MODELING BELIEFS IN DYNAMIC SYSTEMS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Nir Friedman
March, 1997

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Joseph Y. Halpern
(Principal Advisor)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Yoav Shoham

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Johan van Benthem
CSLI

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Daphne Koller

Approved for the University Committee on Graduate Studies:

Preface

Agents must reason and act in an uncertain world. There may be uncertainty about the state of the world, uncertainty about the effects of actions, and uncertainty about other agents' actions. Since agents simply cannot address all possible situations that might occur, they must make defeasible assumptions, that is, assumptions that may turn out to be false. In most situations, we do not focus on the agents' defeasible assumptions, but rather on their consequences. We call the logical consequences of defeasible assumptions *beliefs*. We say that an agent believes φ if she acts as though φ is true. As time passes and new evidence is observed, changes in an agent's defeasible assumptions lead to changes in her beliefs. Thus, the question of *belief change*—that is, how beliefs change over time—is a central one for understanding systems that can make and modify defeasible assumptions.

In this dissertation, we propose a new approach to the question of belief change. This approach is based on developing a semantics for beliefs. This semantics is embedded in a framework that models agents' knowledge (or information) as well as their beliefs, and how these change in time. We argue, and demonstrate by examples, that this framework can naturally model any dynamic system (e.g., agents and their environment). Moreover, the framework allows us to consider what the properties of well-behaved belief change should be.

As we show, such a framework can give us a much deeper understanding of how and why beliefs change. In particular, we can gain a better understanding of the current approaches to belief change—belief revision and belief update. Roughly, revision treats a surprising observation (one that is inconsistent with the agent's current beliefs) as a sign that the beliefs are incorrect, while update treats a surprising observation as an indication that the world has changed. We show how belief revision and belief update can be captured in the proposed framework. This allows us to compare the assumptions made by each method and to better understand the principles underlying them.

This analysis shows that revision and update are only two points on a spectrum. In general, we would expect that an agent making an observation may both want to revise some earlier beliefs and assume that some change has occurred in the world. We describe a novel approach to belief change that allows us to do this, by applying ideas from probability theory in qualitative settings. This approach is based on a qualitative analogue of the Markov assumption, which gives us a well-behaved notion of belief change, without making the occasionally unreasonable assumptions made by belief revision and update. In particular, it allows a user to weigh the relative plausibility that a given observation is due to a change in the world or due to an inaccuracy in previous beliefs.

Acknowledgments

I would like to thank the many people who made this period of my life a memorable one.

First and foremost, I would like to thank my advisor Joe Halpern. In the last four years Joe and I worked on quite a few research papers. Joe always treated me as a colleague and a coauthor of equal status. Yet, at the same time, he managed to instill in me the basic skills of research: how to think critically, how to select problems, how to solve them, and how to present the solution. In particular, Joe showed endless reserves of patience in the, seemingly impossible, undertaking of teaching me how to write clearly. From Joe I learned to strive for quality in every aspect of my scientific work. Aside from being my research advisor, Joe was also a friend, a colleague and a teacher. He always found the time to hear out new half-baked ideas, give a word of advice, or provide detailed and thoughtful comments on a paper. Finally, although Joe was funding me from his own grants, he gave me the freedom to choose my research topics and encouraged me to collaborate with other researchers. In short, I would strongly nominate Joe as the ideal Ph.D. advisor. Thanks Joe!

I would like to thank the members of my reading committee, Johan van Benthem, Daphne Koller and Yoav Shoham, and also Nils Nilsson who was a member of my orals committee. All four provided me with interesting and insightful comments.

I would like to thank Yoav Shoham who was my official advisor at Stanford. Even though I was not his student, he was happy to provide logistic support from his funds, and took care of me as one of his own students. Yoav was always enthusiastic about my research, and at the same time provided critical comments that helped shape up the work reported here.

I would like to thank Daphne Koller. When I started working with Joe, Daphne Koller was in the final year of her Ph.D. studies under Joe's supervision. It did not take long before she declared me her "younger brother" (by academic lineage—in real life, we are unrelated and I am a bit older). Since then she spent two years in Berkeley and came back to Stanford as an assistant professor. Throughout this period, she was a good friend and was always there to provide moral support and words of advice, and to discuss new research ideas. On many different occasions, Daphne found the time to provide thoughtful, detailed, and often critical comments on a paper or a talk. These comments always greatly improved the final product, including this dissertation.

I would like to thank Moises Goldszmidt. My interaction with Moises started as a short summer visit to Rockwell Science Center to chat about nonmonotonic reasoning and belief change. These chats led to a particularly fruitful collaboration on learning probabilistic models from data. Moises instilled in me the importance of structure in probabilistic reasoning and the fundamental concepts of

Bayesian networks. He was instrumental in sparking my interest in machine learning. Together we learned about the induction of Bayesian networks from data, and tried to apply the ideas to realistic problems. This lead to an interesting line of research that could have been a second dissertation. Moises has become a good friend. He was always there to provide a good word of advice and to put academic life in perspective.

During these last four years I had the pleasure of collaborating with several coauthors: Craig Boutilier, Ronen Brafman, Moises Goldszmidt, Joe Halpern, Daphne Koller, and Zohar Yakhini. I learned much from all of them. Craig and I became acquainted when we found out that we have similar views of belief change. Since then we had interesting discussions on a wide range of topics and even wrote few papers together. Discussions with Craig helped focus many of the ideas in this dissertation. When I came to Stanford, Ronen was in his second year of studies under Yoav's supervision. We spent many hours talking about the philosophical foundations of AI. These discussions influenced my views of AI, and motivated my interest in the foundations of decision theory and probability. Zohar and I initially met by virtue of having children of the same age. Our collaboration started when I asked his advice on some properties of the entropy function. From Zohar I learned about information theory and its application to learning of probabilistic models.

I would like to thank Zohar Manna who encouraged me to apply to Stanford. A discussion with him convinced me that this an option I should consider, and eventualy lead to my application to Stanford. This is one decision I certainly do not regret!

During my studies I was fortunate enough to interact with many people at Stanford and in at various conferences and workshops. These interactions shaped, in many ways, my understanding of AI. Much of my initial knowledge of AI came from learning for the qualifying examination. Me passing the quals is largely due to many discussions in the QUAIL reading group. I would like to thank QUAIL members: Marko Balabanovic, Scott Benson, Steve Cousins, Ofer Matan, Illah Nourbakhsh, Karl Pflager, Tomas Uribe, and Michelle Wang. I would also like to thank Ronny Kohavi. Ronny has an encyclopedic knowledge of AI and managed, in his very enthosiastic manner, to explain many of the core ideas of the AI and machine learning. I would also like to thank members and visitors at the Robotics laboratory: Hector Geffner, George John, Steven Ketchpel, James Kittock, Pat Langley, Moshe Tenneholtz, Grisha Schwarz, Mehran Sahami, and Alvaro del Val. Finally, I would like to thank the following for useful discussions: Fahiem Bacchus, Wray Buntine, Adnan Darwice, Tom Dean, Matt Ginsberg, Adam Grove, Leslie Pack Kealbling, Daniel Lehmann, Alon Levy, Ron van der Meyden, and Judea Pearl.

During my studies I had the opportunity to work at two research laboratories. I would like to thank Cynthia Dwork, Ron Fagin, Nimrod Meggido, Moni Naor, Eli Upfal, and Moshe Vardi of IBM Almaden Research Center and Denise Draper, Ken Fertig, Dan Larner, Jim Martin, Mark Peot, Sudhaker Reddy, Corinne Ruokangas, David Smith, Phil Subblefield, and Serdar Uckun of Rockwell Science Center, Palo Alto Laboratory for many interesting discussions. I would also like to thank Ron and Moshe for many useful comments on early versions of the papers that later became parts of this dissertation.

I would like to thank the friends who made this period a fun one: Alona and Ronen Brafman, Jennifer and Steve Cousins, Yifat and Ran Gazit, Galia and Haggai Kupermintz, Merav and Alon

Levy, Vered and Danny Porat, Anat and Yossi Rubner, and Tamar and Zohar Yachini,

Last, but not least, I would like to thank my family. I am grateful to my parents, Noga and Gad, who instilled in me the basic interest in science. Their encouragement and support helped me reach the goal of finishing a Ph.D. I would like to thank my children, Roy and Lior. They brought new dimensions to my life and keep reminding me that there is much more to life than my work. Above all I am grateful to my wife, Yael, for her love, support, patience and encouragement. To her I dedicate this dissertation.

This dissertation was supported by Air Force Office of Scientific Research (AFSC), under Contract F49620-91-C-0080, NSF grant IRI-95-03109, IBN Graduate Fellowship and by Rockwell Science Center.

Contents

Chapter 1

Introduction

1.1 Background and Motivation

Agents must reason and act in an uncertain world. There may be uncertainty about the state of the world, uncertainty about the effects of actions, and uncertainty about other agents' actions. In order to act in the world agents 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. Defeasible assumptions appear in most problems that involve dealing with uncertain world. Such assumptions are crucial—we simply cannot address all possible situations that might occur.

One example of the use of defeasible assumptions is *planning.* Most (classical) planners in artificial intelligence (following [Fikes and Nilsson 1972]) assume that the world is *static*; that is, the only changes that are occur in the world are the ones brought about by the agent's actions. These planners also assume that the agent's actions have deterministic effects in the world. While these assumptions are clearly optimistic, they simplify the planner's task, and thus make it computationally feasible. Since these assumptions might fail, planning systems also involve an *execution monitor* that detects contingencies during execution of the plan. The execution monitor uses sensory information to decide when assumptions should be retracted and new ones should be made, and then it decides whether the current plan is still viable or a new one is needed.

Another example, which we analyze formally in later chapters, is *diagnosis*. The diagnosis task is to determine what failures are responsible for the observed symptoms. (Why is the paper copier jammed? Why is the patient's temperature high?) In most situations, there are many combinations of failures that can cause the observed symptoms. However, diagnosis engines usually make the assumption that failures are rare, and thus only few (say one or two) components actually failed. This assumption is crucial, since the diagnosis engine cannot consider all possible failure modes and their interactions. However, if the symptoms cannot be attributed to a simple failure, the diagnosis engine must withdraw its assumption, and consider more complex failures. Other situations where defeasible assumptions are typically made include multi-agent interactions (e.g., we assume that all drivers obey the traffic laws), game playing, learning, vision, and natural language processing.

In most situations, we do not focus on our defeasible assumptions, but rather on their consequences. We call the logical consequences of defeasible assumptions *beliefs*. Intuitively, we would say that an agent believes φ if she acts as though φ is true. As time passes and new evidence is observed, changes in an agent's defeasible assumptions lead to changes in her beliefs. Thus, the question of *belief change*—that is, how beliefs change over time—is a central one for understanding systems that can make and modify defeasible assumptions. Indeed the problem of belief change has received much attention in recent years in philosophy (e.g., [Gärdenfors 1988; Levi 1988; Spohn 1988]), database theory (e.g., [Keller and Winslett 1985]), game theory (e.g., [Stalnaker 1992]), and artificial intelligence (e.g., [Boutilier 1994c; Darwiche and Pearl 1994; Goldszmidt and Pearl 1992; Katsuno and Mendelzon 1991b; Lehmann 1995; del Val and Shoham 1994]).

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 once the agent decides to accept a new belief. This is the "traditional" approach taken in [Alchourrón, Gärdenfors, and Makinson 1985; Gärdenfors 1988; Katsuno and Mendelzon 1991a], which is predominant in the current belief change literature. Indeed, as the papers in the literature show, much can be done with this framework.

However, there are several problems with this traditional approach. First, it focuses only on beliefs. According to this approach, beliefs change when the agent decides, for some reason, to accept a new belief. The framework does not attempt to relate the acceptance of new beliefs with other aspects of the agent, such as her sensory input. Second, the traditional approach does not treat belief change as a temporal process. Thus, while this approach examines the agent's beliefs before and after the change, it does not consider the effect of previous changes on how the agent would change her beliefs in the future. This is the problem of *iterated belief change*. Finally, the traditional approach is *axiomatic*, that is, it characterizes belief change using axioms. Such a characterization of belief change is *non-robust* in the sense that if we consider enriching the language, the axiomatic characterization might change. Thus, the axioms that characterize belief change in a simple setting will not suffice (and usually will be inconsistent) once we consider a richer language. (This issue is discussed in more detail in Chapter 5.)

These problems with the traditional approach limit its applicability to problems such as plan monitoring. Even if we simplify most of the details, we cannot phrase the plan monitoring problem in the setup of the traditional approach, since it involves an agent's actions and sensory input.

In this dissertation, we propose a different approach to the question of belief change. This approach is based on developing a semantics for beliefs. This semantics is embedded in a framework that models agents' knowledge (or information) as well as their beliefs, and how these change in time. We argue, and demonstrate by examples, that this framework can naturally model any dynamic system (e.g., agents and their environment). Moreover, the framework allows us to consider what the properties of well-behaved belief change should be. As we show, such a framework can give us a much deeper understanding of how and why beliefs change. Moreover, this semantics provides the tools for dealing with complicating factors such actions, external events, and multiple agents.

1.2 Overview of the Dissertation

The first half of the dissertation (Chapters 2–4) develops the formal framework of modeling beliefs in dynamic systems. The second half (Chapters 5 and 6) applies this formal framework to the problem of *belief change*.

1.2.1 The Framework

We start with the semantics of beliefs. 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 [Boutilier 1992; Goldszmidt and Pearl 1992; Katsuno and Mendelzon 1991a; Spohn 1988] have then interpreted "the agent believes φ" as "φ is true in the most plausible worlds that the agent considers possible". This notion of beliefs is closely tied with notions of *non-monotonic* or *default* reasoning that describes what conclusions follow "by default" from a set of premises. Roughly speaking, ψ follows by default from φ in a preference ordering if ψ is true in the most plausible worlds among these where φ is true. In this view, an agent's beliefs are just her default conclusions given the premise *true*. This suggests how a semantics for default reasoning can provide a semantics for beliefs.

There have been many semantics for default reasoning proposed in the literature. The recent literature has been guided by a collection of axioms that have been called the *KLM properties* [Kraus, Lehmann, and Magidor 1990], and many of the recent approaches to default reasoning have been shown to be characterized by these properties. This has been viewed as somewhat surprising, since these approaches seem to capture quite different intuitions.

In Chapter 2, we develop the tools to understand why this should be so. In fact, we show much more. We introduce what is perhaps the most general approach to uncertainty possible, that we call a *plausibility measure*. A plausibility measure associates with a set a *plausibility*, which is just an element in a partially ordered space. The only property that we require is that the plausibility of a set be at least as large as the plausibility of any of its subsets. Every systematic approach for dealing with uncertainty that we are aware of can be viewed as a plausibility measure. In particular, plausibility measures subsume all the approaches to default reasoning.

