforward; left to the reader. $\square$

This result shows that condition A2 is sufficient to get beliefs that satisfy K2. Is it also necessary? In general, the answer is no. However, A2 is the most natural condition that ensures that K2 is satisfied. To see this, note that if K2 is valid in $\mathcal{I}$ then A2 holds for all pairwise disjoint subsets A_1, A_2 and A_3 of points in $\mathcal{I}$ definable in the language such that $\mathcal{R}(\mathcal{K}_i(r,m)) = A_1 \cup A_2 \cup A_3$ for some run r, agent i, and time m. Thus, if we assume that the language is rich enough so that all subsets of $\mathcal{I}$ are definable (in that, for each subset A and agent i, there is a formula ϕ and point (r,m) such that $A = [\![\phi]\!]_{(r,m,i)}$), then K2 forces A2.

In view of this discussion, we focus in this section on synchronous systems with a qualitative prior.

Next, we examine how changes in beliefs are determined by the prior. Using Proposition 22, we now show that we can characterize, within our language, how the agent's beliefs change via conditioning, provided that we can describe in the language what knowledge the agent acquired. We say that a formula ϕ *characterizes agent i's knowledge at* $(r,m+1)$ with respect

[12] Of course, this requires that there be only countably many events of interest.

to his knowledge at (r, m) if, for all $(r', m) \in \mathcal{K}_i(r, m)$, we have $(r', m+1) \models \phi$ if and only if $(r', m+1) \in \mathcal{K}_i(r, m+1)$. That is, among the points that succeed points that are considered possible at time m, exactly these satisfying ϕ are considered possible at time $m+1$. Of course, it is not always possible to characterize the agent's new knowledge by a formula in our language. However, in many applications we can limit our attention to systems where it is possible. (This is the case, for example, in our treatment of revision and update in [17,23].) In such systems, we can characterize within the agent's belief change process in the language.

Proposition 25 *Let $\mathcal{I}$ be a synchronous system satisfying perfect recall and PRIOR. If ϕ characterizes agent i's knowledge at $(r, m+1)$ with respect to his knowledge at (r, m), then $(\mathcal{I}, r, m+1) \models \psi \rightarrow_i \xi$ if and only if $(\mathcal{I}, r, m) \models \bigcirc(\phi \wedge \psi) \rightarrow_i \bigcirc\xi$.*

Proof. See Appendix A.4. $\Box$

Corollary 26 *Let $\mathcal{I}$ be a synchronous system satisfying perfect recall and PRIOR. If ϕ characterizes agent i's knowledge at $(r, m+1)$ with respect to his knowledge at (r, m), then $(\mathcal{I}, r, m+1) \models B_i\psi$ if and only if $(\mathcal{I}, r, m) \models K_i(\bigcirc\phi \Rightarrow (\bigcirc\phi \rightarrow_i \bigcirc\psi))$. Moreover, if $\mathcal{I}$ also satisfies SDP, then $(\mathcal{I}, r, m+1) \models B_i\psi$ if and only if $(\mathcal{I}, r, m) \models \bigcirc\phi \rightarrow_i \bigcirc\psi$.*

We now use this result to relate our approach to other approaches for modeling conditionals in the literature. Boutilier [3], Goldszmidt and Pearl [31], and Lamarre and Shoham [43] give conditional statements similar semantics (using a preference ordering), but $\phi \rightarrow \psi$ is read "after learning ϕ, ψ is believed". Two crucial assumptions are made in these papers. The first is that the agent considers only one plausibility assessment, which in our terminology amounts to SDP. The second is that propositions are static, i.e., their truth value does not change along a run.[13] Formally, a system is *static* if $\pi(r(m)) = \pi(r(0))$ for all runs r and times m. This implies that for any propositional formula ϕ, we have that $\phi \equiv \bigcirc\phi$. These two assumptions lead to a characterization of belief change.

Corollary 27 *Let $\mathcal{I}$ be a synchronous static system satisfying PRIOR, SDP, and perfect recall, and let ϕ and ψ be propositional formulas. If ϕ characterizes agent i's knowledge at $(r, m+1)$ with respect to his knowledge at (r, m), then $(\mathcal{I}, r, m+1) \models B_i\psi$ if and only if $(\mathcal{I}, r, m) \models \phi \rightarrow_i \psi$.*

While this result shows that, in certain contexts, there is a connection between a statement such as "typically ϕ's are ψ's" (which is how we have between interpreting $\phi \rightarrow_i \psi$) and "after learning ϕ, ψ is believed" (which is how it is interpreted in [3,31,43]), the two readings are in general quite different. For one thing, notice that Corollary 27 assumes that ϕ and ψ are propositional formulas. This is a necessary assumption. If ϕ and ψ contain modal formulas, then $\phi \rightarrow \psi$ does not necessarily imply that the agent believes ψ at the next time step. For example, if $(\mathcal{I}, r, m) \models B_i\psi$, then for any formula ϕ, we have $(\mathcal{I}, r, m) \models \phi \rightarrow_i B_i\psi$,

[13] This assumption is only implicit, since none of these papers have an explicit representation of time. Nevertheless, it is clear that this assumption is being made.

regardless of whether $B_i\psi$ is believed at $(r, m+1)$. In [18], we examine conditionals of the form $\phi > \psi$ intended to capture the second interpretation "ψ is believed after learning ϕ". The semantics for these conditionals involves examining future time points, just as our intuitive reading dictates. As we have just seen, $>$ and $\rightarrow$ are quite different when we consider modal formulas in the scope of these conditionals.

This discussion shows one of the benefits of representing time explicitly. In our framework we can distinguish between agents' plausibility assessment and their belief dynamics. Of course, we would like agents to be persistent in their assessment, which is exactly what conditioning captures. In the presence of several assumptions, we get a close connection between agents' conditional beliefs and how their beliefs change. This allows us to identify some of the assumptions implicitly made in previous approaches. For example, all of the approaches we mentioned above would not apply when we consider a changing environment, since they cannot reason about how the environment changes between one time point and the next.

Finally, we examine the work of Battigalli and Bonanno [2]. They consider a logic of knowledge, belief, and time, and attempt to capture properties of "minimal change" of beliefs. Their language is slightly different from ours. Instead of introducing a temporal modality, they define a different belief and knowledge modality for each time step: $B^t\phi$ reads "the agent believes ϕ at time t". Battigalli and Bonanno also assume that propositions are static and do not change in time. Thus, the only changes are in terms of the agent's knowledge and belief. Battigalli and Bonanno propose an axiom system similar to the axioms of Kraus and Lehmann (that is, they use K5 for knowledge is K5, KD45 for belief, and take axioms KB1 and KB2 of Section 2.7 to characterize the connection between knowledge and belief) that also includes two additional axioms that can be written in our language as

BT1. $B_i\bigcirc B_i\phi \Rightarrow B_i\phi$
BT2. $B_i\phi \Rightarrow B_i\bigcirc B_i\phi$

Battigalli and Bonanno claim that these axioms capture the principle that the agent does not change her mind unless new knowledge forces her to do so. Intuitively, this principle also applies to conditioning, and thus it is instructive to understand when these axioms are satisfied in our framework.

It turns out that RANK combined with a minimal assumption implies both BT1 and BT2. We say that a system has *finite branching* if it allows only finitely many "branches" at each local state of an agent (that is there are only finitely many observations that an agent can make at each point).

Lemma 28 *Let $\mathcal{I}$ be a synchronous static system satisfying PRIOR, RANK, SDP, and perfect recall that has finite branching. Then $(\mathcal{I}, r, m) \models B_i\phi \Leftrightarrow B_i\bigcirc B_i\phi$ for all propositional formulas ϕ.*

Proof. See Appendix A.4. $\square$

Are these conditions necessary to characterize BT1 and BT2? The answer is no. First, the proof of Lemma 28 applies to systems with infinite branching, if the agents' prior satisfies an infinitary version of A2. As shown in [25], this infinitary version is satisfied by κ-rankings and preference orderings that are *well founded* (that is, they have no infinite descending sequences $\cdots \prec w_3 \prec w_2 \prec w_1$). Thus, any system with static propositions whose prior is induced by a well-founded preference order satisfies BT1 and BT2. Note that BT1 and BT2 do not characterize RANK, since they put restrictions only on certain events (ones definable by a conjunction of a formula and the agent's new knowledge at some time point). However, RANK is the most natural restriction that implies these axioms.

Thus, we see that Battigalli and Bonanno essentially require systems with minimal change to satisfy conditioning with a prior that is a ranking. As we shall see in the next section, similar requirements are made by the AGM formulation of belief revision [1].

4.2 Properties of Prior Plausibilities

If we take the plausibilities in a system to be generated by a prior, then many of the conditions we are interested in, such as QUAL and REF, can be viewed as being as being induced by the analogous property on the prior. We have considered these properties only in the context of Kripke structures for knowledge and probability, so to make sense of the prior having the "analogous property", we have to be able to view the set of runs as a Kripke structure for knowledge and probability. Let $\mathcal{I}$ be a synchronous system satisfying perfect recall and PRIOR. Define $M_{\mathcal{I}}^r = (\mathcal{R}, \pi^r, \mathcal{K}_1^r, \ldots, \mathcal{K}_n^r, \mathcal{P}_i^r, \ldots, \mathcal{P}_n^r)$, where π^r is an arbitrary truth assignment, $\mathcal{K}_i^r$ is the full relation, i.e., $\mathcal{R} \times \mathcal{R}$, and $\mathcal{P}_i^r(r) = \mathcal{P}_{(r,i)}$, the prior of agent i at run r.

Proposition 29 *Let $\mathcal{I}$ be a synchronous system satisfying perfect recall and PRIOR. If $M_{\mathcal{I}}^r$ satisfies QUAL, REF, SDP, UNIF or RANK, then so does $\mathcal{I}$.*

Proof. Straightforward; left to the reader. $\square$

Thus, by constructing priors that satisfy various properties, we can ensure that the resulting system also satisfies them. In particular, Proposition 29 implies that if $\mathcal{P}_{(r,i)}$ is independent of r, so that agent i's prior is independent of the run he is in, then $\mathcal{I}$ satisfies SDP. A somewhat weaker assumption—that the set of runs can be partitioned into disjoint subsets $\mathcal{R}_1, \ldots, \mathcal{R}_k$ such that for $r, r' \in \mathcal{R}_j$, we have $\mathcal{P}_{(r,i)} = \mathcal{P}_{(r',i)} = (\mathcal{R}_j, \mathrm{Pl}_j)$—ensures that $\mathcal{I}$ satisfies UNIF. Intuitively, the sets $\mathcal{R}_j$ correspond to different settings of parameters. Once we set the parameters, then we fix the plausibility measure (and it is the same at all runs that have the same setting of the parameters).

We conclude this section by examining whether assuming conditioning limits the expressiveness of our belief change operation. A well-known result of Diaconis and Zabell [11] that shows that, in a precise sense, any form of *coherent* probabilistic belief change can be described by conditioning. In particular, they show that, given two probability distributions Pr and Pr$'$ on a finite space W that are coherent in the sense that $\Pr(A)=0$ implies that $\Pr'(A)=0$, there is a space W^* of the form $W\times X$, a subset E of W^*, and a distribution Pr$''$ on W^* such that, for all $A\subseteq W$, we have $\Pr''(A\times X)=\Pr(A)$ (so that Pr$''$ can be viewed as an extension of Pr) and $\Pr'(A)=\Pr''(A\times X|E)$.

We can prove a result in a somewhat similar spirit in our framework. The first step is to define a plausibilistic analogue of coherence in systems.

Let $\mathcal{I}$ be a synchronous system. We say that $\mathcal{I}$ is *coherent* if the following condition is satisfied for all r and m: Suppose $R\subseteq\mathcal{R}$, $A^m\subseteq W_{(r,m,i)}$, $\mathcal{R}(A^m)=R\cap\mathcal{R}(W_{(r,m,i)})$, $A^{m+1}\subseteq W_{(r,m+1,i)}$, and $\mathcal{R}(A^{m+1})=R\cap\mathcal{R}(W_{(r,m+1,i)})$. If $\mathrm{Pl}_{(r,m,i)}(A^m)=\perp$, then $\mathrm{Pl}_{(r,m+1,i)}(A^{m+1})=\perp$. Despite the different formulation, this condition is analogous to the probabilistic coherence of Diaconis and Zabell. Roughly speaking, if a set of runs has plausibility $\perp$ (which is analogous to probability 0 for Diaconis and Zabell) at time m, then it is required to have plausibility $\perp$ at time $m+1$. More precisely, coherence of a system ensures that sets of runs that were considered implausible at (r,m), either by being outside $W_{(r,m,i)}$ or by being given plausibility $\perp_{(r,m,i)}$, are also considered implausible at $(r,m+1)$. Note, this condition does not put any constraints on how the runs that are considered possible are ordered. It is easy to verify that the following axiom is valid in coherent systems:

COH. $N_i\bigcirc\phi\Rightarrow\bigcirc N_i\phi$

Proposition 30 *If $\mathcal{I}$ is a synchronous and coherent system, then COH is valid in $\mathcal{I}$.*

Proof. Straightforward; left to the reader. $\square$

There is a sense in which the converse to Proposition 30 holds as well: Given a synchronous system that is not coherent, we can define a truth assignment π in this system for which COH does not hold. [14]

It is easy to see that coherence is a necessary condition for satisfying PRIOR.

Proposition 31 *If $\mathcal{I}$ is a synchronous system satisfying perfect recall and PRIOR, then $\mathcal{I}$ is coherent.*

[14] We remark that COH is analogous to the axiom $K_i\bigcirc\phi\Rightarrow\bigcirc K_i\phi$ that characterizes perfect recall in synchronous systems [15]. Roughly speaking, this is because coherence ensures that the agent does not forget what she ruled out as implausible.

Proof. Straightforward; left to the reader. $\square$

Thus, PRIOR forces systems to be coherent, and hence to satisfy COH. It also forces systems to satisfy CONS, and hence C5. As we shall see, it also forces some other semantic properties. Nevertheless, we can show that for coherent systems that satisfy CONS, PRIOR does not force any additional properties, by proving an analogue to the Diaconis and Zabell result in our framework.

We say that a formula $\phi \in \mathcal{L}^{KCT}$ is *temporally linear* if temporal modalities in ϕ do not appear in the scope of the $\mathcal{K}_i$ or $\rightarrow_i$ modalities. Thus, for example, a formula such as $(\phi \rightarrow_i \psi) \Rightarrow \bigcirc B_i\psi$ is temporally linear, while $K_i(\bigcirc\phi \rightarrow_i \bigcirc\psi) \Rightarrow \bigcirc B_i\psi$ is not. Temporal linearity ensures that all the temporal connectives in ϕ are evaluated with respect to a single run. The following result says that, at least for temporally linear formulas, we can view belief change in a coherent system $\mathcal{I}$ as coming from conditioning on a prior, in the sense that we can embed $\mathcal{I}$ into a larger system where this is the case.