We then examine the minimal conditions on plausibility measures that are required by KLM properties. We define a notion of *qualitative plausibility measure*, and show that the KLM properties are sound in a plausibility structure if and only if it is qualitative. Moreover, as long as a class $\mathcal{P}$ of qualitative plausibility structures satisfies a minimal richness condition, we show that the KLM properties will completely characterize default reasoning in $\mathcal{P}$. We then show that when we map any of the other structures considered in the literature on defaults into plausibility structures, we get a class of qualitative structures that is easily seen to satisfy the richness conditioning. This explains why the KLM axioms characterize default reasoning in all these frameworks. Far from being surprising that the KLM axioms are complete in all these cases, it is almost inevitable.

This investigation suggests that modeling beliefs using qualitative plausibility measures is the most general approach that satisfies the KLM properties. Using plausibility, we say that an agent believes φ if φ is more plausible than its negation, $\neg\varphi$. By modeling beliefs in this way, there is an assumption that the plausibility ordering is part of the agent's epistemic state. (This assumption is actually made explicitly in [Boutilier 1992; Kraus, Lehmann, and Magidor 1990].) This implies that the plausibility ordering is *subjective*, that is, it describes the agent's estimate of what the plausible events are. But actually, an even stronger assumption is being made: namely, that the agent's epistemic state is characterized by a *single* plausibility ordering. 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 orderings possible; in some it typically does not rain and in others it typically does.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 orderings, 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 orderings she considers possible, φ is true in the most plausible worlds. Notice that in the special case where the agent considers only one ordering possible, this definition of belief is equivalent to of belief as truth in the most plausible worlds.

The properties of belief depend on how the plausibility ordering interacts with the accessibility relation that defines knowledge. In Chapter 3, 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 [1994] 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 [1988]. 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 in Chapter 4. The resulting framework is an extension of the framework of Halpern and Fagin [1989] for modeling knowledge in multi-agent systems.

Halpern and Fagin's framework models how a *system* changes over time. We start by assuming that at each point in time, each component in the system (e.g., the agents and the external environment) is in some *state*. A *run* is a function from time (which we assume ranges over the natural numbers) to global states that describe the state of each component; thus, a run is essentially a sequence of global states. A run can be thought of as a description of how the system changes over time. We describe a system by the set of possible runs. Intuitively, we are identifying the system with its possible behaviors. To this framework, we add a description of each agent's plausibility in each point in time. The resulting framework allows us to talk about knowledge, plausibility (and hence belief), and time. This framework is analogous to combination of knowledge, probability and time studied in [Halpern and Tuttle 1993]. 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 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 her plausibility assessment (and her beliefs) at time $t+1$ to incorporate this new information; otherwise, we would expect her assessment at time $t+1$ to have changed minimally from her 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, by starting with a *prior plausibility* and conditioning. As we show, many situations previously studied in the literature, such as *diagnostic reasoning* [Reiter 1987], can be easily captured by using such prior plausibilities.

To summarize, we construct a general framework that models plausibility and knowledge and how they change over time. We propose one particular mechanism of conditioning that captures the intuition that agents' plausibility assessment should undergo the minimal changes necessary to incorporate new knowledge.

1.2.2 Belief Change

The second half of this dissertation develops one application of this framework: the analysis of *belief change*.

The study of belief change has been an active area in philosophy and artificial intelligence. The focus of this research is to understand how an agent should change her beliefs as a result of getting new information. Two instances of this general phenomenon have been studied in detail. *Belief revision* [Alchourrón, Gärdenfors, and Makinson 1985; Gärdenfors 1988] focuses on how the agent revises her beliefs when she adopts a new belief. *Belief update* [Katsuno and Mendelzon 1991a], on the other hand, focuses on how the agent should change her beliefs when she realizes that the world has changed. Both approaches attempt to capture the intuition that the agent should make minimal changes in order to accommodate the new belief. The difference is that belief revision attempts to decide what beliefs should be discarded to accommodate a new belief, while belief update attempts to decide what changes in the world led to the new observation.

We start by showing how belief revision and update can be represented in our framework. Doing this allows us to compare the assumptions implicit in each method and to understand the principles underlying them.

We are certainly not the first to provide semantic models for belief revision and update. For example, [Alchourrón, Gärdenfors, and Makinson 1985; Grove 1988; Gärdenfors and Makinson

1988; Rott 1991; Boutilier 1992; de Rijke 1992] deal with revision, and [Katsuno and Mendelzon 1991a; del Val and Shoham 1992] deal with update. In fact, there are several works in the literature that capture both using the same machinery [Katsuno and Satoh 1991; Goldszmidt and Pearl 1992], and others that simulate belief revision using belief update [Grahne, Mendelzon, and Rieter 1992; del Val and Shoham 1994]. Our approach is different from most in that we do not construct a specific framework to capture one or both of these belief change paradigms. Instead, we start from a natural framework to model how an agent's knowledge changes over time and add to it machinery that captures a defeasible notion of belief. As we shall see, our framework allows us to clearly bring out the similarities and differences between update and revision.

We start by identifying a class of systems, which we call *belief change systems*, that embody the basic assumptions both AGM revision and KM update make about the setting of belief change. We then examine which belief change systems correspond to AGM revision and which correspond to KM update. We shall see that both can be captured using the same mechanism for changing plausibility measures, namely conditioning. However, while both revision and update are based on the same method of minimal change, they make additional assumptions about the external environment and the form of the agent's plausibility. These additional assumptions are the main source of the difference between revision and update. We can get a better understanding of these assumptions when we examine in more detail what restrictions we have to impose on belief change systems to capture each notion. These restrictions address different aspects of our framework. As we shall see, to capture AGM revision, we need to put strong restrictions on changes of the environment (e.g., environment does not change state) and mild restrictions on the agent's initial plausibility measure. On the other hand, to capture KM update, we need to put only mild restrictions on changes of the environment and strong restrictions on the agent's initial plausibility measure. In particular, the agent's initial plausibility measure has to prefer runs where abnormal events occur as late as possible. This means that the agent would consider a run where there is a slight abnormality at an early stage to be less plausibile than a run where an abnormal event occurs at a later time, even if that latter event is much more "surprising".

As our analysis suggests, revision and update are only two points on a spectrum of belief change operations we might consider. In particular, we claim that the restrictions imposed by KM update on the agent's prior plausibility measures are unreasonable in many situations. Moreover, in general, we would not want to use "pure" revision or update. That is, when the agent makes an observation, she may need both to revise some of her earlier beliefs and to assume that some change has occurred in the world. To see this, consider the following example (which is a variant of Kautz's *stolen car* example [1986]): A car is parked with a full fuel tank at time 0; at time 2, the owner returns to find it parked where she left it. If the owner believes that parked cars tend to stay put, then she would believe that no changes occurred between time 0 and 2. What should she believe when, at time 3, she discovers that the fuel tank is empty? Update treats this observation as an indication of a change between time 2 and 3, for example, a gas leak. Revision, on the other hand, treats it as an indication that previous beliefs, such as the belief that the tank was full at time 2, were wrong. In practice, the agent may want to consider a number of possible explanations for her time-3 observation, depending on what she considers to be the most likely sequence(s) of events between time 0 and time 3. For example, if she has had previous gas leaks, then she may consider a gas leak to be the most plausible explanation. On the other hand, if her spouse has the car keys, she may consider it possible that he used the car in her absence. Is there a reasonable approach that lets us capture these considerations in a natural manner? In Chapter 6, we show that there is and, indeed, we can get one by applying straightforward ideas from probability theory in a qualitative setting.

The approach we propose starts with the agent's prior plausibility measure on runs. As we have described above, to get minimal change of beliefs, we must condition this prior on new knowledge. Thus, the agent's beliefs are determined by her initial prior. One obvious problem with this approach is that, even if there are only two possible states, there are uncountably many possible runs. How can an agent describe a prior plausibility over such a complex space?

A similar problem occurs in probability theory, when we want to describe a probability measure over all possible runs (such a measure often called a *stochastic process*). One standard solution is to assume that state transitions are independent of when they occur, that is, that the probability of the system going from state s to state s' is independent of the sequence of transitions that brought the system to state s. This *Markov assumption* significantly reduces the complexity of the problem. All that is necessary is to describe the probability of state transitions. Moreover, the Markov assumption has been shown to be widely applicable in practice [Kemeny and Snell 1960; Howard 1971].

As we show in Chapter 6, we can use (a qualitative analogue of) the Markov assumption to simplify the description of the prior plausibility. Making a (qualitative) Markov assumption gives us a well-behaved notion of belief change, without making the occasionally unreasonable assumptions made by belief revision and update. In particular, it allows a user to weigh the relative plausibility that a given observation is due to a change in the world or due to an inaccuracy in previous beliefs. In the car example, this means that the agent can decide the relative plausibility of a gas leak and her spouse's taking the car, without making the commitment to one or the other, as required by update and revision.

1.3 Reader's Guide

The rest of the dissertation is organized as follows.

In the first three chapters we formally develop the framework. In Chapter 2 we develop the notion of qualitative plausibility measure and relate it to several well-known approaches to default reasoning and conditional logics. (Portions of this chapter appeared in [Friedman and Halpern 1996c].) In Chapter 3 we use qualitative plausibility measure to model beliefs. We describe a logic for reasoning about knowledge and plausibility, and develop a semantic model that captures both notions. We then examine several semantic variants and their formal properties. In Chapter 4 we introduce time explicitly into the framework. We then develop *conditioning* as a form of *minimal change* in plausibility and analyze it formally. (Portions of Chapters 3 and 4 appeared in [Friedman and Halpern 1997a].)

In next two chapters we examine belief change in the formal framework developed in the preceding chapters. We start, in Chapter 5, by describing a particular class of systems, which we call *belief change systems*, that capture the basic assumptions made in the belief change literature. We then show how AGM revision and KM update can captured in this class of systems, and discuss insight we gain about both notions. (Portions of this chapter appeared in [Friedman and Halpern 1997b].) In Chapter 6 we develop a new approach to belief change based on the Markov assumption. We introduce the necessary machinery to formally define a qualitative Markov assumption and examine the application of this assumption for belief change. (Portions of this chapter appeared in [Friedman and Halpern 1996d].)

Finally, in Chapter 7 we discuss possible extensions of the framework and their implications.

Chapter 2

Representing Uncertainty

2.1 Introduction

As stated in Chapter 1, we are interested in examining how agents deal with uncertainty. The standard approach to modeling uncertainty is probability theory. In recent years, researchers, motivated by varying concerns, including a dissatisfaction with some of the axioms of probability and a desire to represent information more qualitatively, have introduced a number of generalizations and alternatives to probability, such as Dempster-Shafer belief functions [Shafer 1976] and possibility theory [Dubois and Prade 1990]. Our aim here is to introduce what is perhaps the most general approach possible, that we call a *plausibility measure*. A plausibility measure associates with a set a *plausibility*, which is just an element in a partially ordered space. The only property that we require is that the plausibility of a set be at least as large as the plausibility of any of its subsets. Every systematic approach for dealing with uncertainty that we are aware of can be viewed as a plausibility measure. Given how little structure we have required of a plausibility measure, this is perhaps not surprising. Nevertheless, as we hope to demonstrate in this chapter, plausibility measures provide us with a natural setting in which to examine various approaches to reasoning about uncertainty.

The focus of this chapter is (propositional) default reasoning. There have been many approaches to default reasoning proposed in the literature (see [Ginsberg 1987; Gabbay, Hogger, and Robinson 1993] for overviews). The recent literature has been guided by a collection of axioms that have been called the *KLM properties* (since they were discussed in [Kraus, Lehmann, and Magidor 1990]), and many of the recent approaches to default reasoning, including *preferential structures* [Kraus, Lehmann, and Magidor 1990; Shoham 1987], ϵ*-semantics* [Adams 1975; Geffner 1992b; Pearl 1989], *possibilistic structures* [Dubois and Prade 1991], and κ*-rankings* [Goldszmidt and Pearl 1992; Spohn 1988], have been shown to be characterized by these properties. This has been viewed as somewhat surprising, since these approaches seem to capture quite different intuitions. As Pearl [1989] said of the equivalence between ϵ-semantics and preferential structures, "It is remarkable that two totally different interpretations of defaults yield identical sets of conclusions and identical sets of reasoning machinery." As we shall show in this chapter, plausibility measures help us understand why this should be so.

In fact, we show much more. All of these approaches can be understood as giving semantics to defaults by considering a class $\mathcal{S}$ of structures (preferential structures, possibilistic structures, etc.). A default d is then said to follow from a knowledge base Δ of defaults if all structures in $\mathcal{S}$ that satisfy Δ also satisfy d. We define a notion of *qualitative plausibility measure*, and show that the KLM properties are sound in a plausibility structure if and only if it is qualitative. Moreover, as long as a class $\mathcal{S}$ of plausibility structures satisfies a minimal richness condition, we show that the KLM properties completely characterize default reasoning in $\mathcal{S}$. We then show that when we map preferential structures (or possibilistic structures or any of the other structures considered in the literature on defaults) into plausibility structures, we get a class of qualitative structures that is easily seen to satisfy the richness conditioning. This explains why the KLM axioms characterize default reasoning in all these frameworks. Far from being surprising that the KLM axioms are complete in all these cases, it is almost inevitable.

We then examine *conditional logic*. Conditional logic treats the default entailment operator as a modality. Thus, we can write statements that involve negation and disjunction of defaults, as well as nesting of defaults. Plausibility also gives us insight into conditional logic. In particular, we present soundness and completeness proofs for conditional logic with respect to plausibility structures that are substantially simpler than previous ones in the literature. These results allow us to provide soundness and completeness for various proposed semantics of conditional logic.

2.2 Plausibility Measures

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

A plausibility space is a direct generalization of a probability space. We simply replace the probability measure μ by a *plausibility measure* Pl that, rather than mapping sets in $\mathcal{F}$ to numbers in $[0, 1]$, maps them to elements in some arbitrary partially ordered set. We read Pl(A) as "the plausibility of set A". If Pl$(A) \leq$ Pl(B), then B is at least as plausible as A. Formally, a *plausibility space* is a tuple $S = (W, \mathcal{F}, \text{Pl})$, where W is a set of worlds, $\mathcal{F}$ is an algebra of subsets of W, and Pl maps the sets in $\mathcal{F}$ to some set D, partially ordered by a relation $\leq_D$ (so that $\leq_D$ 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 $\bot_D \leq_D d \leq_D \top_D$ for all $d \in D$; we further assume that Pl$(W) = \top_D$ and Pl$(\emptyset) = \bot_D$. The only other assumption we make is

A1. If $A \subseteq B$, then Pl$(A) \leq_D$ Pl(B).

Thus, a set must be at least as plausible as any of its subsets.

[1] Frequently it is also assumed that μ satisfies *countable additivity*, i.e., if A_i, $i > 0$, are pairwise disjoint, then $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$.

Some brief remarks on the definition: We have deliberately suppressed the domain D from the tuple S, since the choice of D is not significant in this chapter. All that matters is the ordering induced by $\leq_D$ on the subsets in $\mathcal{F}$. In dealing with *conditional* plausibility, the domain D plays a more significant role (see Chapter 6). As usual, we define the ordering $<_D$ by taking $d_1 <_D d_2$ if $d_1 \leq_D d_2$ and $d_1 \neq d_2$. We omit the subscript D from $\leq_D$, $<_D$, $\top_D$ and $\bot_D$ whenever it is clear from context.

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

- A *belief function Bel* on W is a function *Bel* : $2^W \rightarrow [0, 1]$ satisfying certain axioms [Shafer 1976]. These axioms certainly imply property A1, so a belief function is a plausibility measure.

- A *fuzzy measure* (or a *Sugeno measure*) f on W [Wang and Klir 1992] is a function f : $2^W \mapsto [0, 1]$, that satisfies A1 and some continuity constraints. A *possibility measure* [Dubois and Prade 1990] Poss is a fuzzy measure such that Poss$(W) = 1$, Poss$(\emptyset) = 0$, and Poss$(A) = \sup_{w\in A}($Poss$(\{w\})$.

- An *ordinal ranking* (or κ*-ranking*) κ on W (as defined by [Goldszmidt and Pearl 1992], based on ideas that go back to [Spohn 1988]) is a function mapping subsets of W to $\mathbb{N}^* = \mathbb{N} \cup \{\infty\}$ 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 $x \leq_{\mathbb{N}^*} y$ if and only if $y \leq x$ under the usual ordering on the ordinals.

- A *preference ordering* on W is a strict partial order $\prec$ over W [Kraus, Lehmann, and Magidor 1990; Shoham 1987]. Intuitively, $w \prec w'$ holds if w is *preferred* to w'.[2] Preference orders have been used to provide semantics for *default* (i.e., conditional) statements. In Section 2.4 we show how to map preference orders on W to plausibility measures on W in a way that preserves the ordering on worlds as well as the truth values of defaults.

- A *parameterized 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* [Pearl 1989; Goldszmidt, Morris, and Pearl 1993]. In Section 2.4 we show how to map PPDs on W to plausibility measures on W in a way that preserves the truth-values of conditionals.

Plausibility structures are motivated by much the same concerns as two other recent symbolic generalizations of probability by Darwiche [1992] and Weydert [1994b]. Their approaches have a great deal more structure though. They start with a domain D and several algebraic operations that have properties similar to the usual arithmetic operations (e.g., addition and multiplication) over $[0, 1]$. The result is an algebraic structure over the domain D that satisfies various properties. Their

[2] We follow the standard notation for preference here [Lewis 1973; Kraus, Lehmann, and Magidor 1990], which uses the (perhaps confusing) convention of placing the more likely world on the left of the $\prec$ operator.

structures are also general enough to capture all of the examples above except preferential orderings. These orderings cannot be captured precisely because of the additional structure. Moreover, as we shall see, by starting with very little structure and adding just what we need, we can sometimes bring to light issues that may be obscured in richer frameworks. We refer the interested reader to [Friedman and Halpern 1995] for a more detailed comparison to [Darwiche 1992; Weydert 1994b].

Given the simplicity and generality of plausibility measures, we were not surprised to discover that Weber [1991] recently defined a notion of *uncertainty measures*, which is a slight generalization of plausibility measure (in that domains more general than algebras of sets are allowed), and that Greco [1987] defined a notion of L-fuzzy measures which is somewhat more restricted than plausibility measures in that the range D is a complete lattice. We expect that others have used variants of this notion as well, although we have not found any further references in the literature. To the best of our knowledge, we are the first to carry out a systematic investigation of the connection between plausibility measures and default reasoning.

2.3 Approaches to Default Reasoning: A Review

Defaults are statements of the form "if φ then typically/likely/by default ψ", which we denote $\varphi{\rightarrow}\psi$. For example, the default "birds typically fly" is represented *Bird*$\rightarrow$*Fly*. Formally, we assume that there is a "base" language $\mathcal{L}$, defined over a set of propositions Φ, that includes the usual propositional connectives, $\wedge, \vee, \neg, \Rightarrow$ and has a consequence relation $\vdash_{\mathcal{L}}$. The language of defaults $\mathcal{L}_{\mathrm{def}}$ contains statements of the form $\varphi{\rightarrow}\psi$, where $\varphi, \psi \in \mathcal{L}$.

There has been a great deal of discussion in the literature as to what the appropriate semantics of defaults should be, and what new defaults should be entailed by a knowledge base of defaults. For the most part, we do not get into these issues here. While there has been little consensus on what the "right" semantics 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, was suggested by [Kraus, Lehmann, and Magidor 1990] and consists of the following axiom and rules of inference.

LLE. If $\vdash_{\mathcal{L}} \varphi \Leftrightarrow \varphi'$, then from $\varphi{\rightarrow}\psi$ infer $\varphi'{\rightarrow}\psi$ (left logical equivalence)

RW. If $\vdash_{\mathcal{L}} \psi \Rightarrow \psi'$, then from $\varphi{\rightarrow}\psi$ infer $\varphi{\rightarrow}\psi'$ (right weakening)

REF. $\varphi{\rightarrow}\varphi$ (reflexivity)

AND. From $\varphi{\rightarrow}\psi_1$ and $\varphi{\rightarrow}\psi_2$ infer $\varphi{\rightarrow}\psi_1 \wedge \psi_2$

OR. From $\varphi_1{\rightarrow}\psi$ and $\varphi_2{\rightarrow}\psi$ infer $\varphi_1 \vee \varphi_2{\rightarrow}\psi$

CM. From $\varphi{\rightarrow}\psi_1$ and $\varphi{\rightarrow}\psi_2$ infer $\varphi \wedge \psi_1{\rightarrow}\psi_2$ (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 $\varphi_2{\rightarrow}\psi$ from $\varphi_1{\rightarrow}\psi$. RW describes a similar property of the consequent: If ψ (logically) entails ψ', then we can deduce $\varphi{\rightarrow}\psi'$ from $\varphi{\rightarrow}\psi$. 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.

This system of rules is called system **P** in [Kraus, Lehmann, and Magidor 1990]. The notation $\Delta \vdash_{\mathbf{P}} \varphi{\rightarrow}\psi$ denotes that $\varphi{\rightarrow}\psi$ can be deduced from Δ using these inference rules.

There are a number of well-known semantics for defaults that are characterized by these rules. We sketch a few of them here, referring the reader to the original references for further details and motivation. All of these semantics involve structures of the form (W, X, π), where W is a set of possible worlds, $\pi(w)$ is a truth assignment consistent with $\vdash_{\mathcal{L}}$ to formulas in $\mathcal{L}$, and X is some "measure" on W such as a preference ordering, a κ-ranking, or a possibility measure. We define a little notation that will simplify the discussion below. Given a structure $M = (W, X, \pi)$ and a formula $\varphi \in \mathcal{L}$, we take $[\![\varphi]\!]_M \subseteq W$ to be the set of of worlds satisfying φ, i.e., $[\![\varphi]\!]_M = \{w \in W : \pi(w)(\varphi) = \textbf{true}\}$. We omit the subscript M when it plays no role or is clear from the context.

The first semantic proposal was provided by Kraus, Lehmann and Magidor [1990], using ideas that go back to [Lewis 1973; Shoham 1987]. Recall that a preference ordering on W is strict partial order (i.e., irreflexive and transitive relation) $\prec$ over W. A *preferential structure* is a tuple $(W, \prec, \pi)$, where $\prec$ is a strict partial order on W.[3] The intuition [Shoham 1987] is that a preferential structure satisfies a conditional $\varphi{\rightarrow}\psi$ if all the the most preferred worlds (i.e., the minimal worlds according to $\prec$) in $[\![\varphi]\!]$ satisfy ψ. However, there may be no minimal worlds in $[\![\varphi]\!]$. This can happen if $[\![\varphi]\!]$ 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 $[\![\varphi]\!]$ whenever $[\![\varphi]\!]$ is not empty; this is the assumption actually made in [Kraus, Lehmann, and Magidor 1990], where it is called the *smoothness* assumption. A yet more general definition—one that works even if $\prec$ is not smooth—is given in [Lewis 1973; Boutilier 1994a]. Roughly speaking, $\varphi{\rightarrow}\psi$ is true if, from a certain point on, whenever φ is true, so is ψ. More formally,

> $(W, \prec, \pi)$ satisfies $\varphi{\rightarrow}\psi$ if, for every world $w_1 \in [\![\varphi]\!]$, there is a world w_2 such that (a) $w_2 \preceq w_1$ (i.e., either $w_2 \prec w_1$ or $w_2 = w_1$) (b) $w_2 \in [\![\varphi \wedge \psi]\!]$, and (c) for all worlds $w_3 \prec w_2$, we have $w_3 \in [\![\varphi \Rightarrow \psi]\!]$ (so any world more preferred 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 knowledge-base Δ *preferentially entails* $\varphi{\rightarrow}\psi$, denoted $\Delta \models_{\mathrm{p}} \varphi{\rightarrow}\psi$, if every preferential structure that satisfies (all the defaults in) Δ also satisfies $\varphi{\rightarrow}\psi$.

Lehmann and Magidor show that preferential entailment is characterized by system **P**.

[3] We note that the formal definition of preferential structures in [Kraus, Lehmann, and Magidor 1990; Lehmann and Magidor 1992] is slightly more complex. Kraus, Lehmann, and Magidor distinguish between *states* and *worlds*. In their definition, a preferential structure is an ordering over states together with a labeling function that maps states to worlds. They take worlds to be truth assignments to primitive propositions. Our worlds thus correspond to states in their terminology, since we allow two worlds $w \neq w'$ such that $\pi(w) = \pi(w')$. Despite these minor differences, all the results that we prove for our version of preferential structures hold (with almost no change in proof) for theirs.

Theorem 2.3.1: [Lehmann and Magidor 1992; Boutilier 1994a] $\Delta \models_{\mathrm{p}} \varphi \rightarrow \psi$ *if and only if* $\Delta \vdash_{\mathbf{P}} \varphi \rightarrow \psi$.

Thus, reasoning with preferential structures corresponds in a precise sense to reasoning with the core properties of default reasoning.

As we mentioned earlier, we usually want to add additional inferences to those sanctioned by the core. Lehmann and Magidor [1992] hoped to do so by limiting attention to a special class of preferential structures. A preferential structure $(W, \prec, \pi)$ is *rational* if $\prec$ is a *modular* order, so that for all worlds $u, v, w \in W$, if $w \prec v$, then either $u \prec v$ or $w \prec u$. It is not hard to show that modularity implies that possible worlds are clustered into equivalence classes, each class consisting of worlds that are incomparable to one another, with these classes being totally ordered. Thus, rational structures form a "well-behaved" subset of preferential structures. Unfortunately, Lehmann and Magidor showed that restricting to rational structures gives no additional properties (at least, as far as the limited language of defaults is concerned). We say that a knowledge base Δ *rationally entails* $\varphi \rightarrow \psi$, denoted $\Delta \models_{\mathrm{r}} \varphi \rightarrow \psi$, if every rational structure that satisfies Δ also satisfies $\varphi \rightarrow \psi$.[4]

Theorem 2.3.2: [Lehmann and Magidor 1992] $\Delta \models_{\mathrm{r}} \varphi \rightarrow \psi$ *if and only if* $\Delta \vdash_{\mathbf{P}} \varphi \rightarrow \psi$.

Thus, we do not gain any new patterns of default inference when we restrict our attention to rational structures.

Pearl [1989] considers a semantics for defaults grounded in probability, using an approach due to Adams [1975]. In this approach, a default $\varphi \rightarrow \psi$ denotes that $\Pr(\psi \mid \varphi)$ is extremely high, i.e., almost 1. Roughly speaking, a collection Δ of defaults implies a default $\varphi \rightarrow \psi$ if we can ensure that $\Pr(\varphi|\psi)$ is arbitrarily close to 1, by taking the probabilities of the defaults in Δ to be sufficiently high.

The formal definition needs a bit of machinery.[5] Recall that a PPD on W is a sequence $\{\Pr_i : i \geq 0\}$ of probability measures over W. A *PPD structure* is a tuple $(W, \{\Pr_i : i \geq 0\}, \pi)$, where $\{\Pr_i\}$ is PPD on W. Intuitively, it satisfies a conditional $\varphi \rightarrow \psi$ if the conditional probability ψ given φ goes to 1 in the limit. Formally, $\varphi \rightarrow \psi$ is satisfied if $\lim_{i \rightarrow \infty} \Pr_i([\![\psi]\!] | [\![\psi]\!]) = 1$ (where $\Pr_i([\![\psi]\!] | [\![\varphi]\!])$ is taken to be 1 if $\Pr_i([\![\varphi]\!]) = 0$). Δ *ϵ-entails* $\varphi \rightarrow \psi$, denoted $\Delta \models_{\epsilon} \varphi \rightarrow \psi$, if every PPD structure that satisfies all the defaults in Δ also satisfies $\varphi \rightarrow \psi$. Surprisingly, Geffner shows that ϵ-entailment is equivalent to preferential entailment.

Theorem 2.3.3: [Geffner 1992b] $\Delta \models_{\epsilon} \varphi \rightarrow \psi$ *if and only if* $\Delta \vdash_{\mathbf{P}} \varphi \rightarrow \varphi$.[6]

Possibility measures and ordinal rankings provide two more semantics for defaults. A *possibility structure* is a tuple $PS = (W, \mathrm{Poss}, \pi)$ such that Poss is a possibility measure on W. We say

[4] Rational entailment should not be confused with the notion of *rational closure*, also defined by Lehmann and Magidor [1992].

[5] We adopt the presentation used in [Goldszmidt, Morris, and Pearl 1993].

[6] Geffner's result is stated in terms of the original formulation of ϵ-entailment, as described in [Pearl 1989]. However, results of [Goldszmidt, Morris, and Pearl 1993] show that the formulation we describe here is equivalent to the original one.