Theorem 32 *Let $\mathcal{A}$ be a subset of $\{QUAL, NORM, REF, RANK\}$ and let $\mathcal{I}$ be a coherent synchronous system satisfying perfect recall, CONS, and $\mathcal{A}$. Then there is a synchronous system $\mathcal{I}'$ satisfying perfect recall, PRIOR, and $\mathcal{A}$, and a mapping $f : \mathcal{R} \mapsto \mathcal{R}'$ such that for all temporally linear formulas $\phi \in \mathcal{L}^{KCT}$, we have $(\mathcal{I}, r, m) \models \phi$ if and only if $(\mathcal{I}', f(r), m) \models \phi$.*[15]

Proof. See Appendix A.5. $\square$

Notice that formulas that just compare an agent's beliefs (or knowledge) at successive time points are temporally linear. All the AGM postulates and the KM postulates (when translated to our language) are of this form. Not surprisingly, as we show in [17,23], these postulates can be captured by systems with the appropriate prior plausibility.

Can we extend Theorem 32 to the full language? We conjecture that Theorem 32 actually holds for all $\phi \in \mathcal{L}^{KCT}$, not just temporally linear formulas. This conjecture implies that a formula is valid with respect to synchronous systems satisfying perfect recall, CONS, and PRIOR if and only if it is valid with respect to synchronous coherent systems satisfying CONS and perfect recall. That is, except for COH and C9, we do not get any new properties by assuming PRIOR and CONS.

Note that the construction described by Theorem 32 does not necessarily preserve SDP or UNIF in the transformation from $\mathcal{I}$ to $\mathcal{I}'$. This is due to the fact that in the presence of SDP or UNIF, PRIOR forces new semantic properties. Recall that UNIF implies that there is a

[15] We note that this result is, in a sense, stronger than Diaconis and Zabell's. They examine only the probability of events, which are essentially propositional formulas (i.e., formulas without modal operators).

partition of possible points such that two points (r, m) and (r', m') are in the same cell if and only if $\mathcal{P}_i(r, m) = \mathcal{P}_i(r', m')$. Let PERSIST be the requirement that this partition changes minimally in time. More precisely, we say that a system satisfies PERSIST if for all runs $r, r' \in \mathcal{R}$ and m such that $(r, m+1) \sim_i (r', m+1)$, we have that $\mathcal{P}_i(r, m+1) = \mathcal{P}_i(r', m+1)$ if and only if $\mathcal{P}_i(r, m) = \mathcal{P}_i(r', m)$. Intuitively, PERSIST (in the presence of synchrony, perfect recall, and CONS) implies that the partition of points at time $m+1$ is determined by the partition of corresponding points at time m and the knowledge relation at time $m+1$.

Proposition 33 *If $\mathcal{I}$ is a synchronous system that satisfies perfect recall and either PRIOR and UNIF, or SDP, then $\mathcal{I}$ satisfies PERSIST.*

Proof. Straightforward; left to the reader. □

It is not clear to us at this stage whether PERSIST forces new properties in our language. However, if we assume that PERSIST holds, we can get a result analogous to Theorem 32.

Theorem 34 *Let $\mathcal{A}$ be a subset of $\{QUAL, NORM, REF, SDP, UNIF, RANK\}$ and let $\mathcal{I}$ be a coherent synchronous system satisfying perfect recall, CONS, PERSIST, and $\mathcal{A}$. Then there is a synchronous system $\mathcal{I}'$ satisfying perfect recall, PRIOR, and $\mathcal{A}$, and a mapping $f : \mathcal{R} \mapsto \mathcal{R}'$ such that for all temporally linear formulas $\phi \in \mathcal{L}^{KCT}$, $(\mathcal{I}, r, m) \models \phi$ if and only if $(\mathcal{I}', f(r), m) \models \phi$.*

Proof. See Appendix A.5. □

Thus, the question of whether PRIOR forces new properties in the presence of UNIF reduces to the question of whether PERSIST forces new properties. Finally, since SDP implies PERSIST, PRIOR does not force new properties in the presence of SDP.

Our discussion of conditioning and priors up to now assumed synchrony and perfect recall. Can we make sense of conditioning when we relax these assumptions? Note that the definition of PRIOR does not rely on perfect recall. PRIOR is well defined even in systems where agents can forget. However, in such systems, the intuitions that motivated the use of PRIOR are no longer valid. In particular, PRIOR does not imply coherence and the analogue to Proposition 22 does not hold: we no longer can construct $\mathcal{P}_i(r, m+1)$ from $\mathcal{P}_i(r, m)$ since runs that are considered impossible at time m might be considered possible at time $m+1$.[16] Dropping the assumption of synchrony also leads to problems, even in the presence of perfect recall. In an asynchronous setting, an agent might consider several points on the same run possible. The question then arises as to how (or whether) we should distribute the plausibility of a run over these points. Two approaches are considered in a probabilistic setting in [51], in

[16] We could, of course, redefine PRIOR so as to guarantee that Proposition 22 holds, but this leads to other complications.

the context of analyzing games with imperfect recall. It would be of interest to see to what extent these approaches can be carried over to the plausibilistic setting.

5 Conclusion

We have proposed a framework for belief dynamics that combines knowledge, time, and plausibility (and hence beliefs), and investigated a number of properties of the framework, such as complete axiomatizations for various sublanguages and various properties of the relationships between the modal operators. Of course, the obvious question is why we should consider this framework at all.

There are two features that distinguish our approach from others. The first is that we use plausibility to model uncertainty, rather than other approaches that have been mentioned in the literature, such as preference orderings on worlds or ϵ-semantics. The second is that we include knowledge and time, as well as belief, explicitly in the framework.

We could have easily modified the framework to use other ways of modeling uncertainty. Indeed, in a preliminary version of this paper [19], we used preference orderings. We have chosen to use plausibility measures for several reasons. First, plausibility measures generalize all approaches to representing uncertainty that we are aware of. The use of plausibility makes it easier to compare our approach, not only to preference-based approaches (e.g., [3]), but also to approaches based on κ-rankings (e.g., [31]), probably measures (e.g., [35]), or any other measure of uncertainty. More importantly, it makes it easier for us to incorporate intuitions from other approaches. We have already seen one example of this phenomenon in the present paper: we defined a plausibilistic analogue of conditioning, and used it to model minimal change. As we show in [23], we can represent the standard approaches to minimal change—belief revision and belief update—in terms of conditioning. Moreover, the semantic characterization of conditioning should allow us to apply it more easily to deal with complications that arise when the language lets us reason about multiple agents, actions, and beliefs about beliefs. Another example of adopting probabilistic intuitions is given in [17,20,22], where plausibilistic analogues of independence and Markov chains are described and used to define a novel approach to belief change. We believe that these notions will have applications elsewhere as well. Finally, plausibility measures have the advantage of greater expressive power than other approaches. For example, work on defaults has mainly focused on properties of structures with a finite number of worlds. In our framework, however, even a simple system with two global states might have an uncountable number of runs. As shown in [25], once we examine structures with infinitely many worlds, qualitative plausibility measures can capture natural ordering of events that cannot be captured by preference orderings, possibility measures, or κ-rankings.

As we have tried to argue throughout the paper, the explicit representation of knowledge and time makes it much easier to study belief dynamics. Most current work in the area examines only the beliefs of an agent and how they change after incorporating a new belief. Many simplifying assumptions are made: that there is a single agent, that the agent's knowledge does not change, that new information can be characterized in the language, and so on. It is useful to study this simple setting in order to get at the basic issues of belief change. However, these simplifying assumptions are not suitable when we want examine belief change in more realistic settings (such as the diagnosis example of Section 3.2). This means that most of the results in the current belief change literature are not directly applicable in many standard AI problems. Our framework dispenses with most of the simplifying assumptions made in the literature, and thus can be viewed as a first step towards providing a model of more realistic settings of belief change.

We have focused here on the foundations of the framework. In the future, we hope to apply the framework to examine more realistic problems. We have already begun to do this. For example, in [19] we provide a detailed analysis of iterated prisoner dilemma games between two agents. It is well-known that the players cannot cooperate when they have common knowledge of rationality. However, we show that they can cooperate when they have common belief of rationality. A recent proposal by van der Meyden [47] for multi-agent belief change can easily be embedded in our framework [48]. We hope to use our framework to study some of the problems considered by van der Meyden, such as speech-act semantics. Another natural application area is reasoning about actions and planning in the presence of uncertainty. We believe that the flexibility and expressive power of the framework will help to clarify what is going on in all these areas.

Acknowledgements

The authors are grateful to Pierpaolo Battigalli, Craig Boutilier, Ronen Brafman, Ron Fagin, Moises Goldszmidt, Ron van der Meyden, Yoav Shoham, and particularly Daphne Koller and Moshe Vardi for comments on previous versions of this paper and useful discussions relating to this work.

A Proofs

A.1 Proofs for Section 2.7

Theorem 7: *K (resp., K45, KD45) is a sound and complete axiomatization for $\mathcal{L}^B$ with respect to $\mathcal{M}$ (resp., $\mathcal{M}^{\mathrm{CONS}}$, $\mathcal{M}^{\mathrm{CONS,NORM}}$).*

Proof. As usual, soundness is straightforward, so we focus on completeness. We prove completeness by showing that for $M \in \mathcal{M}_K$ (resp. $\mathcal{M}_K^{et}$, $\mathcal{M}_K^{est}$) there is a structure $M^+ \in \mathcal{M}$

(resp. M^{CONS}, $\mathcal{M}^{\text{CONS,NORM}}$) such that for all $\phi \in \mathcal{L}^B$, we have $(M,w) \models \phi$ if and only if $(M^+,w) \models \phi$. Completeness then follows from Theorem 1.

Let $M = (W, \pi, \mathcal{B}_1, \ldots, \mathcal{B}_n)$ be a Kripke structure for belief. We construct a Kripke structure for knowledge and plausibility $M^+ = (W, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{P}_1, \ldots, \mathcal{P}_n)$ as follows. We set $\mathcal{K}_i(w)$ to be the set of worlds where agent i's beliefs are the same as in w. Formally, $(w,v) \in \mathcal{K}_i$ if $\mathcal{B}_i(w) = \mathcal{B}_i(v)$. It is easy to verify that $\mathcal{K}_i$ is an equivalence relation. We define $\mathcal{P}_i(w) = (W_{(w,i)}, \mathrm{Pl}_{(w,i)})$, where $W_{(w,i)} = \mathcal{B}_i(w)$ is the set of worlds agent i considers possible, $\mathrm{Pl}_{(w,i)}(\emptyset) = 0$, and $\mathrm{Pl}_{(w,i)}(A)$ is 1 if $A \subseteq W_{(w,i)}$ is not empty It is easy to verify that these (trivial) plausibility measures are qualitative.

We now prove that $(M,w) \models \phi$ if and only $(M^+,w) \models \phi$ for any $\phi \in \mathcal{L}^B$. This is shown by induction on the structure of ϕ. The only interesting case is if ϕ is of the form $B_i\phi'$. Assume $(M,w) \models B_i\phi'$. We want to show that $(M^+,w) \models K_i(\mathit{true} \rightarrow_i \phi')$. We start by noting that $(w,v) \in \mathcal{K}_i$ if and only if $\mathcal{B}_i(v) = \mathcal{B}_i(w)$. This implies that $\mathcal{P}_i(v) = \mathcal{P}_i(w)$. Thus, $(M^+,v) \models \mathit{true} \rightarrow_i \phi'$ if and only if $(M^+,w) \models \mathit{true} \rightarrow_i \phi'$. Thus, it suffices to show that $(M^+,w) \models \mathit{true} \rightarrow_i \phi'$, since this implies that $(M^+,w) \models K_i(\mathit{true} \rightarrow_i \phi')$, i.e., $(M^+,w) \models B_i\phi'$. There are two cases. If $\mathcal{B}_i(w) = \emptyset$, then $W_{(w,i)} = \emptyset$. This implies that $\mathit{true} \rightarrow_i \phi'$ holds vacuously. If $\mathcal{B}_i(w)$ is not empty, then using the induction hypothesis we conclude that $[\![\phi']\!]_{(w,i)} = \mathcal{B}_i(w)$. From the definition of $\mathrm{Pl}_{(w,i)}$ we conclude that $\mathrm{Pl}_{(w,i)}([\![\phi']\!]_{(w,i)}) = 1$ and that $\mathrm{Pl}_{(w,i)}([\![\neg\phi']\!]_{(w,i)}) = 0$. Thus, $(M^+,w) \models \mathit{true} \rightarrow_i \phi'$ and hence $(M^+,w) \models K_i(\mathit{true} \rightarrow_i \phi')$. Now assume $(M,w) \models \neg B_i\phi'$. Then there is some $v \in \mathcal{B}_i(w)$ such that $(M,v) \models \neg\phi'$. Using the induction hypothesis we conclude that $\mathrm{Pl}_{(w,i)}([\![\neg\phi']\!]_{(w,i)}) = 1$. Hence, $(M^+,w) \models \neg(\mathit{true} \rightarrow_i \phi')$ and therefore, $(M^+,w) \models \neg K_i(\mathit{true} \rightarrow_i \phi')$.

It remains to show that if $M \in \mathcal{M}_K^{et}$ then M^+ satisfies CONS, and if $M \in \mathcal{M}_K^{est}$, then M^+ also satisfies NORM. Assume $\mathcal{B}_i$ is transitive and Euclidean. Let w and v be worlds such that $(w,v) \in \mathcal{B}_i$. We claim that $\mathcal{B}_i(w) = \mathcal{B}_i(v)$. If $(w,t) \in \mathcal{B}_i$, then since $\mathcal{B}_i$ is Euclidean we get that $(v,t) \in \mathcal{B}_i$. If $(v,t) \in \mathcal{B}_i$, then since $\mathcal{B}_i$ is transitive we get that $(w,t) \in \mathcal{B}_i$. Thus, $\mathcal{B}_i(v) = \mathcal{B}_i(w)$, as desired. Recall that if $\mathcal{B}_i(v) = \mathcal{B}_i(w)$, then our construction ensures that $v \in \mathcal{K}_i(w)$. Hence, $\mathcal{B}_i(w) \subseteq \mathcal{K}_i(w)$ and M^+ satisfies CONS. Assume that $\mathcal{B}_i$ is serial. This implies that for all w, $\mathcal{B}_i(w)$ is not empty. Thus, our construction guarantees that $W_{(w,i)}$ is not empty and $\mathrm{Pl}_{(w,i)}(W_{(w,i)}) > \perp$. $\square$

Theorem 8: *AX^{KB} (resp., $AX^{\text{KB,CONS}}$, $AX^{\text{KB,CONS,NORM}}$) is a sound and complete axiomatization of $\mathcal{L}^{KB}$ with respect to $\mathcal{M}$ (resp., $\mathcal{M}^{\text{CONS}}$, $\mathcal{M}^{\text{CONS,NORM}}$).*

Proof. Again soundness is straightforward, so we focus on completeness. We sketch a completeness proof following the usual Makinson [46] style of proof. We describe only the parts that are different from the standard proofs. See, for example, Halpern and Moses [34] for details.

In order to prove completeness, we need only show that if the formula ϕ is consistent with the axiom system (i.e., AX^{KB}, $\mathrm{AX}^{\text{KB,CONS}}$ or $\mathrm{AX}^{\text{KB,CONS,NORM}}$) then ϕ is satisfiable in a Kripke structure of the appropriate class (i.e., $\mathcal{M}$, $\mathcal{M}^{\rm CONS}$, or $\mathcal{M}^{\rm CONS,NORM}$, respectively).

Let V be a set of formulas and AX an axiom system. We say that V is *AX-consistent* if for all $\phi_1, \ldots \phi_n \in V$, it is not the case that AX $\vdash \neg(\phi_1 \wedge \ldots \wedge \phi_n)$. The set V is a *maximal* consistent set if it is consistent, and for each formula ϕ, either $\phi \in V$ or $\neg\phi \in V$.

We now build a *canonical model* $M^{\rm KB}$ for $\mathrm{AX}^{\rm KB}$, in which every $\mathrm{AX}^{\rm KB}$-consistent formula is satisfiable. $M^{\rm KB}$ has a world w_V corresponding to every maximal $\mathrm{AX}^{\rm KB}$-consistent set V of formulas; we show that $(M^{\rm KB}, w_V) \models \phi$ if and only if $\phi \in V$.

We proceed as follows. If V is a set of formulas, define $V/K_i = \{\phi : K_i\phi \in V\}$ and $V/B_i = \{\phi : B_i\phi \in V\}$. Let $M^{\rm KB} = (W, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{P}_i, \ldots, \mathcal{P}_n)$, where

- $W = \{w_V : \text{V is a maximal AX}^{\rm KB}\text{-consistent set of formulas}\}$
- $\pi(w_V)(p) = \mathbf{true}$ if and only if $p \in V$
- $\mathcal{K}_i = \{(w_V, w_U) : V/K_i \subseteq U\}$
- $\mathcal{P}_i(w_V) = (W_{(w_V,i)}, \mathrm{Pl}_{(w_V,i)})$, where $W_{(w_V,i)} = \{w_U : V/B_i \subseteq U\}$, $\mathrm{Pl}_{(w_V,i)}(\emptyset) = 0$, and $\mathrm{Pl}_{(w_V,i)}(A) = 1$ for $A \neq \emptyset$.

Using standard arguments, it is easy to show that the $\mathcal{K}_i$'s are equivalence relations (see [34]). Using a standard induction argument, we can verify that $(M^{\rm KB}, w_V) \models \phi$ if and only if $\phi \in V$.

This construction proves completeness for $\mathrm{AX}^{\rm KB}$. To prove completeness for the other two variants we use the same construction, setting W to correspond to the maximal $\mathrm{AX}^{\rm KB,CONS}$-consistent sets (resp. $\mathrm{AX}^{\rm KB,CONS,NORM}$-consistent sets). We must show that the resulting canonical models satisfy CONS and NORM, respectively.

Let $M^{\rm KB,CONS}$ be the canonical model constructed for $\mathrm{AX}^{\rm KB,CONS}$. To show that $M^{\rm KB,CONS}$ satisfies CONS, it is enough to show that $V/K_i \subseteq V/B_i$. To show this, assume $\phi \in V/K_i$. Then $K_i\phi \in V$. Since KB2 $\in \mathrm{AX}^{\rm KB,CONS}$, we conclude that $B_i\phi \in V$, and thus $\phi \in V/B_i$.

Let $M^{\rm KB,CONS,NORM}$ be the canonical model constructed for $\mathrm{AX}^{\rm KB,CONS,NORM}$. The argument above shows that $M^{\rm KB,CONS,NORM}$ satisfies CONS. To show that it satisfies NORM, i.e., $\mathrm{Pl}_{(w,i)}(W_{(w,i)}) > \perp$, it is enough to show that V/B_i is consistent, for then there must be some U such that $V/B_i \in U$. Assume, by way of contradiction, that V/B_i is inconsistent. Then there are formulas $\phi_1, \ldots, \phi_m \in V/B_i$ such that $\vdash \neg(\phi_1 \wedge \ldots \wedge \phi_m)$. Since $\phi_1, \ldots, \phi_n \in V/B_i$, we conclude that $B_i\phi_1, \ldots, B_i\phi_m \in V$. Using the K45 axioms for B_i, standard arguments show that $B_i(\phi_1, \ldots, \phi_n) \in V$, and hence that $B_i(\mathit{false}) \in V$, which contradicts the consistency of V. $\square$

Lemma 10: *Let M be a propositional Kripke structure of knowledge and plausibility satisfying CONS and SDP. Suppose that w, i, and α are such that the most plausible worlds in $\mathcal{P}_i(w)$ are exactly those worlds in $\mathcal{K}_i(w)$ that satisfy α, i.e., $MP(\mathcal{P}_i(w)) = \{w' \in \mathcal{K}_i(w) : (M, w') \models \alpha\}$. Then for any formula $\phi \in \mathcal{L}^{KB}$ that includes only the modalities K_i and B_i,*

$(M, w) \models \phi$ *if and only if* $(M, w) \models \phi^*$*, where* ϕ^* *is the result of recursively replacing each subformula of the form* $B_i\psi$ *in* ϕ *by* $K_i(\alpha \Rightarrow \psi^*)$.

Proof. We prove by induction that for any $w' \in \mathcal{K}_i(w)$, $(M, w') \models \phi$ if and only if $(M, w') \models \phi^*$. The only interesting case is if ϕ has the from $B_i\phi'$. Suppose that $(M, w') \models B_i\phi'$. This implies that $(M, w') \models \mathit{true} \rightarrow_i \phi'$, i.e., for all $w'' \in \mathrm{MP}(\mathcal{P}_i(w'))$ we have $(M, w'') \models \phi'$. Now let $w'' \in \mathcal{K}_i(w')$. If $(M, w'') \models \neg\alpha$, then $(M, w'') \models \alpha \Rightarrow (\phi')^*$. If $(M, w'') \models \alpha$ then, by definition, $w'' \in \mathrm{MP}(\mathcal{P}_i(w))$, and since we assumed SDP, $\mathrm{MP}(\mathcal{P}_i(w')) = \mathrm{MP}(\mathcal{P}_i(w))$. Thus, we conclude that $(M, w'') \models \phi'$, and using the induction hypothesis we get that $(M, w'') \models (\phi')^*$. We conclude that all worlds in $\mathcal{K}_i(w')$ satisfy $\alpha \Rightarrow (\phi')^*$, and thus $(M, w') \models K_i(\alpha \Rightarrow (\phi')^*)$. Now assume that $(M, w') \models K_i(\alpha \Rightarrow (\phi')^*)$. Let w'' be any world in $\mathcal{K}_i(w')$. Since we assumed SDP, we have that $\mathrm{MP}(\mathcal{P}_i(w'')) = \mathrm{MP}(\mathcal{P}_i(w))$ is the set of worlds in $\mathcal{K}_i(w)$ that satisfy α. We conclude, using our induction hypothesis, that all worlds in $\mathrm{MP}(\mathcal{P}_i(w''))$ satisfy ϕ'. Hence, $(M, w'') \models \mathit{true} \rightarrow_i \phi'$. Since this is true for all $w'' \in \mathcal{K}_i(w')$ we conclude that $(M, w') \models B_i\phi'$. $\square$

A.2 Proofs for Section 2.8

Theorem 11: *AX is a sound and complete axiomatization for* $\mathcal{L}^{KC}$ *with respect to* $\mathcal{M}$.

Proof. Again, we just describe the completeness proof. This proof draws on the usual completeness proofs for S5 modal logic, and the completeness proof for conditional logic described in [17,24].

We proceed as follows. If V is a set of formulas, define $V/K_i = \{\phi : K_i\phi \in V\}$ and $V/N_i = \{\phi : N_i\phi \in V\}$. We define a canonical model $M^c = (W, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{P}_i, \ldots, \mathcal{P}_n)$ as follows:

- $W = \{w_V :$ V is a maximal AX-consistent set of formulas$\}$
- $\pi(w_V)(p) =$ **true** if and only if $p \in V$
- $\mathcal{K}_i = \{(w_V, w_U) : V/K_i \subseteq U\}$
- $\mathcal{P}_i(w_V) = (W_{(w_V,i)}, \mathcal{F}_{(w_V,i)}, \mathrm{Pl}_{(w_V,i)})$, where
  - $W_{(w_V,i)} = \{w_U : V/N_i \subseteq U\}$,
  - $\mathcal{F}_{(w_V,i)} = \{[\phi]_{(w_V,i)} : \phi \in \mathcal{L}^{KC}\}$ where $[\phi]_{(w_V,i)} = \{w_U \in W_{(w_V,i)} : \phi \in U\}$, and
  - $\mathrm{Pl}_{(w_V,i)}$ is such that $\mathrm{Pl}_{(w_V,i)}([\phi]_{(w_V,i)}) \le \mathrm{Pl}_{(w_V,i)}([\psi]_{(w_V,i)})$ if and only if $(\phi \vee \psi) \rightarrow_i \psi \in V$.

We need to verify that M^c is indeed a structure in $\mathcal{M}$. Using standard arguments it is easy to show that the $\mathcal{K}_i$ relations are equivalence relations. In [17,24] we prove that $\mathcal{P}_i(w_V)$ is a well-defined qualitative plausibility space.

Finally, we have to show that $(M^c, w_V) \models \phi$ if and only if $\phi \in V$. As usual, this is done by induction on the structure of ϕ. We use the standard argument for formulas of the form $K_i\phi$

and arguments from [17,24] for formulas of the from $\phi \rightarrow_i \psi$. We omit the details here. $\square$

Theorem 12: *Let $\mathcal{A}$ be a subset of {RANK, NORM, REF, UNIF, CONS, SDP} and let A be the corresponding subset of {C5, C6, C7, C8, C9, C10}. Then $AX \cup A$ is a sound and complete axiomatization with respect to the structures in $\mathcal{M}$ satisfying $\mathcal{A}$.*

Proof. Yet again, we focus on completeness. We obtain completeness in each case by modifying the proof of Theorem 11. We construct a canonical model as in that proof, checking consistency with the extended axiom system. The resulting structure is in $\mathcal{M}$ and has the property that $(M, w_V) \models \phi$ if and only if $\phi \in V$. We just need to show that this structure also satisfies the corresponding semantic restrictions.

First, we consider CONS and axiom C9. Assume that C9 is included as an axiom. It is easy to see that this implies that $V/N_i \subseteq V/K_i$. This implies that $W_{(w_V,i)} \subseteq \mathcal{K}_i(w_V)$ in our construction.

Now consider the relationship between SDP and C10. Assume that C10 is included as an axiom. We need to show that if $w_U \in \mathcal{K}_i(w_V)$, then $\mathcal{P}_i(w_U) = \mathcal{P}_i(w_V)$. It is enough to show that $\phi \rightarrow_i \psi \in V$ if and only if $\phi \rightarrow_i \psi \in U$, since these statements determine $\mathcal{P}_i$ in our construction. Assume $\phi \rightarrow_i \psi \in V$. Then, according to C10, $K_i(\phi \rightarrow_i \psi) \in V$, and thus $\phi \rightarrow_i \psi \in V/K_i$. Recall that $w_U \in \mathcal{K}_i(w_V)$ only if $V/K_i \subseteq U$. We conclude that $\phi \rightarrow_i \psi \in U$. The other direction follows from the fact that $\mathcal{K}_i$ is symmetric in our construction, and thus $w_V \in \mathcal{K}_i(w_U)$.

The desired relationship between RANK, NORM, REF, and UNIF and the axioms C5, C6, C7, and C8 is proved in [17,24], for a logic that does not mention knowledge. Since these conditions put restrictions on $\mathcal{P}_i(w)$ and do not involve knowledge, the proof of [17,24] goes through unchanged; we do not repeat it here. $\square$

Theorem 13: *Let $\mathcal{A}$ be a subset of {CONS, NORM, REF, SDP, UNIF, RANK}. The formula ϕ is satisfiable in a Kripke structure satisfying $\mathcal{A}$ if and only if it is satisfiable in a Kripke structure with at most $2^{|Sub(\phi)|}$ worlds.*

Proof. The proof of this theorem relies on techniques from [21]. We sketch only the main steps here. The proof is based on a standard filtration argument.

Suppose there is a structure M and a world w in M such that $(M, w) \models \phi$. Let $Sub^+(\phi) = Sub(\phi) \cup \{\neg\phi : \phi \in Sub(\phi)\}$. We say that $V \subseteq Sub^+(\phi)$ is an *atom* if for each $\phi \in Sub(\phi)$, either $\phi \in V$ or $\neg\phi \in V$. We say that a world w in M satisfies an atom V if for all $\phi \in V$, we have $(M, w) \models \phi$. It is easy to see that each world satisfies exactly one atom. Given a world w', we define $[w]$ to be the equivalence class containing all worlds that satisfy the same atom as w. For each equivalence class $[w]$, we arbitrarily choose a *representative world* $w_{[w]} \in [w]$. We define $M' = (W', \pi', \mathcal{K}'_1, \ldots \mathcal{K}'_n, \mathcal{P}'_1, \ldots, \mathcal{P}'_n)$, where $W' = \{[w] : w \in W\}$, $\pi'([w]) = \pi(w_{[w]})$,

$\mathcal{K}'_i = \{([w],[w']) : (w,w') \in \mathcal{K}_i\}$, and $\mathcal{P}'_i([w]) = (W'_{([w],i)}, \mathrm{Pl}'_{([w],i)})$, where $W'_{([w],i)} = \{[w'] : w' \in W_{(w_{[w]},i)}\}$ and $\mathrm{Pl}'_{([w],i)}(A) \leq \mathrm{Pl}'_{([w],i)}(B)$ if $\mathrm{Pl}_{(w_{[w]},i)}(A^* \cap W_{(w_{[w]},i)}) \leq \mathrm{Pl}_{(w_{[w]},i)}(B^* \cap W_{(w_{[w]},i)})$, where $A^* = \{w'' : \exists [w'] \in A, w'' \in [w']\}$. Arguments essentially identical to those of [21] show that $(M',[w]) \models \psi$ if and only if $(M,w) \models \psi$ for all $\psi \in Sub(\phi)$; we omit details here.