PS $\models_{Poss}$ $\varphi \rightarrow \psi$ if either Poss($[\![\varphi]\!]$) $= 0$ or Poss($[\![\varphi \wedge \psi]\!]$) $>$ Poss($[\![\varphi \wedge \neg\psi]\!]$). Intuitively, $\varphi \rightarrow \psi$ holds vacuously if φ is impossible; otherwise, it holds if $\varphi \wedge \psi$ is more "possible" than $\varphi \wedge \neg\psi$. For example, *Bird*$\rightarrow$*Fly* is satisfied when *Bird* $\wedge$ *Fly* is more possible than *Bird* $\wedge$ $\neg$*Fly*. Similarly, an *ordinal ranking structure* is a tuple $R = (W, \kappa, \pi)$ if κ is a an ordinal ranking on W. We say that $R \models_\kappa \varphi \rightarrow \psi$ if either $\kappa([\![\varphi]\!]) = \infty$ or $\kappa([\![\varphi \wedge \psi]\!]) < \kappa([\![\varphi \wedge \neg\psi]\!])$. We say that Δ *possibilistically entails* $\varphi \rightarrow \psi$, denoted $\Delta \models_{Poss} \varphi \rightarrow \psi$ (resp., Δ κ*-entails* $\varphi \rightarrow \psi$, denoted $\Delta \models_\kappa \varphi \rightarrow \psi$) if all possibility structures (resp., all ordinal ranking structures) that satisfy Δ also satisfy $\varphi \rightarrow \psi$.

These two approaches are again characterized by the KLM properties.

Theorem 2.3.4: [Geffner 1992b; Dubois and Prade 1991] $\Delta \models_{Poss} \varphi \rightarrow \psi$ *if and only if* $\Delta \models_\kappa \varphi \rightarrow \psi$ *if and only if* $\Delta \vdash_{\mathbf{P}} \varphi \rightarrow \psi$.

Why do we always seem to end up with the KLM properties? As we are about to show, thinking in terms of plausibility measures provides the key to understanding this issue.

2.4 Default Reasoning Using Plausibility

We can give semantics to defaults using plausibility measures much as we did using possibility measures. A *plausibility structure* is a tuple *PL* $= (W, \mathcal{F}, \mathrm{Pl}, \pi)$, where $(W, \mathcal{F}, \mathrm{Pl})$ is a plausibility space and π maps each possible world to a truth assignment. Since we will be interested in events that correspond to formulas, we require that $[\![\varphi]\!] \in \mathcal{F}$ for all formula $\varphi \in \mathcal{L}$. For ease of exposition, we omit the $\mathcal{F}$ from here on, always taking it to be $\mathcal{F} = \{[\![\varphi]\!] : \varphi \in \mathcal{L}\}$ when we are considering a plausibility structure, and just denote a plausibility space as a pair (W, Pl), and a plausibility structure as a tuple (W, Pl, π). We define *PL* $\models_{PL} \varphi \rightarrow \psi$ if either Pl($[\![\varphi]\!]$) $=\perp$ or Pl($[\![\varphi \wedge \psi]\!]$) $>$ Pl($[\![\varphi \wedge \neg\psi]\!]$).

Notice that if Pl is a probability function Pr, then $\varphi \rightarrow \psi$ holds exactly if either $\Pr([\![\varphi]\!]) = 0$ or $\Pr([\![\psi]\!] | [\![\varphi]\!]) > 1/2$. How does this semantics for defaults compare to others that have been given in the literature? It is immediate from the definitions that the semantics we give to defaults in possibility structures is the same as that given to them if we view these possibility structures as plausibility structures (using the obvious mapping described in Section 2.3), and similarly for ordinal ranking structures. What about preferential structures and PPD structures? Can we map them into plausibility structures while still preserving the semantics of defaults? As we now show, we can.

In fact, Lemma 2.4.1 shows that there is a general procedure for mapping any approach that satisfies the KLM postulates to plausibility measures. Before describing this general construction, we briefly sketch its instantiation in the case of PPDs and preferential structures.

We start with PPDs. Let $PP = (W, \{\Pr_i\}, \pi)$ be a PPD structure. Let Pl_{PP} be a plausibility measure on W such that

$$\mathrm{Pl}_{PP}(A) \leq \mathrm{Pl}_{PP}(B) \text{ if and only if } \lim_{i \to \infty} \Pr_i(B | A \cup B) = 1. \tag{2.1}$$

It is easy to check that such a plausibility measure exists and that $(W, \mathrm{Pl}_{PP}, \pi)$ satisfies the same defaults as PP. We note that this construction, as well as others in the remainder of the chapter, specifies only the relative order of plausibilities of events, and does not describe the domain of plausibility values. It is easy to check that as long as the ordering constraints are consistent with reflexivity, transitivity, and A1, we can always construct a matching plausibility domain. From here on, we treat such ordering constraints as though they define a plausibility measure.[7]

We stress that this embedding, which is sufficient for the purpose of this work, is not the only one possible. To see this, suppose that A and B are disjoint sets such that $\mathrm{Pr}_i(A) = \mathrm{Pr}_i(B)$ for all i. One might argue that the plausibility of A and B should be equal. Yet our definition would make $\mathrm{Pl}(A)$ and $\mathrm{Pl}(B)$ incomparable since $\mathrm{Pr}_i(B|A \cup B) = 0.5$ for all i.

The construction for mapping preferential structures into plausibility structures is slightly more complex. Suppose we are given a preferential structure $(W, \prec, \pi)$. Let D_0 be the domain of plausibility values consisting of one element d_w for every element $w \in W$. We use $\prec$ to determine the order of these elements: $d_v < d_w$ 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 closed under least upper bounds (so that every set of elements in D has a least upper bound in D). It is not hard to show that D is well-defined (i.e., there is a unique, up to renaming, smallest set) and that taking $\mathrm{Pl}_\prec(A)$ to be the least upper bound of $\{d_w : w \in A\}$ gives us the following property:

$\mathrm{Pl}_\prec(A) \le \mathrm{Pl}_\prec(B)$ if and only if for all $w \in A - B$, there is a world $w' \in B$ such that $w' \prec w$ and there is no $w'' \in A - B$ such that $w'' \prec w'$. (2.2)

Again, it is easy to check that $(W, \mathrm{Pl}_\prec, \pi)$ satisfies the same defaults as $(W, \prec, \pi)$.

We now present our general construction.

Lemma 2.4.1: *Let W be set of possible worlds and let π be a function that maps each world in W to a truth assignment to $\mathcal{L}$. Let $T \subseteq \mathcal{L}_{\mathrm{def}}$ be a set of defaults that is closed under the rules of system* **P** *that satisfies the following condition:*

> $(*)$ *if $\varphi \rightarrow \psi \in T$, $[\![\varphi]\!] = [\![\varphi']\!]$, and $[\![\psi]\!] = [\![\psi']\!]$, then $\varphi' \rightarrow \psi' \in T$, for all formulas $\varphi, \varphi', \psi, \psi' \in \mathcal{L}$.*

There is a plausibility structure $PL_T = (W, \mathrm{Pl}_T, \pi)$ such that $\mathrm{Pl}_T([\![\varphi]\!]) \le \mathrm{Pl}_T([\![\psi]\!])$ if and only if $\varphi \vee \psi \rightarrow \psi \in T$. Moreover, $PL_T \models \varphi \rightarrow \psi$ if and only if $\varphi \rightarrow \psi \in T$.

The proof of this lemma, and all other proofs, can be found in Appendix A.

Theorem 2.4.2: *(a) Let $PP = \{\mathrm{Pr}_i\}$ be a PPD on W. There is a plausibility measure Pl_{PP} on W such that $(W, \{\mathrm{Pr}_i\}, \pi) \models_\epsilon \varphi \rightarrow \psi$ if and only if $(W, \mathrm{Pl}_{PP}, \pi) \models_{PL} \varphi \rightarrow \psi$.*

(b) Let $\prec$ be a preference ordering on W. There is a plausibility measure $\mathrm{Pl}_\prec$ on W such that $(W, \prec, \pi) \models_p \varphi \rightarrow \psi$ if and only if $(W, \mathrm{Pl}_\prec, \pi) \models_{PL} \varphi \rightarrow \psi$.

[7] We can take the domain of the plausibility measure to consist of sets of logically equivalent formulas, partially ordered so as to satisfy the constraints.

Proof: We start with part (a). Set $T_{PP} = \{\varphi{\rightarrow}\psi : (W, PP, \pi) \models_\epsilon \varphi{\rightarrow}\psi\}$. Theorem 2.3.3 implies that T_{PP} is closed under the KLM rules. Moreover, since T is constructed from a PPD over W using π it satisfies the requirement of Lemma 2.4.1. It follows that we can construct a plausibility structure that satisfies the requirements of the theorem. It is also easy to verify that this construction agrees with the construction described earlier, in that it satisfies constraint (2.1). To see this, let $A, B \in \mathcal{F}$. Then, according to our assumptions, there are formulas φ and ψ such that $A = [\![\varphi]\!]$ and $B = [\![\psi]\!]$. By definition, $\mathrm{Pl}_{PP}(A) \leq \mathrm{Pl}_{PP}(B)$ if and only if $(W, PP, \pi) \models_\epsilon (\varphi \vee \psi){\rightarrow}\psi$. From definition of $\models_\epsilon$ we immediately get (2.1).

The proof of part (b) is identical using Theorem 2.3.1. Again, an analogous argument easily shows that $\mathrm{Pl}_\prec$ satisfies (2.2). ∎

Thus, each of the semantic approaches to default reasoning that were considered in Section 2.3 can be mapped into plausibility structures in a way that preserves the semantics of defaults.

Other semantics for defaults can also be mapped into plausibility measures using the general technique of Lemma 2.4.1. In most cases, we can also establish a direct relationship between these semantics and plausibility measures. For example, the *coherent filters* approach of [Ben-David and Ben-Eliyahu 1994; Schlechta 1995] can be mapped to plausibility, as shown by Schlechta [1996], and Weydert's *full ranking measures* [1994a] are easily seen to be a special case of plausibility measures.

2.5 Default Entailment in Plausibility Structures

In this section we characterize default entailment in plausibility structures. To do so, it is useful to have a somewhat more general definition of entailment in plausibility structures.

Definition 2.5.1: If $\mathcal{S}$ is a class of plausibility structures, then a knowledge base Δ entails $\varphi{\rightarrow}\psi$ with respect to $\mathcal{S}$, denoted $\Delta \models_{\mathcal{S}} \varphi{\rightarrow}\psi$, if every plausibility structure $PL \in \mathcal{S}$ that satisfies all the defaults in Δ also satisfies $\varphi{\rightarrow}\psi$. ∎

The classes of structures we are interested in include $\mathcal{S}^{PL}$, the class of all plausibility structures, and $\mathcal{S}^{Poss}$, $\mathcal{S}^{\kappa}$, $\mathcal{S}^{p}$, $\mathcal{S}^{r}$, and $\mathcal{S}^{\epsilon}$, the classes that arise from mapping possibility structures, ordinal ranking structures, preferential structures, rational structures, and PPDs, respectively, into plausibility structures. (In the case of possibility structures and ordinal ranking structures, the mapping is the obvious one discussed in Section 2.2; in the case of preferential, rational and PPD structures, the mapping is the one described in Theorem 2.4.2.) Recall that all these mappings preserve the semantics of defaults.

It is easy to check that our semantics for defaults does *not* guarantee that the axioms of system **P** hold in all structures in $\mathcal{S}^{PL}$. In particular, they do not hold in probability structures. It is easy to construct a plausibility structure, $PL = (W, \mathrm{Pl}, \pi)$, where Pl is actually a probability measure Pr such that $\Pr([\![p \wedge q]\!]) > 0$, $\Pr([\![q]\!]|[\![p]\!]) > .5, \Pr([\![r]\!]|[\![p]\!]) > .5$, but $\Pr([\![q \wedge r]\!]|[\![p]\!]) < .5$ and $\Pr([\![r]\!]|[\![p \wedge q]\!]) < .5$. (For example, if $\Pr([\![q \wedge r]\!]|[\![p]\!]) = \Pr([\![\neg q \wedge \neg r]\!]|[\![p]\!]) = 0.1$ and $\Pr([\![q \wedge \neg r]\!]|[\![p]\!]) = \Pr([\![\neg q \wedge r]\!]|[\![p]\!]) = 0.4$.) Recall that if $\Pr([\![\varphi]\!]) > 0$ then $PL \models_{PL} \varphi{\rightarrow}\psi$

if and only if $\Pr([\![\psi]\!]|[\![\varphi]\!]) > .5$. Thus, $PL \models_{PL} (p \rightarrow q) \wedge (p \rightarrow r)$, but $PL \not\models_{PL} p \rightarrow (q \wedge r)$ and $PL \not\models_{PL} (p \wedge q) \rightarrow r$. This gives us a violation of both AND and CM. We can similarly construct a counterexample to OR. On the other hand, as the following result shows, plausibility structures do satisfy the other three axioms of system **P**. Let system $\mathbf{P}'$ be the system consisting of LLE, RW, and REF.

Theorem 2.5.2: *If $\Delta \vdash_{\mathbf{P}'} \varphi \rightarrow \psi$, then $\Delta \models_{\mathcal{S}^{PL}} \varphi \rightarrow \psi$.*

What extra conditions do we have to place on plausibility structures to ensure that AND, OR, and CM are satisfied? We focus first on the AND rule. We want an axiom that cuts out probability functions, but leaves more qualitative notions. Working at a semantic level, taking $[\![\varphi]\!] = A$, $[\![\psi_1]\!] = B_1$, and $[\![\psi_2]\!] = B_2$, and using $\overline{X}$ to denote the complement of X, the AND rule translates to:

A2′. For all sets A, B_1, and B_2, if $\mathrm{Pl}(A \cap B_1) > \mathrm{Pl}(A \cap \overline{B_1})$ and $\mathrm{Pl}(A \cap B_2) > \mathrm{Pl}(A \cap \overline{B_2})$, then $\mathrm{Pl}(A \cap B_1 \cap B_2) > \mathrm{Pl}(A \cap \overline{(B_1 \cap B_2)})$.

It turns out that in the presence of A1, the following somewhat simpler axiom is equivalent to A2′:

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 $\mathrm{Pl}(A) > \mathrm{Pl}(B \cup C)$.

Proposition 2.5.3: *A plausibility measure satisfies A2 if and only if it satisfies A2′.*

A2 can be viewed as a generalization of a natural requirement of qualitative plausibility: if A, B, and C are pairwise disjoint, $\mathrm{Pl}(A) > \mathrm{Pl}(B)$, and $\mathrm{Pl}(A) > \mathrm{Pl}(C)$, then $\mathrm{Pl}(A) > \mathrm{Pl}(B \cup C)$. Moreover, since A2 is equivalent to A2′, and A2′ is a direct translation of the AND rule into conditions on plausibility measures, any plausibility structure whose plausibility measure satisfies A2 also satisfies the AND rule. Somewhat surprisingly, a plausibility measure Pl that satisfies A2 also satisfies CM. Moreover, Pl satisfies the non-vacuous case of the OR rule. That is, if $\mathrm{Pl}([\![\varphi_1]\!]) > \perp$, then from $\varphi_1 \rightarrow \psi$ and $\varphi_2 \rightarrow \psi$ we can conclude $(\varphi_1 \vee \varphi_2) \rightarrow \psi$.[8] To handle the vacuous case of OR we need an additional axiom:

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

Thus, A2 and A3 capture the essence of the KLM properties. To make this precise, define a plausibility space (W, Pl) to be *qualitative* if it satisfies A2 and A3 in addition to A1. We say $PL = (W, \mathrm{Pl}, \pi)$ is a *qualitative plausibility structure* if (W, Pl) is a qualitative plausibility space. Let $\mathcal{S}^{QPL}$ consist of all qualitative plausibility structures.

[8] We remark that if we dropped requirement A1, then we can define properties of plausibilities measures that correspond precisely to CM and OR. The point is that in the presence of A1, A2—which essentially corresponds to AND—implies CM and the non-vacuous case of OR. Despite appearances, A1 does *not* correspond to RW. Semantically, RW says that if A and B are disjoint sets such that $\mathrm{Pl}(A) > \mathrm{Pl}(B)$, and $A \subseteq A'$, $B' \subseteq B$, and A' and B' are disjoint, then $\mathrm{Pl}(A') > \mathrm{Pl}(B')$. While this follows from A1, it is much weaker than A1.

Theorem 2.5.4: *$\mathcal{S} \subseteq \mathcal{S}^{QPL}$ if and only if for all Δ, φ, and ψ, if $\Delta \vdash_{\mathbf{P}} \varphi {\rightarrow} \psi$ then $\Delta \models_{\mathcal{S}} \varphi {\rightarrow} \psi$.*

Thus, the KLM axioms are sound for qualitative plausibility structures. We remark that Theorem 2.5.4 provides not only a sufficient but a necessary condition for a set of plausibility structures to satisfy the KLM properties: If the KLM axioms are sound with respect to $\mathcal{S}$, then all *PL* $\in \mathcal{S}$ must be qualitative.

This, of course, leads to the question of which plausibility structures are qualitative. All the ones we have been focusing on are.

Theorem 2.5.5: *Each of $\mathcal{S}^{Poss}$, $\mathcal{S}^{\kappa}$, $\mathcal{S}^{\epsilon}$, $\mathcal{S}^{p}$, and $\mathcal{S}^{r}$ is a subset of $\mathcal{S}^{QPL}$.*

It follows from Theorems 2.5.4 and 2.5.5 that the KLM properties hold in all the approaches to default reasoning considered in Section 2.3. While this fact was already known, this result gives us a deeper understanding as to *why* the KLM properties should hold. In a precise sense, it is because A2 and A3 holds for all these approaches.

We now consider completeness. To get soundness, we have to ensure that $\mathcal{S}$ does not contain too many structures, in particular, no structures that are not qualitative. To get completeness, we have to ensure that $\mathcal{S}$ contains "enough" structures. In particular, if $\Delta \not\vdash_{\mathbf{P}} \varphi {\rightarrow} \psi$, we want to ensure that there is a plausibility structure $PL \in \mathcal{S}$ such that $PL \models_{PL} \Delta$ and $PL \not\models_{PL} \varphi {\rightarrow} \psi$. The following weak condition on $\mathcal{S}$ does this.

Definition 2.5.6: We say that $\mathcal{S}$ is *rich* if for every collection $\varphi_1, \ldots, \varphi_n$, $n > 1$, of mutually exclusive formulas, there is a plausibility structure $PL = (W, \mathrm{Pl}, \pi) \in \mathcal{S}$ such that:

$$\mathrm{Pl}([\![\varphi_1]\!]) > \mathrm{Pl}([\![\varphi_2]\!]) > \cdots > \mathrm{Pl}([\![\varphi_n]\!]) = \bot. \blacksquare$$

The richness requirement is quite mild. It says that we do not have *a priori* constraints on the relative plausibilities of a collection of disjoint sets. Every collection of plausibility measures that we have considered thus far can be easily shown to satisfy this richness condition.

Theorem 2.5.7: *Each of $\mathcal{S}^{Poss}$, $\mathcal{S}^{\kappa}$, $\mathcal{S}^{p}$, $\mathcal{S}^{r}$, $\mathcal{S}^{\epsilon}$, and $\mathcal{S}^{QPL}$ is rich.*

More importantly, richness is a necessary and sufficient condition to ensure that the KLM properties are complete.

Theorem 2.5.8: *A set $\mathcal{S}$ of qualitative plausibility structures is rich if and only if for all finite Δ and defaults $\varphi {\rightarrow} \psi$, we have that $\Delta \models_{\mathcal{S}} \varphi {\rightarrow} \psi$ implies $\Delta \vdash_{\mathbf{P}} \varphi {\rightarrow} \psi$.*

Note that Theorem 2.5.8 deals with what is usually considered to be *weak* completeness. The *strong* notion of completeness would require us to remove the restriction that Δ is finite from the statement of the theorem. It is somewhat cumbersome, but not too difficult, to find a notion of richness that corresponds to strong completeness. Note that if $\models_{\mathcal{S}}$ is compact, then weak completeness implies strong completeness.

Putting together Theorems 2.5.4, 2.5.5, and 2.5.8, we get

Corollary 2.5.9: *For* $\mathcal{S} \in \{\mathcal{S}^{Poss}, \mathcal{S}^{\kappa}, \mathcal{S}^{p}, \mathcal{S}^{r}, \mathcal{S}^{\epsilon}, \mathcal{S}^{QPL}\}$*, and all* Δ*,* φ*, and* ψ*, we have* $\Delta \vdash_{\mathbf{P}} \varphi {\rightarrow} \psi$ *if and only if* $\Delta \models_{\mathcal{S}} \varphi {\rightarrow} \psi$*.*

Not only does this result gives us a straightforward and uniform proof that the KLM properties characterize default reasoning in each of the systems considered in Section 2.3, it gives us a general technique for proving completeness of the KLM properties for other semantics as well. All we have to do is to provide a mapping of the intended semantics into plausibility structures, which is usually straightforward, and then show that the resulting set of structures is qualitative and rich.

Theorem 2.5.8 also has important implications for attempts to go beyond the KLM properties (as was the goal in introducing rational structures). It says that any semantics for defaults that proceeds by considering a class $\mathcal{S}$ of structures satisfying the richness constraint, and defining $\Delta \models_{\mathcal{S}} \varphi {\rightarrow} \psi$ to hold if $\varphi {\rightarrow} \psi$ is true in every structure in $\mathcal{S}$ that satisfies Δ, cannot lead to new properties for entailment.

Thus, to go beyond KLM, we either need to consider interesting non-rich classes of structures, or to define a notion of entailment from Δ that does not consider all the structures of a given class. It is possible to construct classes of structures that are arguably interesting and violate the richness constraint. One way is to impose independence constraints. For example, suppose we consider all structures where p is independent of q in the sense that if any of *true*$\rightarrow q$, $p {\rightarrow} q$, and $\neg p {\rightarrow} q$ holds, then the others also do, so that discovering either p or $\neg p$ does not affect whether or not q is believed.[9] Restricting to such structures clearly gives us extra properties. For example, from *true*$\rightarrow q$ we can infer $p {\rightarrow} q$, which certainly does not follow from the KLM properties. Such structures do not satisfy the richness constraint, since we cannot have, for example, $\mathrm{Pl}([\![p \wedge q]\!]) > \mathrm{Pl}([\![p \wedge \neg q]\!]) > \mathrm{Pl}([\![\neg p \wedge \neg q]\!]) > \mathrm{Pl}([\![\neg p \wedge q]\!])$. Much of the recent work in default reasoning [Bacchus, Grove, Halpern, and Koller 1993; Geffner 1992a; Goldszmidt and Pearl 1992; Goldszmidt, Morris, and Pearl 1993; Lehmann and Magidor 1992; Pearl 1990] has taken the second approach. Roughly speaking, this approach can be viewed as taking the basic idea of preferential semantics—placing a preference ordering on worlds—one step further: We try to get from a knowledge base a set of preferred *structures* (where the structures themselves put a preference ordering on worlds)—for example, in [Goldszmidt, Morris, and Pearl 1993], the PPD of maximum entropy is considered—and carry out all reasoning in these preferred structures. We believe that plausibility measures will provide insight into techniques for choosing such preferred structures, particularly through the use of independence, but the discussion of this issue is beyond the scope of this work. (See [Friedman and Halpern 1995] for preliminary results in this direction.)

2.6 Expressiveness of Qualitative Plausibility Measures

In the previous section we showed that all approaches to default reasoning are instances of qualitative plausibility structures. We now show that each of the classes considered in Theorem 2.5.5 is a strict subset of $\mathcal{S}^{QPL}$. This is clearly true in a trivial sense. For example, if we consider a

[9] We remark that if we define independence appropriately in plausibility structures, this property does indeed hold; see [Friedman and Halpern 1995].

qualitative plausibility measure whose range is $[1, 2]$, it cannot be either a possibility measure or a κ-ranking. To get around this problem, we define two plausibility spaces (W_1, Pl_1) and (W_2, Pl_2) to be *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 $\mathrm{Pl}_2(h(A)) \leq \mathrm{Pl}_2(h(B))$. We claim that for each of the classes of plausibility measures considered in Theorem 2.5.5, there is a qualitative plausibility measure that is not order-isomorphic to any element of that class.

We first examine $\mathcal{S}^{Poss}$ and $\mathcal{S}^{\kappa}$. Since both require $\leq$ to be a total order, a qualitative plausibility measure that does not place a total order on the plausibility of subsets cannot be order-isomorphic to an element of $\mathcal{S}^{Poss}$ or $\mathcal{S}^{\kappa}$. In fact, it is not hard to show that there are qualitative plausibility measures that totally order the plausibility of subsets and yet are not order-isomorphic to possibility measures or κ-rankings.

Proposition 2.6.1: *There exists a totally ordered plausibility structure PL $\in \mathcal{S}^{QPL}$ such that PL is not order-isomorphic to any structure in $\mathcal{S}^{Poss}$ and $\mathcal{S}^{\kappa}$.*

Proof: Define a plausibility measure Pl on $\{a, b, c\}$ such that $\mathrm{Pl}(\{a\}) = \mathrm{Pl}(\{b\}) = \mathrm{Pl}(\{c\}) = \mathrm{Pl}(\{b, c\}) = \mathrm{Pl}(\{a, c\}) = 1/2$ and $\mathrm{Pl}(\{a, b\}) = \mathrm{Pl}(\{a, b, c\}) = 1$. Then $\mathrm{Pl}(\{c\}) < \mathrm{Pl}(\{a, b\})$, although neither $\mathrm{Pl}(\{c\}) < \mathrm{Pl}(\{a\})$ nor $\mathrm{Pl}(\{c\}) < \mathrm{Pl}(\{b\})$ hold. It is easy to see that this cannot happen for any plausibility measure in $\mathcal{S}^{Poss}$ or $\mathcal{S}^{\kappa}$. ∎

In fact, the plausibility measure defined in the proof of this proposition cannot be order-isomorphic to any element of $\mathcal{S}^{p}$ and $\mathcal{S}^{\epsilon}$ either. The construction of Lemma 2.4.1 is such that if (W, Pl, π) is in $\mathcal{S}^{p}$ or $\mathcal{S}^{\epsilon}$, and A and B are disjoint subsets of W, then $\mathrm{Pl}(A) \neq \mathrm{Pl}(B)$. Since $\mathrm{Pl}(\{a\}) = \mathrm{Pl}(\{b\})$, this plausibility measure cannot be isomorphic to one in $\mathcal{S}^{p}$ or $\mathcal{S}^{\epsilon}$.

If all we are interested in is default reasoning, then all that matters is the relative plausibility of disjoint sets. We say that two plausibility spaces (W_1, Pl_1) and (W_2, Pl_2) are *default-isomorphic* via h if h is a bijection from W_1 to W_2 such that, for disjoint subsets A and B of W_1, we have $\mathrm{Pl}_1(A) < \mathrm{Pl}_1(B)$ if and only if $\mathrm{Pl}_2(h(A)) < \mathrm{Pl}_2(h(B))$. Two spaces are said to be default-isomorphic if they are default isomorphic via some bijection h. Clearly, if (W_1, Pl_1) and (W_1, Pl_2) are default isomorphic via h, then, for all interpretations π_1 and π_2 such that $\pi_1(w) = \pi_2(h(w))$, the structures $(W_1, \mathrm{Pl}_1, \pi_1)$ and $(W_1, \mathrm{Pl}_2, \pi_2)$ satisfy the same defaults.

Even if we restrict to plausibility measures that totally order all subsets, there are qualitative plausibility spaces that are not default isomorphic to possibility spaces or κ-rankings.

Proposition 2.6.2: *There exists a totally ordered plausibility structure PL $\in \mathcal{S}^{QPL}$ such that PL is not default-isomorphic to any structure in $\mathcal{S}^{Poss}$ and $\mathcal{S}^{\kappa}$.*

Proof: The plausibility measure described in the proof of Proposition 2.6.1 is also a counterexample for default-isomorphism. ∎

In fact, the construction of Proposition 2.6.1 describes a plausibility measure that is not default-isomorphic to any structure in $\mathcal{S}^{p}$ either. The situation is quite different for $\mathcal{S}^{\epsilon}$ though. We say that a plausibility space (W, Pl) is *normal* (following Lewis [1973]) if $\mathrm{Pl}(W) > \perp$. A plausibility structure *PL* $= (W, \mathrm{Pl}, \pi)$ is *normal* if the plausibility space (W, Pl) is.

Theorem 2.6.3: *Let PL $\in \mathcal{S}^{QPL}$ be a normal plausibility structure with a countable algebra $\mathcal{F}$. Then there is a structure PL$'$ $\in \mathcal{S}^{\epsilon}$ default-isomorphic to PL.*

Since the evaluation of default statements is determined by the (strict) ordering of disjoint events and the algebra defined by formulas (that is, the algebra $\{[\![\varphi]\!] : \varphi \in \mathcal{L}\}$) is countable if the language $\mathcal{L}$ is countable, we get as an immediate corollary to Theorem 2.6.3 a converse to Theorem 2.4.2 in the case of ϵ-semantics.

Corollary 2.6.4: *Let $\mathcal{L}$ be a countable language and $(W, \mathrm{Pl}, \pi) \in \mathcal{S}^{QPL}$ be a normal plausibility structure. Then there exists a structure $(W, \mathrm{Pl}', \pi) \in \mathcal{S}^{\epsilon}$ such that $(W, \mathrm{Pl}, \pi) \models \varphi {\rightarrow} \psi$ if and only if $(W, \mathrm{Pl}', \pi) \models \varphi {\rightarrow} \psi$ for all $\varphi, \psi \in \mathcal{L}$.*

Thus, with respect to conditional statements, $\mathcal{S}^{\epsilon}$ is as expressive (in a strong sense) as $\mathcal{S}^{QPL}$. We do note that for uncountable languages there is a difference $\mathcal{S}^{QPL}$ and $\mathcal{S}^{\epsilon}$: Probability distributions can assign positive weight only to a countable number of pairwise disjoint events, while qualitative plausibility measures do not suffer from such constraints.