We now have to describe how to modify this argument to ensure that M' satisfies $\mathcal{A}$. The modifications for NORM, REF, UNIF and RANK are described in [21]. Suppose that M satisfies CONS. Let $[w'] \in W'_{([w],i)}$. By definition, $w' \in W_{(w_{[w]},i)}$. But since M satisfies CONS, we have that $w' \in \mathcal{K}_i(w_{[w]})$. By definition, we get that $[w'] \in \mathcal{K}'_i([w])$. We conclude that M' satisfies CONS. Finally, suppose that M satisfies SDP. We force M' to satisfy SDP as follows. For all worlds w, we choose a representative world $w_{\mathcal{K}_i([w])} \in \mathcal{K}'_i([w])$ such that if $(w,w') \in K_i$, then $w_{\mathcal{K}_i([w])} = w_{\mathcal{K}_i([w'])}$. We then modify the construction so that, for each world $v \in \mathcal{K}_i(w)$, we have $\mathcal{P}'_i([v]) = \mathcal{P}'_i(w_{\mathcal{K}_i([w])})$. It is easy to see that for all $\psi \rightarrow_i \chi \in Sub(\phi)$, we have that $(M,w) \models \psi \rightarrow_i \chi$ if and only if $(M, w_{\mathcal{K}_i([w])}) \models \psi \rightarrow_i \chi$. Thus, it is easy to show that after this modification we still have that $(M',[w]) \models \psi$ if and only if $(M,w) \models \psi$ for all $\psi \in Sub(\phi)$. $\square$

Theorem 15: *Let $\mathcal{A}$ be a subset of $\{CONS, NORM, REF, SDP, UNIF, RANK\}$ containing CONS and either SDP or UNIF. If ϕ talks about the knowledge and plausibility of only one agent, then ϕ is satisfiable in a Kripke structure satisfying $\mathcal{A}$ if and only if it is satisfiable in a preferential Kripke structure satisfying $\mathcal{A}$ with at most $|\mathrm{Sub}(\phi)|^3$ worlds.*

Proof. Assume $M = (W, \pi, \mathcal{K}_1, \mathcal{P}_1)$ is a structure satisfying ϕ. Since CONS is in $\mathcal{A}$, we must have that $W_{(w,1)} \subseteq \mathcal{K}_1(w))$. Without loss of generality, we can assume that $\mathcal{K}_1$ consists of one equivalence class, that is, that $\mathcal{K}_1 = W \times W$. Since CONS and SDP imply UNIF, and since $\mathcal{A}$ contains CONS and either SDP or UNIF, we conclude that M satisfies UNIF. Using techniques from [21] we can assume, without loss of generality, that for each world w, the plausibility space $\mathcal{P}_1(w)$ is preferential (i.e., induced by some preference ordering) and that $W_{(w,1)}$ has at most $|Sub(\phi)|^2$ worlds.

Choose $w_0 \in W$ such that $(M,w_0) \models \phi$. For each formula $\neg K_1\psi \in Sub(\phi)$ such that $(M,w_0) \models \neg K_1\psi$, we select a world w_ψ such that $(M,w_\psi) \models \neg\psi$. Let T be $\{w_0\} \cup \{w_\psi : \neg K_1\psi \in Sub(\phi)\}$. Note that the cardinality of T is at most $|Sub(\phi)|$. Define $M' = (W', \pi', \mathcal{K}'_1, \mathcal{P}'_1)$ by taking W' to be the union of $W'_{(w,1)}$ for each $w \in T$, taking π' to be π restricted to W', and taking $\mathcal{P}'_1(w) = \mathcal{P}_1(w)$. Clearly $|W'|$ is at most $|Sub(\phi)|^3$. A straightforward argument for all subformulas ψ of ϕ and all worlds $w' \in W'$, we have $(M,w') \models \psi$ if and only if $(M',w') \models \psi$. It follows that $(M',w_0) \models \phi$, so ϕ is satisfiable in a small preferential structure. $\square$

Theorem 16: *Let $\mathcal{A}$ be a subset of $\{CONS, NORM, REF, SDP, UNIF, RANK\}$. If $CONS \in \mathcal{A}$, but it is not the case that UNIF or SDP is in $\mathcal{A}$, then the validity problem with respect to structures satisfying $\mathcal{A}$ is complete for exponential time. Otherwise, the validity problem is complete for polynomial space.*

Proof. The proof combines ideas from [14,21,34]. We briefly sketch the main ideas here, referring the reader to the other papers for details.

The polynomial space lower bound follows from the polynomial space lower bound for logics of knowledge alone [34]. For the exponential lower bound we use exactly the lower bound described Fagin and Halpern [14] for the combination of knowledge and probability (which is in turn based on the lower bound for PDL [16]). This lower bound construction uses only formulas involving K_i and probabilistic statements of the form $w_i(\phi) = 1$ (i.e., the probability of ϕ is 1). Since $N_i\phi$ has exactly the same properties as $w_i(\phi) = 1$, the same construction applies to our logic.

In the cases where we claim a polynomial space upper bound, this is shown by proving that if a formula ϕ is satisfiable at all, it is satisfiable in a structure that looks like a tree, with polynomial branching and depth no greater than the depth of nesting of $\mathcal{K}_i$ and $\rightarrow_i$ operators in ϕ. The result now follows along similar lines to corresponding results for logics of knowledge.

Finally, the exponential time upper bound follows by showing that if a formulas is satisfiable at all, it is satisfiable in an exponential size structure that can be constructed in deterministic exponential time; the technique is similar to that used to show that logics of knowledge with common knowledge are decidable in deterministic exponential time [34] or that PDL is decidable in deterministic exponential time [52]. $\square$

Theorem 17: *Let $\mathcal{A}$ be a subset of $\{CONS, NORM, REF, SDP, UNIF, RANK\}$ containing CONS and either UNIF or SDP. For the case of one agent, the validity problem in structures satisfying $\mathcal{A}$ is co-NP-complete.*

Proof. We show that the satisfiability problem is NP-complete. It follows that the validity problem is co-NP-complete. The lower bound is immediate, since clearly the logic is at least as hard as propositional logic. For the upper bound, by Theorem 15, ϕ is satisfiable in a structure satisfying $\mathcal{A}$ if and only if ϕ is satisfiable in a structure M of size polynomial in $|\phi|$. We simply guess a structure M and check that ϕ is satisfiable. It is easy to show that model checking can be done in polynomial time (see [34,21]). $\square$

A.3 Proofs for Section 3.3

Theorem 20: *The axiom system AX^T is a sound and complete axiomatization of $\mathcal{L}^{KCT}$ with respect to $\mathcal{C}$.*

Proof. As usual, we focus on completeness. Again, we construct a canonical interpreted system $\mathcal{I}$ such that if $\phi \in \mathcal{L}^{KCT}$ is consistent, then ϕ is satisfied in $\mathcal{I}$. The outline of the proof is similar to that of Theorem 11.

We proceed as follows. Let V be a maximal AX^T-consistent set of formulas in $\mathcal{L}^{KCT}$. We define $V/\bigcirc = \{\phi : \bigcirc\phi \in V\}$. We claim that $V/\bigcirc$ is also a maximal AX^T-consistent set. To show that $V/\bigcirc$ is maximal, assume that $\phi \notin V/\bigcirc$. Then $\bigcirc\phi \notin V$. From axiom T2, we have that $\bigcirc\neg\phi \in V$, and thus, $\neg\phi \in V/\bigcirc$. This shows that $V/\bigcirc$ is maximal. To show that $V/\bigcirc$ is AX^T-consistent, assume that there are formulas $\phi_1, \ldots \phi_n \in V/\bigcirc$ such that $\vdash_{\mathrm{AX}^T} \neg(\phi_1 \wedge \ldots \wedge \phi_n)$. From K1, T1 and RT1 we get that $\mathit{false} \in V/\bigcirc$. Thus, $\bigcirc \mathit{false} \in V$. Using T2 we get that $\neg\bigcirc \mathit{true} \in V$. Using RT1, however, we get that $\bigcirc \mathit{true} \in V$, which contradicts the assumption that V is consistent. Thus, $V/\bigcirc$ is AX^T-consistent. Finally, we define $V/\bigcirc^m$ to the result of m applications of $/\bigcirc$. Repeated applications of the above argument show that $V/\bigcirc^m$ is a maximal AX^T-consistent set for all $m \geq 0$.

We construct a canonical interpreted system as follows. Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}_1, \ldots, \mathcal{P}_n)$, where

- $\mathcal{R} = \{r^V : V \subseteq \mathcal{L}^{KCT}$ is a maximal AX^T-consistent set$\}$ such that
  - $r_e^V(m) = V/\bigcirc^m$, and
  - $r_i^V(m) = (V/\bigcirc^m)/K_i$,
- $\pi(r^V, m)(p) = \mathbf{true}$ if and only if $p \in r_e^V(m)$, and
- $\mathcal{P}_i(r^V, m) = (W_{(r^V,m,i)}, \mathrm{Pl}_{(r^V,m,i)})$, where
  - $W_{(r^V,m,i)} = \{(r^U, n) : (V/\bigcirc^m)/N_i \subseteq U/\bigcirc^n\}$, and
  - $\mathrm{Pl}_{(r^V,m,i)}$ is such that $\mathrm{Pl}_{(r^V,m,i)}([\phi]_{(r^V,m,i)}) \leq \mathrm{Pl}_{(r^V,m,i)}([\psi]_{(r^V,m,i)})$ if and only if $(\phi \vee \psi) \rightarrow_i \psi \in V/\bigcirc^m$, where $[\phi]_{(r^V,m,i)} = \{(r^U, k) \in W_{(r^V,m,i)} : \phi \in U/\bigcirc^k\}$.

Using the arguments in the completeness proof for conditional logic of [17,24], we can show that $\mathcal{P}_i(r, m)$ is well-defined for all i. Finally, we have to show that $(\mathcal{I}, r^V, m) \models \phi$ if and only if $\phi \in r_e^V(m)$. As usual, this is done by induction on the structure of ϕ. This is identical to the proof in of Theorem 11 except for the $\bigcirc$ modality, which is handled by standard arguments. We omit the details here. □

Theorem 21: *Let $\mathcal{A}$ be a subset of $\{$RANK, NORM, REF, UNIF, CONS, SDP$\}$ and let A be the corresponding subset of $\{$C5, C6, C7, C8, C9, C10$\}$. Then AX$^T \cup A$ is a sound and complete axiomatization with respect to systems in $\mathcal{C}$ satisfying $\mathcal{A}$.*

Proof. Again, we focus on completeness. We obtain completeness in each case by modifying the proof of Theorem 20. We construct a canonical system as in that proof, checking consistency with the extended axiom system. The resulting system has the property that $(\mathcal{I}, r^V, m) \models \phi$ if and only if $\phi \in V/\bigcirc^m$. We just need to show that this system satisfies the corresponding semantic restrictions. The desired relationship between these semantic properties and axioms is proved in [17,24] and the proof of Theorem 12. □

A.4 Proofs for Section 4.1

Proposition 25: *Let $\mathcal{I}$ be a synchronous system satisfying perfect recall and PRIOR. If ϕ characterizes agent i's knowledge at $(r, m+1)$ with respect to his knowledge at (r, m), then $(\mathcal{I}, r, m+1) \models \psi \rightarrow_i \xi$ if and only if $(\mathcal{I}, r, m) \models \bigcirc(\phi \wedge \psi) \rightarrow_i \bigcirc\xi$.*

Proof. Expanding the definition we get that $\mathcal{R}([\![\bigcirc(\phi \wedge \psi)]\!]_{(r,m)}) = \{r' \in W_{(r,i)} : (r', m) \sim_i (r, m), (r', m+1) \models \phi \wedge \psi\}$. Similarly, we get that $\mathcal{R}([\![\psi]\!]_{(r,m+1)}) = \{r' \in W_{(r,i)} : (r', m+1) \sim_i (r, m+1), (r', m+1) \models \psi\}$. However, since ϕ characterizes agent i's knowledge at time $m+1$ with respect to his knowledge at time m, we get that $(r', m+1) \sim_i (r, m+1)$ if and only if $(r', m) \sim_i (r, m)$ and $(r, m+1) \models \phi$. We conclude that $\mathcal{R}([\![\bigcirc(\phi \wedge \psi)]\!]_{(r,m)}) = \mathcal{R}([\![\psi]\!]_{(r,m+1)})$. The lemma now follows directly from Proposition 22. $\Box$

Lemma 28: *Let $\mathcal{I}$ be a synchronous static system satisfying PRIOR, RANK, SDP, and perfect recall that has finite branching. Then $(\mathcal{I}, r, m) \models B_i\phi \equiv B_i\bigcirc B_i\phi$ for all propositional formulas ϕ.*

Proof. For all points (r, m) in $\mathcal{I}$, note that $W_{(r,m,i)} = \cup\{A_\psi\}$, where A_ψ is the set of points $(r', m) \sim_i (r, m)$ such that the agent's new knowledge at time $m+1$ is ψ. If $\mathcal{I}$ has finite branching, this is a finite partition of $W_{(r,m,i)}$. Additionally, note that if $\mathrm{Pl}_{(r,m,i)}$ is a ranking, and $C_1, \ldots, C_k$ is a finite partition of C, then since $\mathrm{Pl}_{(r,m,i)}(C) = \max_{1 \le j \le k} \mathrm{Pl}_{(r,m,i)}(C_j)$, there must be some j such that $\mathrm{Pl}_{(r,m,i)}(C_j) = \mathrm{Pl}_{(r,m,i)}(C)$. In particular, for all $C \subseteq W_{(r,m,i)}$, either $\mathrm{Pl}_{(r,m,i)}(C) = \top$ or $\mathrm{Pl}_{(r,m,i)}(W_{(r,m,i)} - C) = \top$.

For the "$\Rightarrow$" part, suppose that $(\mathcal{I}, r, m) \models B_i\phi$. If $\mathrm{Pl}_{(r,m,i)}(W_{(r,m,i)}) = \bot$, then $(\mathcal{I}, r, m) \models B_i\bigcirc B_i\phi$ vacuously. If $\mathrm{Pl}_{(r,m,i)}(W_{(r,m,i)}) \neq \bot$, then $\mathrm{Pl}_{(r,m,i)}([\![\phi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\neg\phi]\!]_{(r,m,i)})$. Assume that ψ is such that $\mathrm{Pl}_{(r,m,i)}(A_\psi) = \top$. It is easy to verify that since $\mathrm{Pl}_{(r,m,i)}$ is a ranking, we get that $\mathrm{Pl}_{(r,m,i)}(A_\psi \cap [\![\phi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}(A_\psi \cap [\![\neg\phi]\!]_{(r,m,i)})$. Let r' be a run such that $(r', m) \in A_\psi$. By SDP, we get that $\mathrm{Pl}_{(r,m,i)} = \mathrm{Pl}_{(r',m,i)}$, and thus $\mathrm{Pl}_{(r',m,i)}(A_\psi \cap [\![\phi]\!]_{(r',m,i)}) > \mathrm{Pl}_{(r',m,i)}(A_\psi \cap [\![\neg\phi]\!]_{(r',m,i)})$. By definition of A_ψ, we have that $(r'', m+1) \sim_i (r', m+1)$ if and only if $(r'', m) \in A_\psi$. Since $\mathcal{I}$ satisfies PRIOR, $\mathrm{Pl}_{(r',m+1,i)}$ is the result of conditioning $\mathrm{Pl}_{(r,m,i)}$ on A_ψ. Moreover, since propositions are static, we get that $\mathrm{Pl}_{(r',m+1,i)}([\![\phi]\!]_{(r',m+1,i)}) > \mathrm{Pl}_{(r',m+1,i)}([\![\neg\phi]\!]_{(r',m+1,i)})$. Thus, $(\mathcal{I}, r', m) \models \bigcirc B_i\phi$. We conclude that $A_\psi \subseteq [\![\bigcirc B_i\phi]\!]_{(r,m,i)}$, and thus $\mathrm{Pl}_{(r,m,i)}([\![\bigcirc B_i\phi]\!]_{(r,m,i)}) = \top$. Moreover, since $A_\psi \subseteq [\![\bigcirc B_i\phi]\!]_{(r,m,i)}$ for all A_ψ such that $\mathrm{Pl}_{(r,m,i)}(A_\psi) = \top$, we get that $\mathrm{Pl}_{(r,m,i)}([\![\neg\bigcirc B_i\phi]\!]_{(r,m,i)}) \le \max\{\mathrm{Pl}_{(r,m,i)}(A_\psi) : \mathrm{Pl}_{(r,m,i)}(A_\psi) < \top\} < \top$. We conclude that $\mathrm{Pl}_{(r,m,i)}([\![\bigcirc B_i\phi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\neg\bigcirc B_i\phi]\!]_{(r,m,i)})$, and thus, $(\mathcal{I}, r, m) \models B_i\bigcirc B_i\phi$.