2.7 Epistemic Entrenchment

There has been much work related to defaults and plausibility. It can roughly be divided into three categories. The first consists of various approaches to dealing with uncertainty such as the ones mentioned in Section 2.2. For a more detailed comparison to such approaches see [Friedman and Halpern 1995]. The second category consists of semantics for defaults that are discussed at length in Section 2.4. The final category includes semantics for defaults that are linguistic in nature. The most well known approach of this kind is *epistemic entrenchment* [Gärdenfors and Makinson 1988; Grove 1988]. This has been proposed as a semantics for *belief revision* [Gärdenfors 1988]. Recently, Gärdenfors and Makinson [1989] proposed using a similar notion of expectation ordering as a semantics for default reasoning. We briefly review their approach here.

Let $\mathcal{L}$ be some logical language that includes the usual propositional connectives with a compact consequence relation $\vdash_{\mathcal{L}}$ that satisfies the axioms of propositional logic. An *expectation ordering* $\trianglelefteq$ on $\mathcal{L}$ is a relation over formulas in L that satisfies the following requirements:

E1. $\trianglelefteq$ is transitive,

E2. if $\vdash_{\mathcal{L}} \varphi \Rightarrow \psi$ then $\varphi \trianglelefteq \psi$,

E3. for any φ and ψ, either $\varphi \trianglelefteq \varphi \wedge \psi$ or $\psi \trianglelefteq \varphi \wedge \psi$.

Intuitively, $\varphi \trianglelefteq \psi$ if ψ is as at least as expected as φ, so the agent would not retract his belief in ψ before retracting his belief in φ. We do not go here into the motivation for E1–E3. It is not hard to verify that E1–E3 imply that $\trianglelefteq$ is a total preorder on $\mathcal{L}$.

An *expectation structure* is a pair $E = (\mathcal{L}, \trianglelefteq)$, where $\trianglelefteq$ is an expectation ordering on $\mathcal{L}$. Intuitively, E satisfies $\varphi {\rightarrow} \psi$ if ψ is the consequence of formulas that are expected given φ. This definition hinges on the choice of formulas that are expected given φ. Gärdenfors and Makinson take these to be the formulas that are more expected than $\neg\varphi$. Formally, an expectation structure $E = (\mathcal{L}, \trianglelefteq)$ satisfies a default $\varphi \rightarrow \psi$ if $\{\varphi\} \cup \{\xi : \neg\varphi \triangleleft \xi\} \vdash_{\mathcal{L}} \psi$, where $\varphi \triangleleft \psi$ holds if $\varphi \trianglelefteq \psi$ and not $\psi \trianglelefteq \varphi$. This definition can restated as follows:

Theorem 2.7.1: [Gärdenfors and Makinson 1989] *Let $E = (\mathcal{L}, \trianglelefteq)$ be an expectation structure. $E \models \varphi \rightarrow \psi$ if and only if $\vdash_{\mathcal{L}} \varphi \Rightarrow \psi$ or $(\varphi \Rightarrow \neg\psi) \triangleleft (\varphi \Rightarrow \psi)$.*

While this definition seems quite different than the one described in Section 2.4, the two are in fact closely related. Notice that $\varphi \Rightarrow \neg\psi$ is equivalent to $\neg(\varphi \wedge \psi)$, while $\varphi \Rightarrow \psi$ is equivalent to $\neg(\varphi \wedge \neg\psi)$. Thus, the second clause in the theorem above, which says $(\varphi \Rightarrow \neg\psi) \triangleleft (\varphi \Rightarrow \psi)$, can be viewed as saying $\neg(\varphi \wedge \psi)$ is less expected than $\neg(\varphi \wedge \neg\psi)$; this is clearly much in the spirit of the second clause in the definition of defaults in plausibility, which says that $(\varphi \wedge \psi)$ must be more plausible than $(\varphi \wedge \neg\psi)$. If we identify p being more plausible than q with $\neg p$ being less expected than $\neg q$, they are equivalent. The first clause in the theorem, $\vdash_L \varphi \Rightarrow \psi$, corresponds to the vacuous case that φ has plausibility $\perp$ in our definition. However, as we now show, Gärdenfors and Makinson treat the vacuous case in a somewhat nonstandard manner (which can still be captured using plausibility).

To make the relationship between expectation orderings and plausibility precise, let $E = (\mathcal{L}, \trianglelefteq)$ be an expectation structure. We say that a set $V \subseteq \mathcal{L}$ is *consistent* if for all $\varphi_1, \ldots, \varphi_n \in V$, we have $\not\vdash_{\mathcal{L}} \neg(\varphi_1 \wedge \ldots \wedge \varphi_n)$. V is a *maximal consistent* set if it is consistent and for each $\varphi \in \mathcal{L}$, either $\varphi \in V$ or $\neg\varphi \in V$. We now construct a plausibility structure $PL_E = (W_E, \mathrm{Pl}_E, \pi_E)$. We define $W_E = \{w_V : V$ is a maximally consistent subset of $\mathcal{L}\}$ and $\pi_E(w_V)(p) = \mathbf{true}$ if $p \in V$. Finally, we need to define Pl_E. The obvious choice is to define $\mathrm{Pl}_E([\![\varphi]\!]) \leq \mathrm{Pl}_E([\![\psi]\!])$ if and only if $(\neg\psi) \trianglelefteq (\neg\varphi)$. It is easy to see that this implies that $(\varphi \Rightarrow \neg\varphi) \triangleleft (\varphi \Rightarrow \psi)$ if and only if $\mathrm{Pl}_E([\![\varphi \wedge \psi]\!]) > \mathrm{Pl}_E([\![\varphi \wedge \neg\psi]\!])$. Thus, E and PL_E agree on non-vacuous defaults. However, suppose that $\mathrm{Pl}_E([\![\varphi]\!]) = \perp$ for some consistent formula φ. It follows that $PL_E \models \varphi \rightarrow \psi$ for any ψ. On the other hand, it is not hard to show that $E \models \varphi \rightarrow \psi$ if and only if $\vdash_L \varphi \Rightarrow \psi$.[10] We can easily modify the definition of Pl_E to avoid this problem: We define $\mathrm{Pl}_E([\![\varphi]\!]) \leq \mathrm{Pl}_E([\![\psi]\!])$ either if $(\neg\psi) \trianglelefteq (\neg\varphi)$ and it is not the case that *true* $\trianglelefteq \neg\psi$ or if $[\![\varphi]\!] \subseteq [\![\psi]\!]$. With this modified definition, we get the desired result.

Proposition 2.7.2: *If E is an expectation structure, then PL_E is a plausibility structure. Furthermore, $E \models \varphi \rightarrow \psi$ if and only if $PL_E \models \varphi \rightarrow \psi$.*

We now examine default entailment with respect to expectation orderings. Let $\mathcal{S}^E$ be the set of plausibility structures that correspond to expectation structures. It is not hard to prove that

Theorem 2.7.3: *$\mathcal{S}^E$ is a subset of $\mathcal{S}^{QPL}$.*

[10] Proof sketch: The "if" direction follows from Theorem 2.7.1. For the "only if" direction, $\mathrm{Pl}_E([\![\varphi]\!]) = \perp$, so we must have *true* $\trianglelefteq \neg\varphi$. Since $\vdash_L \neg\varphi \Rightarrow (\varphi \Rightarrow \neg\psi)$, it follows from E1 and E2 that *true* $\trianglelefteq (\varphi \Rightarrow \neg\psi)$. Moreover, since $\vdash_L (\varphi \Rightarrow \psi) \Rightarrow$ *true* we have $(\varphi \Rightarrow \psi) \trianglelefteq$ *true*. Thus, we cannot have $(\varphi \Rightarrow \neg\psi) \triangleleft (\varphi \Rightarrow \psi)$, for otherwise, by E2, we would have *true* $\triangleleft$ *true*, a contradiction.

It immediately follows that the KLM properties are sound for default entailment with expectation orderings, i.e., with respect to $\mathcal{S}^E$. The KLM properties, however, are not complete with respect to $\mathcal{S}^E$. For example, if p and q are arbitrary primitive propositions, then $p{\rightarrow}$*false* entails $q{\rightarrow}$*false*. This example is a consequence of a property Gärdenfors and Makinson call *consistency preservation*:

- $E \models \varphi{\rightarrow}$*false* if and only if $\vdash_{\mathcal{L}} \varphi \Rightarrow$ *false*.

This property states that φ is totally *implausible* —that is, has plausibility $\perp$—if and only if it is inconsistent. This implies that no $\mathrm{Pl}_E \in \mathcal{S}^E$ satisfies $p{\rightarrow}$*false*, and hence $p{\rightarrow}$*false* entails, vacuously, all other defaults. Thus, one cannot specify that events, such as *white*$\wedge$*black*, are impossible in the database Δ; these constraints must be somehow embedded in $\vdash_{\mathcal{L}}$.

We note that expectation orderings are similar to plausibility measures in that they order events. However, there are several differences. First, expectation orderings use formulas to denote events. (We remark that there are similar formulations of probability theory [Jeffreys 1961] that are based on a linguistic description of events.) Secondly, as shown by our construction of PL_E, expectation orderings order events according to the implausibility of their complements. (This type of ordering is usually called the *dual order* [Dubois and Prade 1990; Friedman and Halpern 1995; Shafer 1976].) Thirdly, the treatment of the vacuous case is slightly different. This difference leads to additional properties of default entailment.

2.8 Conditional Logic

Up to now, we focused on whether a set of defaults implies another default. We have not considered a full logic of defaults, with negated defaults, nested defaults, and disjunctions of defaults. It is easy to extend all the approaches we defined so far to deal with such a logic. *Conditional logic* is a logic that treats $\rightarrow$ as a modal operator. The syntax of the logic is simple: let $\mathcal{L}^C$ be the language defined by starting with primitive propositions, and closing off under $\wedge$, $\neg$, and $\rightarrow$. Formulas can describe logical combination of defaults (e.g., $(p{\rightarrow}q) \vee (p{\rightarrow}\neg q)$) as well as nested defaults (e.g., $(p{\rightarrow}q){\rightarrow}r$).

We note that the connections between default reasoning and conditional logics are well-known; see [Boutilier 1994a; Crocco and Lamarre 1992; Kraus, Lehmann, and Magidor 1990; Katsuno and Satoh 1991]. We gloss over the subtle philosophical differences between the two here.

The semantics of conditional logic is similar to the semantics of defaults. As with defaults, we evaluate conditional statements such as $\varphi{\rightarrow}\psi$ by comparing the plausibility of those worlds that satisfy $\varphi \wedge \psi$ to the plausibility of those worlds that satisfy $\varphi \wedge \neg\psi$. Unlike default reasoning, conditional logic allows us to combine defaults with propositional statements. Thus, $p \wedge (q{\rightarrow}r)$ is a formula of conditional logic, and is satisfied if both p and $q{\rightarrow}r$ are satisfied. The truth of a formula such as $p \wedge (q{\rightarrow}r)$ depends on the world; $p \wedge (q{\rightarrow}r)$ might be true in w_1 and false in w_2 if p is true at w_1 and not at w_2.

Conditional logic also allows us to consider nested conditionals. For example, to evaluate $(p{\rightarrow}q){\rightarrow}r$, we need to consider the plausibility of the worlds that satisfy r and $p{\rightarrow}q$ and compare them to the plausibility of worlds that satisfy $\neg r$ and $p \rightarrow q$. In the structures we considered in the preceding sections, a statement such as $p \rightarrow q$ is determined by the global plausibility measure. Thus, the set of worlds that satisfy $p \rightarrow q$ is either the empty set or W (i.e., all possible worlds). The usual definition of conditional logic [Lewis 1973] provides a non-trivial semantics for nested conditionals by associating with each world a different preferential order over worlds.

We now give a similar definition based on plausibility measures. A *(generalized) plausibility structure* is a tuple $(W, \mathcal{P}, \pi)$ where W and π are, as usual, a set of worlds and a mapping from worlds to truth assignments, and $\mathcal{P}$ maps each world w to a plausibility space $(W_w, \mathcal{F}_w, \mathrm{Pl}_w)$ where $W_w \subseteq W$. Intuitively, $(W_w, \mathcal{F}_w, \mathrm{Pl}_w)$ describes the agent's plausibility when she is in world w. We can view the plausibility structures we defined in previous sections to be a special case of generalized plausibility structures where $\mathcal{P}(w)$ is the same for all worlds w. From now on we focus on generalized plausibility structures, and will refer to them as plausibility structures where it is clear that we are not referring to plausibility structures as defined in Section 2.2.

Given a plausibility structure $PL = (W, \mathcal{P}, \pi)$, we define what it means for a formula φ to be true at a world w in PL. The definition for the propositional connectives is standard, and for $\rightarrow$, we use the definition already given:

- $(PL, w) \models p$ if $\pi(w) \models p$ for a primitive proposition p
- $(PL, w) \models \neg\varphi$ if $(PL, w) \not\models \varphi$
- $(PL, w) \models \varphi \wedge \psi$ if $(PL, w) \models \varphi$ and $(PL, w) \models \psi$
- $(PL, w) \models \varphi \rightarrow \psi$ if either $\mathrm{Pl}_w([\![\varphi]\!]_{(PL,w)}) = \perp$ or $\mathrm{Pl}_w([\![\varphi \wedge \psi]\!]_{(PL,w)}) > \mathrm{Pl}_w([\![\varphi \wedge \neg\psi]\!]_{(PL,w)})$, where we define $[\![\varphi]\!]_{(PL,w)} = \{w \in W_w : (\mathrm{Pl}, w) \models \varphi\}$.

It easy to conceive a generalized structure that uses preferential orderings, κ-rankings, ϵ-semantics or possibility measures instead of plausibility measures. As before, all of these structures can be embedded in qualitative plausibility structures. We denote by $\mathcal{S}_c^{QPL}$ the class of all qualitative generalized plausibility structures, and similarly denote the subclasses that correspond to various semantics (e.g., $\mathcal{S}_c^{p}$ is the class that consists of plausibility structures based on preference orderings).

We now want to axiomatize default reasoning in this framework. Clearly we need axioms and inference rules that generalize those of system **P**. Let system **C** be the system consisting of the following axioms and inference rules:

C0. All substitution instances of propositional tautologies

C1. $\varphi \rightarrow \varphi$

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

C3. $((\varphi_1 \rightarrow \psi) \wedge (\varphi_2 \rightarrow \psi)) \Rightarrow ((\varphi_1 \vee \varphi_2) \rightarrow \psi)$

C4. $((\varphi_1 \rightarrow \varphi_2) \wedge (\varphi_1 \rightarrow \psi)) \Rightarrow ((\varphi_1 \wedge \varphi_2) \rightarrow \psi)$

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

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

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

It is easy to see that system **C** is very similar to system **P**. The richer language lets us replace a rule like AND by the axiom C2. Similarly, C1, C3, C4, RC1, and RC2 are the analogues of REF, OR, CM, LLE, and RW, respectively. We need C0 and R1 to deal with propositional reasoning.