For the "$\Leftarrow$" part, suppose that $(\mathcal{I}, r, m) \models B_i\bigcirc B_i\phi$. If $\mathrm{Pl}_{(r,m,i)}(W_{(r,m,i)}) = \bot$, then $(\mathcal{I}, r, m) \models B_i\phi$ vacuously. If $\mathrm{Pl}_{(r,m,i)}(W_{(r,m,i)}) \neq \bot$, then $\mathrm{Pl}_{(r,m,i)}([\![\bigcirc B_i\phi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\neg\bigcirc B_i\phi]\!]_{(r,m,i)})$. Thus, since $\mathrm{Pl}_{(r,m,i)}$ is a ranking, $\mathrm{Pl}_{(r,m,i)}([\![\bigcirc B_i\phi]\!]_{(r,m,i)}) = \top$. Let (r', m) be some point in A_ψ for some ψ. By SDP, we have that $(\mathcal{I}, r', m) \models \bigcirc B\phi$ if and only if $(\mathcal{I}, r'', m) \models \bigcirc B\phi$ for all points $(r'', m) \in A_\psi$. Thus, $[\![\bigcirc B_i\phi]\!]_{(r,m,i)} = A_{\psi_1} \cup \ldots \cup A_{\psi_k}$ for some $\psi_1, \ldots, \psi_k$. Since $\mathrm{Pl}_{(r,m,i)}([\![\bigcirc B_i\phi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\neg\bigcirc B_i\phi]\!]_{(r,m,i)})$, we get that $\mathrm{Pl}_{(r,m,i)}(A_\psi) = \top$ only if $\psi = \psi_j$ for some $1 \leq j \leq k$. Moreover, since $A_{\psi_1}, \ldots, A_{\psi_k}$ is a finite partition of $[\![\bigcirc B_i\phi]\!]_{(r,m,i)}$, there must be at least one $1 \leq j \leq k$ such that $\mathrm{Pl}_{(r,m,i)}(A_{\psi_j}) = \top$. Let ψ_j be such that $\mathrm{Pl}_{(r,m,i)}(A_{\psi_j}) = \top$. Suppose that $(r', m) \in A_{\psi_j}$. Then we have that $\mathrm{Pl}_{(r',m+1,i)}([\![\phi]\!]_{(r',m+1,i)}) > \mathrm{Pl}_{(r',m+1,i)}([\![\neg\phi]\!]_{(r',m+1,i)})$. Since $\mathcal{I}$ is synchronous, static, and satisfies perfect recall, PRIOR, and SDP, we get that $\mathrm{Pl}_{(r,m,i)}(A_{\psi_j} \cap [\![\phi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}(A_{\psi_j} \cap [\![\neg\phi]\!]_{(r,m,i)})$. Since $\mathrm{Pl}_{(r,m,i)}$ is a ranking, we get that $\mathrm{Pl}_{(r,m,i)}(A_{\psi_j} \cap [\![\phi]\!]_{(r,m,i)}) = \top$, and thus, $\mathrm{Pl}_{(r,m,i)}([\![\phi]\!]_{(r,m,i)}) = \top$. Finally, if $\mathrm{Pl}_{(r,m,i)}(A_\psi) < \top$, then $\mathrm{Pl}_{(r,m,i)}(A_\psi \cap [\![\neg\phi]\!]_{(r,m,i)}) < \top$. Thus, since $\mathrm{Pl}_{(r,m,i)}([\![\neg\phi]\!]_{(r,m,i)}) = \max_\psi \mathrm{Pl}_{(r,m,i)}(A_\psi \cap [\![\neg\phi]\!]_{(r,m,i)})$, we get that $\mathrm{Pl}_{(r,m,i)}([\![\neg\phi]\!]_{(r,m,i)}) < \top$. We conclude that $(\mathcal{I}, r, m) \models B_i\phi$. $\square$

A.5 Proofs for Section 4.2

Theorem 32: *Let $\mathcal{A}$ be a subset of $\{QUAL, NORM, REF, RANK\}$ and let $\mathcal{I}$ be a coherent synchronous system satisfying perfect recall, CONS, and $\mathcal{A}$. Then there is a synchronous system $\mathcal{I}'$ satisfying perfect recall, PRIOR, and $\mathcal{A}$, and a mapping $f : \mathcal{R} \mapsto \mathcal{R}'$ such that for all temporally linear formulas $\phi \in \mathcal{L}^{KCT}$, we have $(\mathcal{I}, r, m) \models \phi$ if and only if $(\mathcal{I}', f(r), m) \models \phi$.*

Proof. To construct $\mathcal{I}'$, we use a general technique for taking a "sum" of a sequence of plausibility spaces. Let λ be an ordinal and let $\{S_i : 0 \leq i < \lambda\}$ be a sequence of plausibility spaces, where $S_i = (W_i, \mathrm{Pl}_i)$ and the W_i's are pairwise disjoint. Define $\oplus_i S_i$ as $(\cup_i W_i, \mathrm{Pl}_{\oplus S_i})$, where $\mathrm{Pl}_{\oplus S_i}(A) \geq \mathrm{Pl}_{\oplus S_i}(B)$ if either $\mathrm{Pl}_i(A \cap W_i) = \mathrm{Pl}_i(B \cap W_i) = \perp$ for all i, or there exists some i such that $\mathrm{Pl}_i(A \cap W_i) \geq \mathrm{Pl}_i(B \cap W_i)$, $\mathrm{Pl}_i(A \cap W_i) > \perp$, and $\mathrm{Pl}_j(A \cap W_j) = \mathrm{Pl}_j(B \cap W_j) = \perp$ for all $j < i$. We can think of $\oplus_i S_i$ as a lexicographic combination of the S_i's.

Lemma 35 *(a) $\oplus_i S_i$ is a plausibility space,*

(b) if S_i is qualitative for all i, then $\oplus_i S_i$ is qualitative,

(c) if S_i is ranked for all i, then $\oplus_i S_i$ is ranked,

(d) $(\oplus_i S_i)|_C$ is isomorphic to $\oplus_i (S_i|_C)$ under the identity mapping.

(e) $(\oplus_i S_i)|_{W_j}$ is isomorphic to S_j under the identity mapping.

(f) If $W_1, \ldots, W_k = \emptyset$, then $\oplus_i S_i$ is isomorphic to $\oplus_{i \geq k+1} S_i$.

Proof. We have to show that $\leq$ is reflexive, transitive, and satisfies A1. It is easy to see that, by definition, $\leq$ is reflexive. Next, we consider transitivity. Suppose that $\mathrm{Pl}_{\oplus S_i}(A) \geq \mathrm{Pl}_{\oplus S_i}(B)$ and $\mathrm{Pl}_{\oplus S_i}(B) \geq \mathrm{Pl}_{\oplus S_i}(C)$. If $\mathrm{Pl}_i(B \cap W_i) = \perp_i$ for all i, then clearly $\mathrm{Pl}(C \cap W_i) = \perp_i$ for all i (since $\mathrm{Pl}_{\oplus S_i}(B) \geq \mathrm{Pl}_{\oplus S_i}(C)$), so $\mathrm{Pl}_{\oplus S_i}(A) \geq \mathrm{Pl}_{\oplus S_i}(C)$. So suppose that $\mathrm{Pl}(B \cap W_i) > \perp_i$ for some i. Let i and j be the smallest indexes such that $\mathrm{Pl}_i(A \cap W_i) > \perp_i$ and $\mathrm{Pl}_j(B \cap W_j) > \perp_j$.

It is easy to see that $i \leq j$, and that $\mathrm{Pl}_k(C \cap W_k) = \bot_k$ for all $k \leq j$. If $i < j$, we conclude that $\mathrm{Pl}_i(A \cap W_i) \geq \mathrm{Pl}_i(C \cap W_i) = \bot_i$, and thus $\mathrm{Pl}_{\oplus S_i}(A) \geq \mathrm{Pl}_{\oplus S_i}(C)$. On the other hand, if $i = j$, then by definition $\mathrm{Pl}_i(A \cap W_i) \geq \mathrm{Pl}_i(B \cap W_i)$, and $\mathrm{Pl}_i(B \cap W_i) \geq \mathrm{Pl}_i(C \cap W_i)$. Since $\leq$ is transitive in S_i, we get that $\mathrm{Pl}_i(A \cap W_i) \geq \mathrm{Pl}_i(C \cap W_i)$. Thus, we conclude that $\mathrm{Pl}_{\oplus S_i}(A) \geq \mathrm{Pl}_{\oplus S_i}(C)$, as desired. Finally, we consider A1. Suppose that $A \subseteq B$. Then $A \cap W_i \subseteq B \cap W_i$ for all i. Since each S_i satisfies A1, we have that $\mathrm{Pl}_i(A \cap W_i) \leq \mathrm{Pl}_i(C \cap W_i)$ for all i. It easily follows that $\mathrm{Pl}_{\oplus S_i}(A) \leq \mathrm{Pl}_{\oplus S_i}(B)$.

Suppose that S_i is qualitative for all i. We have to show that $\oplus_i S_i$ is also qualitative. We start by considering A2. Suppose that A, B, and C are pairwise disjoint sets such that $\mathrm{Pl}_{\oplus S_i}(A \cup B) > \mathrm{Pl}_{\oplus S_i}(C)$ and $\mathrm{Pl}_{\oplus S_i}(A \cup C) > \mathrm{Pl}_{\oplus S_i}(B)$. Let i and j be the minimal indexes such that $\mathrm{Pl}_i((A \cup B) \cap W_i) > \bot_i$ and $\mathrm{Pl}_j((A \cup C) \cap W_j) > \bot_j$. We claim that $i = j$. Assume, by way of contradiction, that $i < j$. Then, $\mathrm{Pl}_i((A \cup C) \cap W_i) = \bot_i$ and hence $\mathrm{Pl}_i(A \cap W_i) = \bot_i$. Moreover, since $\mathrm{Pl}_{\oplus S_i}(A \cup C) > \mathrm{Pl}_{\oplus S_i}(B)$, we get that $\mathrm{Pl}_i(B \cap W_i) = \bot_i$. Using A3 in S_i, we conclude that $\mathrm{Pl}_i((A \cup B) \cap W_i) = \bot_i$, which contradicts our assumption that $\mathrm{Pl}_i((A \cup B) \cap W_i) > \bot_i$. Symmetric arguments show that we also cannot have $j < i$. Thus, $i = j$. By definition $\mathrm{Pl}_i((A \cup B) \cap W_i) > \mathrm{Pl}_i(C \cap W_i)$ and $\mathrm{Pl}_i((A \cup C) \cap W_i) > \mathrm{Pl}_i(B \cap W_i)$. Using A2 we conclude that $\mathrm{Pl}_i(A \cap W_i) > \mathrm{Pl}_i((B \cup C) \cap W_i)$. It is also easy to verify, using A3, that $\mathrm{Pl}_j((B \cup C) \cap W_j) = \bot_j$ for all $j < i$. Thus, we get that $\mathrm{Pl}_{\oplus S_i}(A) > \mathrm{Pl}_{\oplus S_i}(B \cup C)$, as desired. Next, consider A3. The construction of $\oplus S_i$ is such that $\mathrm{Pl}_{\oplus S_i}(A) = \bot$ if and only if $\mathrm{Pl}_1(A \cap W_i) = \bot$ for all i. It is easy to see that A3 follows from A3 in each S_i.

Finally, part (c) follows immediately from the definition, part (d) follows immediately from COND, part (e) is a special case of part (d), and part (f) follows immediately from the definition. $\square$

Returning to the proof of Theorem 32, first suppose that REF is not in $\mathcal{A}$. Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}_1, \ldots, \mathcal{P}_n)$ be a coherent synchronous system satisfying perfect recall and CONS. Roughly speaking, the proof goes as follows. We construct a system $\mathcal{I}'$ which consists of countably many copies of $\mathcal{R}$. The runs in $\mathcal{R}^m$, the mth copy of $\mathcal{R}$, are used to simulate the agent's plausibility assessment at time m. More precisely, for all times m, we define a prior on $\mathcal{R}^m$ that corresponds to the agent's plausibility measure at time m in $\mathcal{I}$. These priors are then combined using $\oplus$ to construct the agent's prior in $\mathcal{I}'$. Since $\oplus$ orders the priors lexicographically, if $m < m'$, the priors on $\mathcal{R}^m$ dominate those on $\mathcal{R}^{m'}$. The construction guarantees that at time m, the agent considers possible only runs in $\mathcal{R}^m \cup \mathcal{R}^{m+1} \cup \ldots$. Since the prior on $\mathcal{R}^m$ dominates the rest, the agent's plausibility measure at time m is similar to that at time m in $\mathcal{I}$. This similarity is what guarantees that conditional formulas are evaluated in the same way in $\mathcal{I}$ and $\mathcal{I}'$. This "peeling away" of copies of $\mathcal{R}$ ensures that all temporally linear formulas holding in runs in $\mathcal{I}$ are also satisfied in the corresponding runs in $\mathcal{I}'$.

The formal construction proceeds as follows. Let $R \subseteq \mathcal{R}$ and $l \in I\!N^*$ (recall that $I\!N^* = I\!N \cup \{\infty\}$). Define $R^l = \{r^l : r \in R\}$, where, for each $i \in \{e, 1 \ldots, n\}$, we have

$$r_i^l(m) = \begin{cases} \langle r_i(m), m\rangle & \text{if } l \geq m \\ \langle r_i(l), m\rangle & \text{if } l < m. \end{cases}$$

Let $\mathcal{I}' = (\mathcal{R}', \pi', \mathcal{P}'_1, \ldots, \mathcal{P}'_n)$, where $\mathcal{R}' = \cup_{l \in I\!N^*} \mathcal{R}^l$, π' is defined so that if $m \leq l$ then $\pi'(r^l, m) = \pi(r, m)$ and if $m > l$ then $\pi(r^l, m) = \pi(r, l)$, and $\mathcal{P}'_i$ is defined by the priors described below.

To define a prior on $\mathcal{R}'$, we first define a plausibility space $\mathcal{P}^m_{(r,i)}$ on $\mathcal{R}^m$ for each $m \in I\!N$, run $r \in \mathcal{R}$, and agent i. We want the time m projection of $\mathcal{P}^m_{(r,i)}$ to be isomorphic to $\mathcal{P}_i(r, m)$. To achieve this, we define $\mathcal{P}^m_{(r,i)} = (\mathcal{R}^m_{(r,i)}, \mathrm{Pl}^m_{(r,i)})$, where $\mathcal{R}^m_{(r,i)} = \mathcal{R}(W_{(r,m,i)})^m$ and $\mathrm{Pl}^m_{(r,i)}$ is defined so that for $A \subseteq W_{(r,m,i)}$, we have $\mathrm{Pl}^m_{(r,i)}((\mathcal{R}(A))^m) = \mathrm{Pl}_{(r,m,i)}(A)$. For $l \in I\!N^*$, we define the prior of agent i at run r^l to be the combination of these priors for all time points: $\mathcal{P}'_{(r^l,i)} = \oplus_m \mathcal{P}^m_{(r,i)}$.

It is easy to see that $\mathcal{I}'$ is synchronous. It is also easy to check that $\mathcal{I}'$ satisfies perfect recall: From the definition, we have that

$$\mathcal{K}'_i(r^l, m) = \begin{cases} \{(r'^{l'}, m) : (r', m) \in \mathcal{K}_i(r, m), l' \geq m\} & \text{if } l \geq m \\ \{(r'^l, m) : (r', l) \in \mathcal{K}_i(r, l)\} & \text{if } l < m. \end{cases}$$

Moreover, since $\mathcal{I}$ satisfies perfect recall, we have that $\mathcal{R}(\mathcal{K}_i(r, m+1)) \subseteq \mathcal{R}(\mathcal{K}_i(r, m))$. We conclude that $\mathcal{R}'(\mathcal{K}'_i(r^l, m+1)) \subseteq \mathcal{R}'(\mathcal{K}'_i(r^l, m))$, which is just what we need for perfect recall.

Let $\phi \in \mathcal{L}^{KC}$ (so that ϕ does not include any temporal modalities) and $l \geq m$. We show that $(\mathcal{I}', r^l, m) \models \phi$ if and only if $(\mathcal{I}, r, m) \models \phi$. As usual we prove this by induction on the structure of ϕ. The only interesting cases are these that directly involve modalities.

We start with the K_i modality. Suppose that $(\mathcal{I}, r, m) \models K_i\phi$. Then for all points $(s, m) \in \mathcal{K}_i(r, m)$, we have $(\mathcal{I}, s, m) \models \phi$. Let $(s^k, m) \in \mathcal{K}'_i(r^l, m)$. From the definition of $\mathcal{I}'$ we get that $(s, m) \in K_i(r, m)$ and $k \geq m$. Using the induction hypothesis, we get that $(\mathcal{I}', s^k, m) \models \phi$. We conclude that $(\mathcal{I}', r^l, m) \models K_i\phi$. Now suppose that $(\mathcal{I}, r, m) \not\models K_i\phi$. Then there is a point $(s, m) \in \mathcal{K}_i(r, m)$ such that $(\mathcal{I}, s, m) \models \neg\phi$. Using the induction hypothesis we conclude that $(\mathcal{I}, s^m, m) \models \neg\phi$. Since $(s^m, m) \in \mathcal{K}'_i(r^l, m)$, we conclude that $(\mathcal{I}', r^l, m) \not\models K_i\phi$.

We now turn to the $\rightarrow_i$ modality. The definition of PRIOR implies that $\mathcal{P}'_i(r^l, m)$ is the projection of $\mathcal{P}'_{(r^l,i)}$ conditioned on $\mathcal{R}'(\mathcal{K}'_i(r^l, m))$. Now $\mathcal{P}'_{(r^l,i)} = \oplus_m \mathrm{Pl}^m_{(r,i)}$. Parts (d) and (f) of Lemma 35 imply that $\mathcal{P}'_{(r^l,i)}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))}$ is isomorphic to $\oplus_{k \geq m}(\mathcal{P}^k_{(r,i)})|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))})$. Consider the first term in the "sum", $\mathcal{P}^m_{(r,i)}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))}$. Since $\mathcal{I}$ satisfies CONS, we have that $W_{(r,m,i)} \subseteq \mathcal{K}_i(r, m)$. Thus, conditioning on $\mathcal{R}'(K'_i(r^l, m))$ does not remove any runs from $\mathcal{R}^m_{(r,i)} = (\mathcal{R}(W_{(r,m,i)})^m$. It follows that $\mathcal{P}^m_{(r,i)}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))} = \mathcal{P}^m_{(r,i)}$ which is isomorphic to $\mathcal{P}_i(r, m)$

under the mapping $r'^m \mapsto (r', m)$. Finally, since $\mathcal{P}^m_{(r,i)}$ is the first plausibility space in the "sum", it determines the ordering of all pairs of sets, unless both of them are assigned plausibility $\perp$ by $\mathrm{Pl}^m_{(r,i)}$. Putting all this together, we conclude that if $A', B' \subseteq W'_{(r^l,m,i)}$ and $A, B \subseteq W_{(r,m,i)}$ such that $(\mathcal{R}(A))^m = \mathcal{R}'(A') \cap \mathcal{R}^m$ and $(\mathcal{R}(B))^m = \mathcal{R}'(B') \cap \mathcal{R}^m$, and if $\mathrm{Pl}_{(r,m,i)}(A) > \perp$, then $\mathrm{Pl}'_{(r^l,m,i)}(A') \geq \mathrm{Pl}'_{(r^l,m,i)}(B')$ if and only if $\mathrm{Pl}_{(r,m,i)}(A) \geq \mathrm{Pl}_{(r,m,i)}(B)$.

Assume that $(\mathcal{I}, r, m) \models \phi \rightarrow_i \psi$. Thus, either $\mathrm{Pl}_{(r,m,i)}([\![\phi]\!]_{(r,m,i)}) = \perp$ or $\mathrm{Pl}_{(r,m,i)}([\![\phi \wedge \psi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\phi \wedge \neg\psi]\!]_{(r,m,i)})$. If $\mathrm{Pl}_{(r,m,i)}([\![\phi]\!]_{(r,m,i)}) = \perp$, then from the coherence of $\mathcal{I}$ it follows that if $A \subseteq W_{(r,l',i)}$ and $\mathcal{R}(A) \subseteq \mathcal{R}([\![\phi]\!]_{(r,m,i)})$, then $\mathrm{Pl}_{(r,l',i)}(A) = \perp$. This implies that $\mathrm{Pl}^{l'}_{(r,i)}((\mathcal{R}([\![\phi]\!]_{(r,m,i)})^{l'} \cap \mathcal{R}(W_{(r,l',i)})^{l'}) = \perp$ for all $l' \geq m$. Since $\mathcal{K}'_i(r^l, m)$ contains only points from $\mathcal{R}^{l'}$ for $l' \geq m$, we get that $\mathrm{Pl}'_{(r^l,m,i)}([\![\phi]\!]_{(r^l,m,i)}) = \perp$. Thus, we conclude that $(\mathcal{I}', r^l, m) \models \phi \rightarrow_i \psi$ in this case. Now suppose that $\mathrm{Pl}_{(r,m,i)}([\![\phi \wedge \psi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\phi \wedge \neg\psi]\!]_{(r,m,i)})$. If we could show that $(\mathcal{R}([\![\phi]\!]_{(r,m,i)}))^m = \mathcal{R}'([\![\phi]\!]_{(r^l,m,i)}) \cap \mathcal{R}^m$, and similarly for ψ, then we could apply the argument of the previous paragraph to show that $\mathrm{Pl}'_{(r^l,m,i)}([\![\phi \wedge \psi]\!]_{(r^l,m,i)}) > \mathrm{Pl}'_{(r^l,m,i)}([\![\phi \wedge \neg\psi]\!]_{(r^l,m,i)})$. This, in turn, would allow us to conclude that $(\mathcal{I}', r^l, m) \models \phi \rightarrow_i \psi$. The fact that $(\mathcal{R}([\![\phi]\!]_{(r,m,i)}))^m = \mathcal{R}'([\![\phi]\!]_{(r^l,m,i)}) \cap \mathcal{R}^m$ follows from the following chain of equivalences:

$$\begin{array}{ll}
 & s^m \in (\mathcal{R}([\![\phi]\!]_{(r,m,i)}))^m \\
\text{iff} & (s, m) \in [\![\phi]\!]_{(r,m,i)} \\
\text{iff} & (s, m) \in W_{(r,m,i)} \text{ and } (\mathcal{I}, s, m) \models \phi \\
\text{iff} & s^m \in (\mathcal{R}(W_{(r,m,i)}))^m = \mathcal{R}^m_{(r,i)} \text{ and (by the induction hypothesis) } (\mathcal{I}', s^m, m) \models \phi \\
\text{iff} & (s^m, m) \in W_{(r^l,m,i)} \text{ and } (\mathcal{I}', s^m, m) \models \phi \\
\text{iff} & (s^m, m) \in [\![\phi]\!]_{(r^l,m,i)} \\
\text{iff} & s^m \in \mathcal{R}([\![\phi]\!]_{(r^l,m,i)}) \cap \mathcal{R}^m.
\end{array}$$

Thus, in either case, we conclude that $(\mathcal{I}', r^l, m) \models \phi \rightarrow_i \psi$, as desired.

For the converse, suppose that $(\mathcal{I}, r, m) \not\models \phi \rightarrow_i \psi$. Then $\mathrm{Pl}_{(r,m,i)}([\![\phi]\!]_{(r,m,i)}) > \perp$ and $\mathrm{Pl}_{(r,m,i)}([\![\phi \wedge \psi]\!]_{(r,m,i)}) \not> \mathrm{Pl}_{(r,m,i)}([\![\phi \wedge \neg\psi]\!]_{(r,m,i)})$. By the same arguments as above, we get that $\mathrm{Pl}'_{(r^l,m,i)}([\![\phi \wedge \psi]\!]_{(r^l,m,i)}) > \perp$ and $\mathrm{Pl}'_{(r^l,m,i)}([\![\phi \wedge \psi]\!]_{(r^l,m,i)}) \not> \mathrm{Pl}'_{(r^l,m,i)}([\![\phi \wedge \neg\psi]\!]_{(r^l,m,i)})$. Thus, $(\mathcal{I}', r^l, m) \not\models \phi \rightarrow_i \psi$, as desired.

Finally, for $r \in \mathcal{R}$, define $f(r) = r^\infty$. We have proved that if $\phi \in \mathcal{L}^{KC}$, then $(\mathcal{I}, r, m) \models \phi$ if and only if $(\mathcal{I}', f(r), m) \models \phi$. Since this holds for all m, a straightforward argument by induction on structure shows that this holds, not just for formulas in $\mathcal{L}^{KC}$, but for all temporally linear formulas.

We now have to ensure that $\mathcal{I}'$ satisfies $\mathcal{A}$. Suppose that $\mathcal{I}$ satisfies QUAL. Thus, $\mathcal{P}_i(r, m)$ is qualitative for all agents i, runs $r \in R$, and times m. Using part (b) of Lemma 35, we conclude that the prior $\mathcal{P}'_{(r,i)}$ is qualitative for all agents i and runs $r \in R$. This implies, using Proposition 29, that $\mathcal{I}'$ satisfies QUAL. Similarly, if $\mathcal{I}$ satisfies RANK, using part (c) of Lemma 35 and Proposition 29, we get that $\mathcal{I}'$ satisfies RANK.

Suppose that $\mathcal{I}$ satisfies NORM. Then $\mathrm{Pl}_{(r,m,i)}([\![true]\!] > \perp$ for all agents i, runs $r \in \mathcal{R}$, and times m. This implies that $\neg(true \rightarrow_i false)$ is valid in $\mathcal{I}$. Suppose that $l \geq m$. Then since $\neg(true \rightarrow_i false) \in \mathcal{L}^{KC}$, we conclude from the proof above that $(\mathcal{I}', r^l, m) \models \neg(true \rightarrow_i false)$. Thus, $\mathrm{Pl}'_{(r^l,m,i)}([\![true]\!]_{(r^l,m,i)}) > \perp$. Suppose that $l < m$. By definition, we have that $\mathcal{R}'(\mathcal{K}'_i(r^l, m)) = (\mathcal{R}(\mathcal{K}_i(r, l)))^l$. Using part (e) of Lemma 35, we get that $\mathcal{P}'_{(r,i)}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))}$ is isomorphic to $\mathcal{P}^l_{(r,i)}$. However, the latter plausibility space is isomorphic to $\mathcal{P}_i(r, l)$. Thus, it satisfies $\top > \perp$. We conclude that $\mathcal{I}'$ satisfies NORM, as desired.

Up to now we have assumed that REF is not in $\mathcal{A}$. If REF is in $\mathcal{A}$, then REF does not hold for $\mathcal{A}$, although it does hold at many points. To understand the issue, suppose that REF holds in $\mathcal{I}$. Since $\mathcal{I}'$ satisfies PRIOR, to show that REF holds in $\mathcal{I}'$, according to Proposition 29 it suffices to show that all priors satisfy REF. This is indeed the case if $l \neq \infty$. For suppose that $r^l \in A \subseteq \mathcal{R}'$. We want to show that $\mathrm{Pl}_{(r^l,i)}(A) > \perp$. Recall that $\mathcal{P}'_{(r^l,i)} = \oplus_m \mathcal{P}^m_{(r,i)}$. From the definition of $\oplus$, it easily follows that if $\mathrm{Pl}^l_{(r,i)}(A \cap \mathcal{R}^l) > \perp$, then $\mathrm{Pl}'_{(r^l,i)}(A) > \perp$. By definition, we have that $\mathrm{Pl}'_{(r^l,i)}(A \cap \mathcal{R}^l) = \mathrm{Pl}_{(r,l,i)}(A')$, where $A' = \{(s, l) : s^l \in A\}$. Clearly $(r, l) \in A'$, since $(r^l, m) \in A$. Since $\mathcal{I}$ satisfies REF, we must have that $\mathrm{Pl}_{(r,l,i)}(A') > \perp$. It follows that $\mathrm{Pl}'_{(r^l,i)}$ satisfies REF if $l \neq \infty$. This argument breaks down if $l = \infty$. Indeed, it is clear that $\mathcal{P}'_{(r^\infty,i)}$ does *not* satisfy REF. Since $\mathcal{R}^\infty$ is disjoint from $\mathcal{R}^m$ for $m < \infty$, and we only "sum" $\mathcal{P}^m_{(r,i)}$ for $m < \infty$ to obtain $\mathcal{P}'_{(r^\infty,i)}$, it follows that $\mathcal{R}^\infty$ is disjoint from $W'_{(r^\infty,i)}$, so REF does not hold.