System **C** is essentially the axiom system described by Burgess [1981] for conditional logic defined over preferential structures.

Theorem 2.8.1: [Burgess 1981] *System* **C** *is a sound and complete axiomatization of* $\mathcal{L}^C$ *with respect to* $\mathcal{S}_c^p$.

Since the axioms of system **C** are clearly valid in all the structures in $\mathcal{S}_c^{QPL}$ and $\mathcal{S}_c^p \subseteq \mathcal{S}_c^{QPL}$, we immediately get that System **C** is a complete axiomatization of $\mathcal{L}^C$ with respect to $\mathcal{S}_c^{QPL}$. However, the proof of Theorem 2.8.1 is quite complicated, while there is a straightforward proof of the completeness of System **C** using plausibility.

We start with a little notation that is useful in this proof and in the following discussion. Let $N\varphi$ be an abbreviation for $\neg\varphi \rightarrow$*false*. (This operator is called the *outer modality* in [Lewis 1973].) Expanding the definition of $\rightarrow$, we get that $N\varphi$ holds at w if and only if Pl($[\![\neg\varphi]\!]$) $=\perp$. Thus, $N\varphi$ holds if $\neg\varphi$ is considered completely implausible. Thus, it implies that φ is true "almost everywhere".

Theorem 2.8.2: *System* **C** *is a sound and complete axiomatization of* $\mathcal{L}^C$ *with respect to* $\mathcal{S}_c^{QPL}$.

Proof: It is easy to verify that System **C** is sound in $\mathcal{S}_c^{QPL}$. To prove completeness, we have to show that if $\models_{\mathcal{S}_c^{QPL}} \varphi$, then $\vdash_{\mathbf{C}} \varphi$. This is equivalent to showing that if $\not\vdash_{\mathbf{C}} \varphi$ (i.e., $\neg\varphi$ is consistent) then $\not\models_{\mathcal{S}_c^{QPL}} \varphi$ (i.e., $\neg\varphi$ is satisfiable).

We now construct a *canonical* qualitative plausibility structure *PL* such that for all $\xi \in \mathcal{L}^C$ we have that if ξ is consistent, then $(PL, w) \models \xi$ for some world w.

We proceed as follows. Recall that a set of formulas $V \subseteq \mathcal{L}^C$ is a *maximal* $\vdash_{\mathbf{C}}$*-consistent* set if it is consistent with respect to $\vdash_{\mathbf{C}}$ and for each $\varphi \in \mathcal{L}^C$, either $\varphi \in V$ or $\neg\varphi \in V$. Let V be a set of formulas. We define $V/N = \{\varphi : N\varphi \in V\}$.

We construct a canonical model $PL = (W, \mathcal{P}, \pi)$ as follows:

- $W = \{w_V : V \subseteq \mathcal{L}^C$ is a maximal $\vdash_{\mathbf{C}}$-consistent set of formulas $\}$,
- $\mathcal{P}(w_V) = (W_w, \mathcal{F}, \mathrm{Pl}_{w_V})$ where
  - $W_{w_V} = \{w_U : V/N \subseteq U\}$,
  - $\mathcal{F}_{w_V} = \{[\varphi]_{w_V} : \varphi \in \mathcal{L}^C\}$, where $[\varphi]_{w_V} = \{w_U \in W_{w_V} : \varphi \in U\}$,
  - Pl_{w_V} is such that $\mathrm{Pl}_{w_V}([\varphi]_{w_V}) \leq \mathrm{Pl}_{w_V}([\psi]_{w_V})$ if and only if $(\varphi \vee \psi) \rightarrow \psi \in V$, and
- $\pi(w_V)(p) =$ **true** if and only if $p \in V$.

We can think of $[\varphi]_{w_V}$ as $[\![\varphi]\!]_{(PL,w)}$; as we shall see, both turn out to be equivalent.

Lemma 2.8.3: *PL is a qualitative plausibility structure. Moreover, $(PL, w_V) \models \varphi$ if and only if $\varphi \in V$ for all formulas φ and worlds w_V.*

The proof of the lemma appears in Appendix A. We note that the proof that *PL* is a plausibility structure is similar that of Lemma 2.4.1. The fact that *PL* is qualitative follows from Theorem 2.5.4.

Lemma 2.8.3 implies that $PL \in \mathcal{S}_c^{QPL}$. Let ξ be a $\vdash_{\mathbf{C}}$-consistent formula. Using standard arguments, it is easy to show that there is some maximal $\vdash_{\mathbf{C}}$-consistent set V_ξ such that $\xi \in V_\xi$. Thus, according to Lemma 2.8.3 $(PL, w_{V_\xi}) \models \xi$, and thus, ξ is satisfiable in $\mathcal{S}_c^{QPL}$. This concludes the proof of Theorem 2.8.2. ∎

This result shows that, at least as far as the language $\mathcal{L}^C$ goes, plausibility structures are no more expressive than preferential structures in the sense that both classes of structures are characterized by the same set of axioms.

As we saw in Section 2.4, the language of defaults is weak. As a result, the same axioms characterize reasoning in a wide range of semantics. The language $\mathcal{L}^C$ is richer than the language of defaults, and allows us to make additional distinctions. For example, consider the following axiom:

C5. $\varphi \rightarrow \psi \wedge \neg(\varphi \rightarrow \neg\xi) \Rightarrow \varphi \wedge \xi \rightarrow \psi$

Axiom C5 corresponds to the rule of *rational monotonicity* discussed in [Kraus, Lehmann, and Magidor 1990; Lehmann and Magidor 1992]. It is known that C5 characterizes rational preferential orderings. We now examine necessary and sufficient conditions for C5 to hold. We say that a plausibility measure Pl is *rational* if it satisfies the following two properties:

A4. For all pairwise disjoint sets A, B and C, if $\mathrm{Pl}(A) < \mathrm{Pl}(B)$, then $\mathrm{Pl}(A) < \mathrm{Pl}(C)$ or $\mathrm{Pl}(C) < \mathrm{Pl}(B)$.

A5. For all pairwise disjoint sets A, B and C, if $\mathrm{Pl}(A) < \mathrm{Pl}(B \cup C)$, then $\mathrm{Pl}(A) < \mathrm{Pl}(B)$ or $\mathrm{Pl}(A) < \mathrm{Pl}(C)$.

A4 says that ordering of disjoint sets is modular. (Recall that an ordering is modular if there are no three elements x, y, z such that $x > y$ and z incomparable to both x and y.) A5 says that the plausibility of $B \cup C$ is essentially the maximum of the plausibility of B and C. Thus, $B \cup C$ cannot be more plausible than A, unless one of the components is more plausible than A.

It is not hard to show that C5 is valid in, and only in, systems where the plausibility ordering is rational.

Proposition 2.8.4: *Let $\mathcal{S} \subseteq \mathcal{S}_c^{QPL}$. C5 is valid in $\mathcal{S}$ if and only if all structures in $\mathcal{S}$ are rational.*

In particular, C5 is valid in each of $\mathcal{S}_c^r$, $\mathcal{S}_c^{Poss}$, and $\mathcal{S}_c^\kappa$, although it is not valid in $\mathcal{S}_c^p$.

As we said above, condition A4 states that the ordering is "almost" modular, in the sense that, when restricted to pairwise disjoint sets, it is modular. It is not surprising to see modularity arise in this context. It is well known that a modular ordering induces a total order. More precisely, if $<$ is a modular order on some set, then if we define $x \leq y$ as $y \not< x$, then $\leq$ is a *total* order, that is, a partial order such that either $x \leq y$ or $y \leq x$ for all x and y. All approaches that satisfy rational monotonicity that have been proposed in the literature involve structures where there is a total or modular order on worlds (e.g., rational preference orderings, κ-rankings, and possibility measures).

We say that a plausibility measure Pl is a *ranking* if it satisfies the following two properties:

A4$'$. $\leq_D$ is a total order; that is, either $\mathrm{Pl}(A) \leq_D \mathrm{Pl}(B)$ or $\mathrm{Pl}(B) \leq_D \mathrm{Pl}(A)$ for all sets A, B.

A5$'$. $\mathrm{Pl}(A \cup B) = \max(\mathrm{Pl}(A), \mathrm{Pl}(B))$ for all sets A, B.

It is easy to see that A4$'$ implies A4, and that in the presence of A4$'$, A5$'$ implies A5 (A4$'$ is required to ensure that the two plausibilities values have a maximum). Thus, any ranking is a rational measure. The opposite, however, is not true. It is easy to verify that A4 and A5 do not imply A4$'$ and A5$'$. Thus, there is a discrepancy between the properties that are necessary to satisfy C5 and those studied in the literature. However, we now show that if we care only about defaults, then there is no difference between rational structures and *ranked structures*, that is, plausibility structures where the the plausibility measure is a ranking.

Theorem 2.8.5: *Let* $(W, \mathcal{F}, \mathrm{Pl})$ *be a rational qualitative plausibility space. Then there is a plausibility space* $(W, \mathcal{F}, \mathrm{Pl}')$ *such that* Pl' *is a ranking, and* Pl *and* Pl' *are default-isomorphic; that is,* $\mathrm{Pl}(A) > \mathrm{Pl}(B)$ *if and only if* $\mathrm{Pl}'(A) > \mathrm{Pl}'(B)$ *for all disjoint sets* $A, B \in \mathcal{F}$.

Corollary 2.8.6: *Let* $PL = (W, \mathcal{P}, \pi)$ *be a rational plausibility structure. Then, there is a ranked plausibility structure* $PL' = (W, \mathcal{P}', \pi)$ *such that* $(PL, w) \models \varphi$ *if and only if* $(PL', w) \models \varphi$ *for all worlds* w *and formulas* $\varphi \in \mathcal{L}^C$.

It is also easy to show that rankings are always qualitative:

Proposition 2.8.7: *Let* Pl *be a plausibility measure. If* Pl *satisfies A1, A4$'$ and A5$'$, then it also satisfies A2 and A3.*

Another property that can be captured in $\mathcal{L}^C$ is *normality*, that is, that $\mathrm{Pl}(W) > \perp$. This corresponds to the axiom

C6. $\neg(\mathit{true}{\rightarrow}\mathit{false})$.

It is not hard to show that C6 is valid in each of $\mathcal{S}_c^{Poss}$, $\mathcal{S}_c^{\kappa}$ and $\mathcal{S}^{\epsilon}$.

Both rationality and normality are properties of particular plausibility measures. We can also examine properties that relate the plausibility measures associated with different worlds. For example, we might require that w be considered plausible according to Pl_w; i.e., $\mathrm{Pl}_w(\{w\}) > \perp$. Stating this condition, however, leads to a technical problem. Recall that Pl_w is defined over the set of measurable subsets of W_w. In general, singletons may not be measurable. Thus, we examine a slightly weaker condition, which we call *reflexivity* (following [Lewis 1973]): a structure *PL* is reflexive if for all worlds w, we have that $w \in W_w$ and $\mathrm{Pl}_w(A) > \perp$ for all $A \in \mathcal{F}_w$ such that $w \in A$.

Another condition we might require is that the plausibility assignment function be uniform in the sense that if $w' \in W_w$, then the plausibility assessment at w' is the same as that in w. This condition, intuitively, says that the agent considers possible only worlds where she has the same plausibility assessment. Formally, we say that *PL* is *uniform* if $w' \in W_w$ implies $\mathcal{P}(w') = \mathcal{P}(w)$ for all worlds w and w'.[11]

The next result shows that C5 characterizes rational structures, C6 characterizes normal structures, and

C7. $N\varphi \rightarrow \varphi$

C8. $[(\varphi \rightarrow \psi) \Rightarrow N(\varphi \rightarrow \psi)] \wedge [\neg(\varphi \rightarrow \psi) \Rightarrow N\neg(\varphi \rightarrow \psi)]$

characterize reflexive and uniform structures, respectively.

We now make these statements formal. Let *PL* be a plausibility structure. We consider the following restriction applied to *PL*:

R *rationality:* Pl_w is rational for all worlds w.

N *normality:* Pl_w in normal for all worlds w.

X *reflexivity:* *PL* is reflexive.

U *uniformity:* *PL* is uniform.

Thus, for example, ϵ semantics, possibility measures, and κ-rankings are all normal orderings. Similarly, rational preferential orderings, possibility measures and κ-rankings are all rational. Finally, the plausibility structures we considered in the previous sections are all uniform, since $\mathcal{P}(w)$ is the same for all worlds.

Theorem 2.8.8: *Let $\varphi \in \mathcal{L}^C$, let $\mathcal{A}$ be a subset of $\{R, N, X, U\}$, and let A be the corresponding subset of $\{C5, C6, C7, C8\}$. Then φ is valid in the subset of $\mathcal{S}_c^{QPL}$ that satisfies $\mathcal{A}$ if and only if φ is provable from system* $\mathbf{C}$*+A.*

Since the language $\mathcal{L}^C$ can only reason about the relative plausibility of disjoint sets, we can apply Theorem 2.8.5 and get the same conclusion where we replace rational by ranked.

Corollary 2.8.9: *Let $\varphi \in \mathcal{L}^C$, let $\mathcal{A}$ be a subset of $\{N, X, U\}$, and let A be the corresponding subset of $\{C6, C7, C8\}$. Then φ is valid in the subset of ranked structures in $\mathcal{S}_c^{QPL}$ that satisfies $\mathcal{A}$ if and only if φ is provable from system* $\mathbf{C}$*+ C5 +A.*

Using this theorem, we can easily prove sound and complete axiomatizations for $\mathcal{S}_c^{\kappa}$, $\mathcal{S}_c^{Poss}$ and $\mathcal{S}_c^{\epsilon}$.

Theorem 2.8.10: *(a)* $\mathbf{C}$*+$\{C6\}$ is a sound and complete axiomatization of $\mathcal{S}_c^{\epsilon}$.*

(b) $\mathbf{C}$ *+$\{C5, C6\}$ is a sound and complete axiomatization of $\mathcal{S}_c^{\kappa}$ and $\mathcal{S}_c^{Poss}$.*

[11] This restriction is called *absoluteness* by Lewis [1973], and should not be confused with the (weaker) restriction he calls uniformity.