Fortunately, a slight modification of the construction of $\mathcal{I}'$ can be used to deal with the case REF $\in \mathcal{A}$. Define $\mathcal{P}^\infty_{(r,i)} = (R^\infty_{(r,i)}, \mathrm{Pl}^\infty_{(r,i)})$, where $R^\infty_{(r,i)} = \{r^\infty\}$ and $\mathrm{Pl}^\infty_{(r,i)}(\{r^\infty\}) > \perp$. Modify the construction of $\mathcal{I}'$ so that the prior of agent i in run r^l is $\mathcal{P}''_{(r^l,i)} = \mathcal{P}'_{(r^l,i)} \oplus \mathcal{P}^\infty_{(r^l,i)}$. (Thus, $\mathcal{P}''_{(r^l,i)} = \oplus_{m \leq \infty} \mathcal{P}^m_{(r^l,i)}$.) It is easy to check that $\mathcal{I}'$ now does satisfy REF. The argument in the case that $l \neq \infty$ remains unchanged. On the other hand, if $r^\infty \in A \subseteq \mathcal{R}'$, it is immediate that $\mathcal{P}^\infty(A \cap \mathcal{R}^\infty) > \perp$, so we can now deal with this case as well. If QUAL, RANK, or NORM is in $\mathcal{A}$, it is easy to see (using the same argument as above) that $\mathcal{I}'$ also satisfies QUAL, RANK, or NORM.

It remains to show that this modification of the prior does not affect the evaluation of formulas. That is, we must show that $(\mathcal{I}, r, m) \models \phi$ if and only if $(\mathcal{I}', r^l, m) \models \phi$ for all $l \geq m$. Again, we proceed by induction on the structure of formulas. The argument for formulas of the form $K_i\phi$ goes through unchanged, since the changes to Pl' did not affect the $\mathcal{K}_i$ relations. The argument for formulas of the form $\phi \rightarrow_i \psi$ goes through with almost no change. The only case that requires attention is if $(\mathcal{I}, r, m) \models \phi \rightarrow_i \psi$ and $[\![\phi]\!]_{(r,m,i)} = \perp$. Our earlier arguments showed that $\mathrm{Pl}^{l'}_{(r,i)}((\mathcal{R}([\![\phi]\!]_{(r,m,i)})^{l'} \cap \mathcal{R}(W_{(r,l',i)}))^{l'}) = \perp$ for all $l' \geq m$, $l' \neq \infty$. These arguments go through without change. We must now show that this also holds if $l' = \infty$. But, from the definition of Pl^∞, we get that $\mathrm{Pl}^\infty_{(r,i)}((\mathcal{R}([\![\phi]\!]_{(r,m,i)})^\infty \cap \mathcal{R}^\infty) = \perp$ unless $r^\infty \in \mathcal{R}([\![\phi]\!]_{(r,m,i)})^\infty$. This implies that $(r, m) \in [\![\phi]\!]_{(r,m,i)}$. But this cannot happen, since $\mathrm{Pl}_{(r,m,i)}([\![\phi]\!]_{(r,m,i)}) = \perp$ and $\mathcal{I}$ satisfies REF. $\square$

Theorem 34: *Let $\mathcal{A}$ be a subset of $\{QUAL, NORM, REF, SDP, UNIF, RANK\}$ and let $\mathcal{I}$ be a coherent synchronous system satisfying perfect recall, CONS, PERSIST, and $\mathcal{A}$. Then there is a synchronous system $\mathcal{I}'$ satisfying perfect recall, PRIOR, and $\mathcal{A}$, and a mapping $f : \mathcal{R} \mapsto \mathcal{R}'$ such that for all temporally linear formulas $\phi \in \mathcal{L}^{KCT}$, $(\mathcal{I}, r, m) \models \phi$ if and only if $(\mathcal{I}', f(r), m) \models \phi$.*

Proof. Suppose that $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}_1, \ldots, \mathcal{P}_n)$ is a coherent synchronous system satisfying perfect recall, CONS, PERSIST, and $\mathcal{A}$. If neither CONS nor UNIF are in $\mathcal{A}$, then Theorem 32 guarantees that there is a system $\mathcal{I}'$ that satisfies the stated properties.

Suppose that UNIF $\in \mathcal{A}$, but SDP, REF $\notin \mathcal{A}$. (We sketch the modifications required to deal with SDP and REF below.) It does not follow that the system $\mathcal{I}'$ constructed in the proof satisfies UNIF. To see why, suppose r, r' and $m > k$ are such that $(r', k) \in W_{(r,k,i)}$ but $(r, m) \not\sim_i (r', m)$. UNIF implies that $\mathcal{P}_i(r, k) = \mathcal{P}_i(r', k)$ and (since $\mathcal{I}$ also satisfies CONS) that $W_{(r,m,i)} \cap W_{(r',m,i)} = \emptyset$. Hence, our construction guarantees that $\mathcal{P}'_{(r^k,i)} \neq \mathcal{P}'_{(r'^k,i)}$, although $r'^k \in W'_{(r^k,i)}$. Thus, the prior in $\mathcal{I}'$ does not satisfy UNIF. It follows that $\mathcal{I}'$ does not satisfy UNIF either, for $\mathcal{P}'_i(r^k, k) \neq \mathcal{P}'_i(r'^k, k)$, although $(r'^k, k) \in W'_{(r^k,k,i)}$.

The solution to this problem is relatively straightforward. We modify our construction so that the prior does indeed satisfy UNIF. In particular, we modify the prior $\mathcal{P}'$ to ensure that if $\mathcal{P}_i(r, k) = \mathcal{P}_i(r', k)$, then $\mathcal{P}'_{(r^k,i)} = \mathcal{P}'_{(r'^k,k)}$. Of course, we have to do so carefully, so as to make sure that nothing goes wrong with the rest of the argument in Theorem 32.

We start with a modification of the construction of $\oplus$ that takes sets (rather than sequences) of plausibility spaces and returns a new plausibility space.

Lemma 36 *Let $\mathcal{S}$ be a set of plausibility spaces such that the sets $\{W : (W, \mathrm{Pl}) \in \mathcal{S}\}$ are pairwise disjoint. Then there is a plausibility space $\otimes\mathcal{S}$ such that*

(a) if $S = (W, \mathrm{Pl}) \in \mathcal{S}$, then $\otimes\mathcal{S}|_W$ is isomorphic to S under the identity mapping,
(b) if S is qualitative for all $S \in \mathcal{S}$, then $\otimes\mathcal{S}$ is qualitative,
(c) if S is ranked for all $S \in \mathcal{S}$, then $\otimes\mathcal{S}$ is ranked.

Proof. Without loss of generality there is an ordinal λ and a sequence $\{S_i : 0 \leq i < \lambda\}$ such that $S_i \in \mathcal{S}$ for all i, and for all $S \in \mathcal{S}$, exists an i such that $S = S_i$.[17] Define $\otimes\mathcal{S} = \oplus_i S_i$. Part (a) of Lemma 35 guarantees that $\otimes\mathcal{S}$ is a plausibility space. Parts (a), (b), and (c) follow immediately from parts (e), (b), and (c) of Lemma 35, respectively. $\square$

[17] If $\mathcal{S}$ is uncountable, this construction may require the axiom of choice. There is a variant of the construction that does not require the axiom of choice, but the additional complexities involved do not seem worth the trouble.

Recall that to satisfy UNIF and PRIOR, it suffices to find a partition of R such that all the runs in each cell have the same prior. We now examine a possible way of partitioning the runs in the system. Let $r \in R$. Define $[r,m]_i = \{(r',m) : (r',m) \sim_i (r,m), \mathcal{P}_i(r',m) = \mathcal{P}_i(r,m)\}$. Thus, $[r,m]_i$ is the set of points in which agent i has the same knowledge state and plausibility assessment as at (r,m). (Note that if $W_{(r,m,i)} \neq \emptyset$, then since $\mathcal{I}$ satisfies CONS, $\mathcal{P}_i(r',m) = \mathcal{P}_i(r,m)$ implies that $(r',m) \sim_i (r,m)$.)

Lemma 37 *(a) For all times m, the collection $\{\mathcal{R}([r,m]_i) : r \in \mathcal{R}\}$ is a partition of $\mathcal{R}$.*

(b) For all times m and runs r, $W_{(r,m,i)} \subseteq [r,m]_i$.

(c) For all times m and runs r, $\mathcal{R}([r,m+1]_i) \subseteq \mathcal{R}([r,m]_i)$.

(d) For all times m and runs r, r' such that $(r',0) \in [r,0]_i$, if $(r',m) \sim_i (r,m)$, then $(r',m) \in [r,m]_i$.

Proof. By definition, if $(r',m) \in [r,m]_i$, then $[r',m] = [r,m]_i$. Thus, if $[r,m]_i \neq [r',m]_i$, then $[r,m]_i \cap [r',m]_i = \emptyset$. Part (a) follows immediately. For part (b), suppose that $(r',m) \in W_{(r,m,i)}$. Since $\mathcal{I}$ satisfies CONS, we have that $(r',m) \sim_i (r,m)$. Moreover, since $\mathcal{I}$ satisfies UNIF, we have that $\mathcal{P}_i(r',m) = \mathcal{P}_i(r,m)$. Thus, $(r',m) \in [r,m]_i$. We conclude that $W_{(r,m,i)} \subseteq [r,m]_i$, as desired. For part (c), suppose that $(r',m+1) \in [r,m+1]_i$. This implies that $(r',m+1) \sim_i (r,m+1)$ and $\mathcal{P}_i(r',m+1) = \mathcal{P}_i(r,m+1)$. Since $\mathcal{I}$ satisfies perfect recall, we get that $(r',m) \sim_i (r,m)$. Moreover, since $\mathcal{I}$ satisfies PERSIST, we get that $\mathcal{P}_i(r',m) = \mathcal{P}_i(r,m)$. We conclude that $(r',m) \in [r,m]_i$. Thus, $\mathcal{R}([r,m+1]_i) \subseteq \mathcal{R}([r,m]_i)$, as desired. Finally, we prove part (d) by induction on m. When $m = 0$, part (d) obviously holds. Suppose that $m > 0$, $(r',0) \in [r,0]_i$, and $(r',m) \sim_i (r,m)$. Since $\mathcal{I}$ satisfies perfect recall, we have that $(r',m-1) \sim_i (r,m-1)$. Using the induction hypothesis, we get that $(r',m-1) \in [r,m-1]$. This implies that $\mathcal{P}_i(r',m-1) = \mathcal{P}_i(r,m-1)$. Using PERSIST, we conclude that $\mathcal{P}_i(r',m) = \mathcal{P}_i(r,m)$. Thus, $(r',m) \in [r,m]_i$, as desired. $\square$

Using both $\oplus$ and $\otimes$, we now construct a prior over $\mathcal{R}'$ that satisfies UNIF. For $r \in R$, let $[r]_i$ abbreviate $\mathcal{R}([r,0]_i)$. Define $\mathcal{P}^m_{[r]_i} = \otimes\{\mathcal{P}^m_{(r',i)} : r' \in [r]_i\}$, where $\mathcal{P}^m_{(r,i)} = (\mathcal{R}^m_{(r,i)}, \mathrm{Pl}^m_{(r,i)})$ is the prior defined in the proof of Theorem 32 that is isomorphic to $\mathcal{P}_i(r,m)$ under the mapping $r'^m \mapsto (r',m)$. We must show that $\mathcal{P}^m_{(r,i)}$ is well defined; that is, we must show that if $\mathcal{P}^m_{(r',i)} \neq \mathcal{P}^m_{(r'',i)}$, then $\mathcal{R}^m_{(r',i)}$ is disjoint from $\mathcal{R}^m_{(r'',i)}$. Note that if $(r',m) \in [r'',m]_i$, then $\mathcal{P}^m_{(r',i)}$ and $\mathcal{P}^m_{(r'',i)}$ are identical. Using part (b) of Lemma 37 we get that if $(r',m) \notin [r'',m]_i$, then $\mathcal{R}^m_{(r',i)} \cap \mathcal{R}^m_{(r'',i)} = \emptyset$, as desired. Thus, $\mathcal{P}^m_{[r]_i}$ is indeed well defined. We now define $\mathcal{P}'_{(r^l,i)} = \oplus_m \mathcal{P}^m_{[r]_i}$ as the prior of agent i in run r^l.

We claim that this family of priors satisfies UNIF. Notice that $W'_{(r^l,i)} = \cup_{m,r' \in [r]_i} \mathcal{R}^m_{(r',i)}$. If $r'^m \in W'_{(r^l,i)}$ then, by definition, $r' \in W_{(r,m,i)}$. Using parts (a) and (b) of Lemma 37, we get that $r' \in [r]_i$. It easily follows that $[r']_i = [r]_i$, so indeed the construction guarantees that $\mathcal{P}'_{(r^l,i)} = \mathcal{P}'_{(r'^m,i)}$, as desired. Since the family of priors satisfies UNIF, so does $\mathcal{I}'$.

Let $\phi \in \mathcal{L}^{KC}$ and $l \geq m$. As in the proof of Theorem 32, we now proceed by induction on the structure of formulas to show that $(\mathcal{I}', r^l, m) \models \phi$ if and only if $(\mathcal{I}, r, m) \models \phi$. The only difference arises in dealing with the $\rightarrow_i$ modality.

As before, parts (d) and (f) of Lemma 35 imply that $\mathcal{P}'_{(r^l,i)}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))}$ is isomorphic to $\oplus_{k\geq m}(\mathcal{P}^k_{[r]_i}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))})$. Again, we consider the first term in the "sum", $\mathcal{P}^m_{[r]_i}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))}$. We want to show that $\mathcal{P}^m_{[r]_i}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))} = \mathcal{P}^m_{(r,i)}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))}$. Recall that $\mathcal{P}^m_{(r,i)}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))}$ is the first term in the analogous "sum" in the proof of Theorem 32. Thus, even though we are using a different prior from that of the proof of Theorem 32, after conditioning, they are essentially the same. By Lemma 36, we have that $\mathcal{P}^m_{[r]_i}|_{\mathcal{R}^m_{(r,i)}} = \mathcal{P}^m_{(r,i)}$. Thus, it suffices to show that $\cup_{r'\in[r]_i}\mathcal{R}^m_{(r',i)} \cap \mathcal{R}'(\mathcal{K}'_i(r^l,m)) = \mathcal{R}^m_{(r,i)} \cap \mathcal{R}'(\mathcal{K}'_i(r^l,m)$. The inclusion from right to left is immediate. For the opposite inclusion, suppose that $s^m \in \cup_{r'\in[r]_i}\mathcal{R}^m_{(r',i)} \cap \mathcal{R}'(\mathcal{K}'_i(r^l,m))$. Since $s^m \in \mathcal{R}'(\mathcal{K}'_i(r^l,m))$, we must have $(r,m) \sim_i (s,m)$. Since $s^m \in \cup_{r'\in[r]_i}\mathcal{R}^m_{(r',i)}$, there must also be some run $r' \in [r]_i$ such that $s \in \mathcal{R}^m_{(r',i)}$. Since $s \in \mathcal{R}^m(r',i)$, we have that $(s,m) \in W_{(r',m,i)}$. By part (b) of Lemma 37, $(s,m) \in [r',m]_i$. By part (c) of Lemma 37, we get that $(s,0) \in [r',0]_i$. Since $(r',0) \in [r,0]_i$, it immediately follows that $[r',0]_i = [r,0]_i$. Hence, $(s,0) \in [r,0]_i$. Now by part (d) of Lemma 37, we get that $(s,m) \in [r,m]_i$. Thus, $\mathcal{P}_i(s,m) = \mathcal{P}_i(r,m)$. Since $\mathcal{I}$ satisfies UNIF and $(s,m) \in W_{(r',m,i)}$, it follows that $\mathcal{P}_i(s,m) = \mathcal{P}_i(r',m)$. Hence, $(s,m) \in W_{(r,m,i)}$. Finally, we can conclude that $s \in \mathcal{R}^m_{(r,i)}$, as desired. Given this equivalence, we can deal with the $\rightarrow_i$ case just as we did in the proof of Theorem 32.

Finally, we need to ensure that $\mathcal{I}'$ satisfies $\mathcal{A}$. The proof of Theorem 32 shows that if $\mathcal{I}$ satisfies NORM, then so does $\mathcal{I}'$. Using parts (b) and (c) of Lemma 36, it easily follows that if $\mathcal{I}$ satisfies QUAL or RANK, then so does $\mathcal{I}'$.

If REF and UNIF are both in $\mathcal{A}$ (but SDP is not), then we need a further modification of the prior, in the same spirit of that in the proof of Theorem 32. Define $\mathcal{P}^\infty_{[r]_i} = ([r]^\infty_i, \mathrm{Pl}^\infty_{[r]_i})$, where $\mathrm{Pl}^\infty_{[r]_i}(\emptyset) = \bot$ and $\mathrm{Pl}^\infty_{[r]_i}(A) = \top$ for all $A \neq \emptyset$. We now take the prior of the agent to be $\mathcal{P}''_{(r^l,i)} = \mathcal{P}'_{(r^l,i)} \oplus \mathcal{P}^\infty_{[r]_i}$. It is straightforward to show that the resulting system satisfies REF and the requirements of the theorem, using essentially the same arguments for dealing with REF as in the proof of Theorem 32.

Finally, suppose SDP $\in \mathcal{A}$ but REF is not. Note that, since CONS and SDP imply UNIF, $\mathcal{I}$ satisfies UNIF, so we can assume without loss of generality that UNIF is also in $\mathcal{A}$. To get $\mathcal{I}'$ to satisfy SDP, we further modify $\mathcal{P}'$ so that it depends only on the agent, and not the run. Thus, we define $\mathcal{P}^m_i = \otimes\{\mathcal{P}^m_{[r]_i} : r \in \mathcal{R}\}$, and define $\mathcal{P}'_{(r^l,i)} = \oplus_m \mathcal{P}^m_i$. Clearly, with this prior, $\mathcal{I}'$ satisfies SDP. Again, we need to check that this change in prior does not affect the rest of our argument. Once more, the only difficulty comes in dealing with the $\rightarrow_i$ case. Just as in the case of UNIF, we proceed by showing that $\mathcal{P}^m_i|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))} = \mathcal{P}^m_{(r,i)}|_{\mathcal{R}'(\mathcal{K}'_i(r^l,m))}$. The argument is actually even easier than that for UNIF: We show that $\cup_{r'}\mathcal{R}^m_{(r',i)} \cap \mathcal{R}'(\mathcal{K}'_i(r^l,m)) = \mathcal{R}^m_{(r,i)} \cap \mathcal{R}'(\mathcal{K}'_i(r^l,m)$. Again, the inclusion from right to left is immediate. For the opposite inclusion, suppose that $s^m \in \cup_{r'}\mathcal{R}^m_{(r',i)} \cap \mathcal{R}'(\mathcal{K}'_i(r^l,m))$. Since $s^m \in \mathcal{R}'(\mathcal{K}'_i(r^l,m))$, we must have $(r,m) \sim_i (s,m)$. Since $s^m \in \cup_{r'}R^m_{(r',i)}$, there must also be some run r' such that $s \in \mathcal{R}^m_{(r',i)}$. Thus, $(s,m) \in W_{(r',m,i)}$. Since $\mathcal{I}$ satisfies CONS, we have

$(s,m) \sim_i (r',m)$. It follows that $(r',m) \sim_i (r,m)$. Since $\mathcal{I}$ satisfies SDP, we must have that $W_{(r,m,i)} = W_{(r',m,i)}$, so $(s,m) \in W_{(r,m,i)}$. Therefore, $s \in \mathcal{R}^m_{(r,i)}$, as desired.

The modifications to deal with the case where both SDP and REF are in $\mathcal{A}$ are identical to the case with UNIF, and are omitted here. $\square$

References

[1] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet functions for contraction and revision. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] P. Battigalli and G. Bonanno. The logic of belief persistency. *Economics and philosophy*, 1997. To appear.

[3] C. Boutilier. Normative, subjective and autoepistemic defaults: adopting the Ramsey test. In B. Nebel, C. Rich, and W. Swartout, editors, *Proc. Third International Conference on Principles of Knowledge Representation and Reasoning (KR '92)*, pages 685–696. Morgan Kaufmann, San Francisco, 1992.

[4] C. Boutilier. Conditional logics of normality: a modal approach. *Artificial Intelligence*, 68:87–154, 1994.

[5] C. Boutilier. Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence*, 68:33–85, 1994.

[6] R. Brafman, J.-C. Latombe, Y. Moses, and Y. Shoham. Applications of a logic of knowledge to motion planning under uncertainty. *Journal of the ACM*, 1997. To appear.

[7] J. Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22:76–84, 1981.

[8] R. Davis and W. Hamscher. Model-based reasoning: troubleshooting. In H. Shrobe and The American Association for Artificial Intelligence, editors, *Exploring AI*, pages 297–346. Morgan Kaufmann, SF, 1988.

[9] B. De Finetti. *Probability, Induction and Statistics.* John Wiley & Sons, Inc., New York, 1972.

[10] J. de Kleer. Using crude probability estimates to guide diagnosis. *Artificial Intelligence*, 45:381–392, 1990.

[11] P. Diaconis and S. L. Zabell. Updating subjective probability. *Journal of the American Statistical Society*, 77(380):822–830, 1982.

[12] D. Dubois and H. Prade. An introduction to possibilistic and fuzzy logics. In G. Shafer and J. Pearl, editors, *Readings in Uncertain Reasoning*, pages 742–761. Morgan Kaufmann, San Francisco, Calif., 1990.

[13] D. Dubois and H. Prade. Possibilistic logic, preferential models, non-monotonicity and related issues. In *Proc. Twelfth International Joint Conference on Artificial Intelligence (IJCAI '91)*, pages 419–424. Morgan Kaufmann, San Francisco, 1991.

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

[15] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, Mass., 1995.

[16] M. J. Fischer and R. E. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, 18(2):194–211, 1979.

[17] N. Friedman. *Modeling Beliefs in Dynamic Systems*. PhD thesis, Stanford, 1997.

[18] N. Friedman and J. Y. Halpern. Conditional logics of belief change. In *Proc. National Conference on Artificial Intelligence (AAAI '94)*, pages 915–921. AAAI Press, Menlo Park, Calif., 1994.

[19] N. Friedman and J. Y. Halpern. A knowledge-based framework for belief change. Part I: foundations. In R. Fagin, editor, *Theoretical Aspects of Reasoning about Knowledge: Proc. Fifth Conference*, pages 44–64. Morgan Kaufmann, San Francisco, Calif., 1994.

[20] N. Friedman and J. Y. Halpern. Plausibility measures: a user's manual. In P. Besnard and S. Hanks, editors, *Proc. Eleventh Conference on Uncertainty in Artificial Intelligence (UAI '95)*, pages 175–184. Morgan Kaufmann, San Francisco, 1995.

[21] N. Friedman and J. Y. Halpern. On the axiomatization and complexity of conditional logics. In preperation. A preliminary version appeared in J. Doyle, E. Sandewall, and P. Torasso, editors. *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*, Morgan Kaufmann, San Francisco, 1994., 1996.

[22] N. Friedman and J. Y. Halpern. A qualitative Markov assumption and its implications for belief change. In E. Horvitz and F. Jensen editors. *Proc. Twelfth Conference on Uncertainty in Artificial Intelligence (UAI '96)*, pages 263–273, 1996.

[23] N. Friedman and J. Y. Halpern. Modeling belief in dynamic systems. Part II: revision and update. Submitted for publication. A preliminary version appears in J. Doyle, E. Sandewall, and P. Torasso, editors. *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*, 1994, pp. 190–201, under the title "A knowledge-based framework for belief change. Part II: revision and update.", 1997.

[24] N. Friedman and J. Y. Halpern. Plausibility measures and default reasoning. *Journal of the ACM*, 1997. To appear. A preliminary version of this work appeared in *Proc. National Conference on Artificial Intelligence (AAAI '96)*, 1996, pages 1297–1304.

[25] N. Friedman, J. Y. Halpern, and D. Koller. Conditional first-order logic revisited. In *Proc. National Conference on Artificial Intelligence (AAAI '96)*, pages 1305–1312. AAAI Press, Menlo Park, Calif., 1996.

[26] D. Gabbay, A. Pnueli, S. Shelah, and J. Stavi. On the temporal analysis of fairness. In *Proc. 7th ACM Symp. on Principles of Programming Languages*, pages 163–173, 1980.

[27] P. Gärdenfors. *Knowledge in Flux*. MIT Press, Cambridge, Mass., 1988.

[28] P. Gärdenfors and D. Makinson. Revisions of knowledge systems using epistemic entrenchment. In M. Vardi, editor, *Proc. Second Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 83–95. Morgan Kaufmann, San Francisco, 1988.

[29] M. L. Ginsberg. Counterfactuals. *Artificial Intelligence*, 30:35–79, 1986.

[30] M. Goldszmidt, P. Morris, and J. Pearl. A maximum entropy approach to nonmonotonic reasoning. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 15(3):220–232, 1993.

[31] M. Goldszmidt and J. Pearl. Rank-based systems: A simple approach to belief revision, belief update and reasoning about evidence and actions. In B. Nebel, C. Rich, and W. Swartout, editors, *Proc. Third International Conference on Principles of Knowledge Representation and Reasoning (KR '92)*, pages 661–672. Morgan Kaufmann, San Francisco, 1992.

[32] A. Grove. Two modelings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[33] J. Y. Halpern and R. Fagin. Modelling knowledge and action in distributed systems. *Distributed Computing*, 3(4):159–179, 1989. A preliminary version appeared in *Proc. 4th ACM Symposium on Principles of Distributed Computing*, 1985, with the title "A formal model of knowledge, action, and communication in distributed systems: preliminary report".

[34] J. Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[35] J. Y. Halpern and M. R. Tuttle. Knowledge, probability, and adversaries. *Journal of the ACM*, 40(4):917–962, 1993.

[36] J. Y. Halpern and M. Y. Vardi. The complexity of reasoning about knowledge and time, I: lower bounds. *Journal of Computer and System Sciences*, 38(1):195–237, 1989.

[37] J. Hintikka. *Knowledge and Belief*. Cornell University Press, Ithaca, N.Y., 1962.

[38] H. Katsuno and A. Mendelzon. On the difference between updating a knowledge base and revising it. In J. A. Allen, R. Fikes, and E. Sandewall, editors, *Proc. Second International Conference on Principles of Knowledge Representation and Reasoning (KR '91)*, pages 387–394. Morgan Kaufmann, San Francisco, 1991.

[39] H. Katsuno and A. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3):263–294, 1991.

[40] S. Kraus and D. Lehmann. Knowledge, belief, and time. *Theoretical Computer Science*, 58:155–174, 1988.

[41] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[42] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:76–86, 1951.

[43] P. Lamarre and Y. Shoham. Knowledge, certainty, belief, and conditionalisation. In J. Doyle, E. Sandewall, and P. Torasso, editors, *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*, pages 415–424. Morgan Kaufmann, San Francisco, 1994.

[44] H. J. Levesque. A logic of implicit and explicit belief. In *Proc. National Conference on Artificial Intelligence (AAAI '84)*, pages 198–202, 1984.

[45] D. K. Lewis. *Counterfactuals*. Harvard University Press, Cambridge, Mass., 1973.

[46] D. Makinson. On some completeness theorems in modal logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 12:379–384, 1966.

[47] R. van der Meyden. Mutual belief revision (preliminary report). In J. Doyle, E. Sandewall, and P. Torasso, editors, *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*, pages 595–606. Morgan Kaufmann, San Francisco, Calif., 1994.

[48] R. van der Meyden. Personal communication. 1994.

[49] Y. Moses and Y. Shoham. Belief as defeasible knowledge. *Artificial Intelligence*, 64(2):299–322, 1993.

[50] J. Pearl. Probabilistic semantics for nonmonotonic reasoning: a survey. In R. J. Brachman, H. J. Levesque, and R. Reiter, editors, *Proc. First International Conference on Principles of Knowledge Representation and Reasoning (KR '89)*, pages 505–516, 1989. Reprinted in *Readings in Uncertain Reasoning*, G. Shafer and J. Pearl (eds.), Morgan Kaufmann, San Francisco, Calif., 1990, pp. 699–710.

[51] M. Piccione and A. Rubinstein. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, 1997. To appear.

[52] V. R. Pratt. Models of program logics. In *Proc. 20th IEEE Symp. on Foundations of Computer Science*, pages 115–122, 1979.

[53] F. P. Ramsey. Truth and probability. In R. B. Braithwaite, editor, *The Foundations of Probability and other Logical Essays*. Harcourt Brace, New York, 1931.

[54] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987. Reprinted in in *Readings in Nonmonotonic Reasoning*, M. L. Ginsberg (ed.), Morgan Kaufman, San Francisco, CA. 1987, pp. 352–371.

[55] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J., 1976.

[56] Y. Shoham. A semantical approach to nonmonotonic logics. In *Proc. 2nd IEEE Symp. on Logic in Computer Science*, pages 275–279, 1987. Reprinted in M. L. Ginsberg (Ed.), *Readings in Nonmonotonic Reasoning*, Morgan Kaufman, San Francisco, Calif., 1987, pp. 227–250.

[57] W. Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. In W. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change and Statistics*, volume 2, pages 105–134. Reidel, Dordrecht, Netherlands, 1988.

[58] F. Voorbraak. Generalized Kripke models for epistemic logic. In Y. Moses, editor, *Theoretical Aspects of Reasoning about Knowledge: Proc. Fourth Conference*, pages 214–228. Morgan Kaufmann, San Francisco, Calif., 1992.

[59] Z. Wang and G. J. Klir. *Fuzzy Measure Theory*. Plenum Press, New York, 1992.