2.9 Discussion

In this chapter we made three major contributions: the introduction of plausibility measures, the unification of all earlier results regarding the KLM properties into one framework, and a general result showing the inevitability of these properties. Do we really need plausibility measures? If all we are interested in is propositional default reasoning and the KLM properties, then the results of Section 2.8 show that preferential structures provide us all the expressive power we need. Roughly speaking, this is so because when doing propositional reasoning, we can safely restrict to finite structures. (Technically, this is because we have a *finite model property*: if a formula in $\mathcal{L}^C$ is satisfiable, it is satisfiable in a finite plausibility structure [Friedman and Halpern 1994c].) As we show in [Friedman, Halpern, and Koller 1996], preferential structures and plausibility structures are no longer equally expressive once we move to a first-order logic, precisely because infinite structures now play a more important role. Another instance where infinite structures appear is when we consider time explicitly in the framework. As we shall see in Chapter 4, this results in infinite structures even in the propositional case.

Chapter 3

Adding Knowledge

3.1 Introduction

In the previous chapter, we developed qualitative plausibility measures for representing defaults. In this chapter, we use qualitative plausibility measures as a semantics for beliefs. Moreover, we combine this semantics for beliefs with a semantics of knowledge. This results in a model that allows us to express what the agents are certain of (knowledge) and what they assume (beliefs).

We start with a brief review of the standard models for knowledge and beliefs (see [Halpern and Moses 1992] for further motivation and details). We then show how to combine knowledge with plausibility. Finally, we compare the derived notion of belief with previous definitions of belief.

3.2 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 $K_i\varphi$ is read "agent i knows φ". The logic of belief is the result of replacing the K_i operator by B_i. The formula $B_i\varphi$ 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 binary relations on the worlds in W, called *accessibility* relations. 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 $(M, w) \models \varphi$ 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) = $ *true*,

- $(M, w) \models \neg\varphi$ if $(M, w) \not\models \varphi$,
- $(M, w) \models \varphi \wedge \psi$ if $(M, w) \models \varphi$ and $(M, w) \models \psi$,
- $(M, w) \models K_i\varphi$ if $(M, w') \models \varphi$ 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 $B_i\varphi$ is exactly the same as for $K_i\varphi$.

Let $\mathcal{M}_K$ be the class of Kripke structures described above. We say that $\varphi \in \mathcal{L}^K$ is valid in some $M \in \mathcal{M}_K$ if $(M, w) \models \varphi$ for all w in M. We say that $\varphi \in \mathcal{L}^K$ 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 $(M, w) \models \varphi$.

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. $K_i\varphi \wedge K_i(\varphi \Rightarrow \psi) \Rightarrow K_i\psi$

K3. $K_i\varphi \Rightarrow \varphi$

K4. $K_i\varphi \Rightarrow K_iK_i\varphi$

K5. $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$

K6. $\neg K_i$*false*

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

RK2. From φ infer $K_i\varphi$

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 3.2.1: [Halpern and Moses 1992] *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 work, we take knowledge to satisfy the axioms of S5. 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 3.2.1, 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 $\mathcal{M}_K^{est}$.[1]

3.3 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 $\varphi \rightarrow_i \psi$ as "under agent i's plausibility ordering, φ typically implies ψ".

We define a *(Kripke) structure (for knowledge and plausibility)* to be 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 ordering 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 $K_i\varphi$ are evaluated according to $\mathcal{K}_i$:

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

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

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

We now define beliefs. Recall that *true*$\rightarrow_i \varphi$ means that φ is more plausible than not 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*$\rightarrow_i \varphi$ while

[1]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.

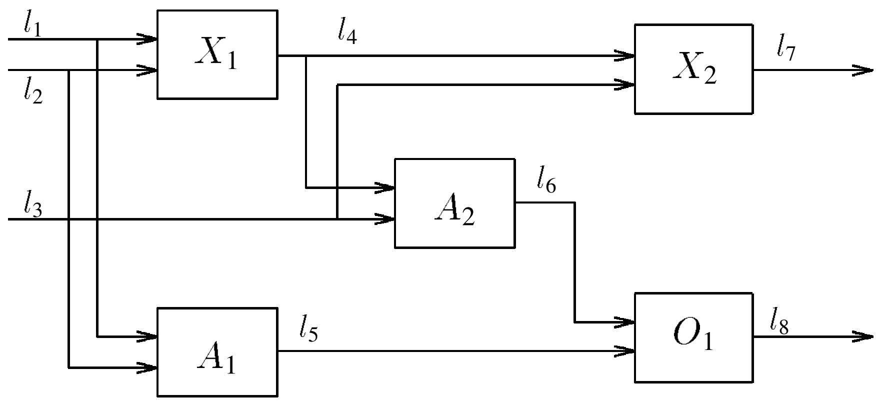

Figure 3.1: A full adder. X_1 and X_2 are XOR gates, A_1 and A_2 are AND gates, and O_1 is an OR gate.

$(M, w') \models \neg(\mathit{true} \rightarrow_i \varphi)$. That is, φ is plausible in one of the worlds the agent considers possible, but is not plausible 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 true in the most plausible worlds. Thus, we define $B_i\varphi$, read "agent i believes φ", as an abbreviation for $K_i(\mathit{true} \rightarrow_i \varphi)$. This means that the agent believes φ if φ is plausible in all the worlds he considers possible.

3.4 Example: Circuit Diagnosis

The following example illustrates some of the expressive power of this language.

The circuit diagnosis problem has been well studied in the literature (see [Davis and Hamscher 1988] 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 3.1.[2] The diagnosis task is to to identify which 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.

[2]The "full adder" example is classic in the diagnosis literature. In our discussion here we loosely follow the examples of Reiter [1987].

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 $value(w)$, the value of all the lines in the circuit. We consider only worlds where the values assigned to lines are consistent with the set of faulty components. For example, in the circuit of Figure 3.1, 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 $value(w)$ 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 $\langle l_i, 1\rangle \in value(w)$.

Next, we need to define the agent's knowledge. We define $o_w \subseteq value(w)$ 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 score.

Plausibility To model the agent's beliefs, we need to decide on the plausibility ordering 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 ordering 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 plausibility ordering as follows: If f_1 and f_2 are two failure sets, we say that f_1 is more plausible than f_2 if $|f_1| < |f_2|$, that is, if f_1 consists of fewer faulty components than f_2. This definition leads to a preference ordering over failure sets. This preference ordering induces a plausibility measure using the construction of Theorem 2.4.2. 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 $(\Pr_0, \Pr_1, \ldots)$, where in $\Pr_m$ the probability of a single failure is $1/(m+1)$. It is not hard to check that $\lim_{m\to\infty} \Pr_m(F_2)/\Pr_m(F_1) = 0$ if and only if $\text{Pl}(F_2) < \text{Pl}(F_1)$ 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 $|f_1| = |f_2|$, $\text{Pl}(\{f_1\})$ is incomparable to $\text{Pl}(\{f_2\})$ 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 $I\!N^m$ to $I\!N$. If $\vec{m} = (m_1, \ldots, m_n)$, let $\Pr_{g(\vec{m})}$ 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 $\lim_{m\to\infty} \Pr_m(f_2)/\Pr_m(f_1) = 0$ if 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 Theorem 2.4.2, we can again consider the plausibility measure Pl induced by $\prec$. It is not hard to see that $\text{Pl}(F_1) \leq \text{Pl}(F_2)$ 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 orderings we just described, we can construct two possible structures $M_{\mathit{diag},1}$ and $M_{\mathit{diag},2}$. In both structures we set $W_{(w,1)} = \mathcal{K}_{\mathit{diag}}(w)$, and in both $M_{\mathit{diag},1}$ and $M_{\mathit{diag},2}$ the plausibility measure is induced from a preference ordering on failures (using the construction of Theorem 2.4.2). In $M_{\mathit{diag},1}$ we set the plausibility ordering so that $\text{Pl}_{(w,1)}(\{w\}) \geq \text{Pl}_{(w,1)}(\{w'\})$ if and only if $|\mathit{fault}(w)| \leq |\mathit{fault}(w')|$, and in $M_{\mathit{diag},2}$ so that $\text{Pl}_{(w,1)}(\{w\}) \geq \text{Pl}_{(w,1)}(\{w'\})$ 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 $\text{Bel}(M, w) = \{f : (M, w) \models \neg B_1 \neg D_f\}$. 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 $o_w = o$.[3]

[3]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.

Proposition 3.4.1: *(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.

Thus, both Bel$(M_{diag,1}, w)$ and 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 [de Kleer 1990], while $M_{diag,2}$ captures the assumptions made in [Reiter 1987]. More concretely, in our example, if the agent observes $hi(l_1) \land \neg hi(l_2) \land hi(l_3) \land hi(l_7) \land \neg hi(l_8)$, 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 $\{X_2, A_2\}$.

3.5 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 [1994]. The only difference is that in Kripke structures for knowledge and probability, $\mathcal{P}_i(w)$ is a probability space rather than a 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. We also examine some conditions on plausibility orderings without regard to the knowledge, which we discussed in Section 2.8.

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 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 [Fagin and Halpern 1994]):

CONS $W_{(w,i)} \subseteq \mathcal{K}_i(w)$ for all worlds w and all agents i.[4]

[4]We remark that CONS is inappropriate if we use $\rightarrow$ to model, not plausibility, but counterfactual conditions, as is done by Lewis [1973]. If CONS holds, then it is easy to see that $K_i\varphi \Rightarrow K_i(\neg\varphi \rightarrow_i \psi)$ is valid, for all ψ. That is, if agent i knows φ, then he knows that in the most plausible worlds where $\neg\varphi$ is true, ψ is vacuously true, because there are no plausible worlds where $\neg\varphi$ 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."

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.)

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 $\top = \bot$ in plausibility spaces. If this happens, the agent considers all sets to be completely implausible.

NORM $\mathcal{P}(w, i)$ is normal, that is, $\top_{(w,i)} > \bot_{(w,i)}$, 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 requiring that $\mathrm{Pl}_{(w,i)}$ is reflexive. The definition of reflexivity from Section 2.8 is easily restated for structures of knowledge and plausibility:

REF For all worlds w and all agents i,

- $w \in W_{(w,i)}$, and
- $\mathrm{Pl}_{(w,i)}(A) > \bot$ 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 ordering possible. This amounts to assuming that the plausibility ordering is a function of the agent's epistemic state. This is captured by an assumption called SDP (following [Fagin and Halpern 1994]) 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 ordering. In particular, as we shall see, with SDP we have that $\varphi \rightarrow_i \psi$ implies $K_i(\varphi \rightarrow_i \psi)$.

It is easy to verify that the structures described in the diagnosis example of Section 3.4 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 orderings. To capture this, we need to generalize SDP. The following example might help motivate the formal definition.

Example 3.5.1: 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.

We can capture this by partitioning the worlds Alice considers possible into two sets, according to Bob's tribe. Alice has a plausibility ordering on the set W_R of worlds she considers possible where Bob is a Rightfoot and another on the set W_L of worlds she considers possible where Bob is a Leftfoot. For each $w' \in W_j$, $j \in \{L, R\}$, we have $W_{(w',Alice)} = W_j$ and for each $w'' \in W_j$, we have $\mathcal{P}_{Alice}(w') = \mathcal{P}_{Alice}(w'')$. In such a structure, the formula $\neg K_{Alice} \neg(tell(\varphi) \rightarrow_{Alice} \neg\varphi) \wedge$

$\neg K_{Alice}\neg(tell(\varphi) \rightarrow_{Alice} \varphi)$ is satisfiable, where $tell(\varphi)$ is the formula that holds when Bob tells Alice φ. On the other hand, in structures satisfying SDP this formula is satisfiable only when $tell(\varphi)$ is false in all the worlds Alice considers plausible. ∎

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 4.2), 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).

This example motivates suggest using *uniform* structures, as defined in Section 2.8.

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

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 any specific plausibility structure can compare all possible worlds, 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 $\mathrm{Pl}_w(A \cup B) = \max(\mathrm{Pl}_w(A), \mathrm{Pl}_w(B))$.

While rankings are quite natural, they have often been rejected as being too inexpressive [Ginsberg 1986]. 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. If the plausibility ordering 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.

3.6 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 $\mathcal{M}^{\text{CONS}}$ (resp. $\mathcal{M}^{\text{CONS,NORM}}$) be the structures satisfying QUAL and CONS (resp. QUAL, CONS and NORM).

Work on belief and knowledge in the literature [Halpern and Moses 1992; Hintikka 1962; Levesque 1984] has focused on the modal systems S5, KD45, K45, and K with semantics based on Kripke structures as described in Section 3.2. 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 $(M, w) \models B_i\varphi$ if and only if $(M, v) \models \varphi$ for all $v \in \mathcal{B}_i(w)$? We show that this is possible in some structures, but not in general.

Let $S = (W, \text{Pl})$ be a qualitative plausibility space. We say that $A \subseteq W$ is a set of *most plausible* worlds if $\text{Pl}(A) > \text{Pl}(\overline{A})$ (where $\overline{A}$ is the complement of A, i.e., $W - A$) and for all $B \subset A$, $\text{Pl}(B) \not> \text{Pl}(\overline{B})$. 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 $\text{Pl}(A \cap A') > \text{Pl}(\overline{A \cap A'})$. Since A and A' are both most plausible sets of worlds, it follows that we must have $A = A'$. Suppose that $\text{Pl}(A) > \text{Pl}(\overline{A})$ and $\text{Pl}(A') > \text{Pl}(\overline{A'})$. Note that $A \cap A'$, $A - A'$ and $\overline{A}$ are pairwise disjoint. Now $\text{Pl}((A \cap A') \cup (A - A')) = \text{Pl}(A) > \text{Pl}(\overline{A})$, and $\text{Pl}((A \cap A') \cup \overline{A}) \geq \text{Pl}((A \cap A') \cup (A' - A)) = \text{Pl}(A') > \text{Pl}(\overline{A'}) \geq \text{Pl}(A - A')$. We can apply A2 to get that $\text{Pl}(A \cap A') > \text{Pl}((A - A') \cup \overline{A}) = \text{Pl}(\overline{A \cap A'})$.

In general, however, a set of most plausible worlds does not necessarily exist. For example, consider the space $S_0 = (W, \text{Pl})$, where $W = \{w_i : i \geq 0\}$ and Pl is defined as follows: $\text{Pl}(A) = \infty$ if A contains an infinite number of worlds, and $\text{Pl}(A) = \max_{w_i \in A}(i)$ otherwise. Suppose that $\text{Pl}(A) > \text{Pl}(\overline{A})$. $\overline{A}$ must be finite, for otherwise $\text{Pl}(\overline{A}) = \infty$. 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, $\text{Pl}(A - \{w_i\}) > \text{Pl}(\overline{A - \{w_i\}})$. 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 $\pi(w_i)(p_j) = \textbf{true}$ 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 $(M, w_0) \models B_1\varphi$ if and only if $[\![\neg\varphi]\!]_{(w_0,i)}$ is a finite set, i.e., there is an index i such that for all $j \geq i$, we have $(M, w_j) \models \varphi$. 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 MP(S) to be the set of most plausible worlds in S if it exists, and $\emptyset$ if $\mathrm{Pl}(W) = \perp$. Otherwise MP(S) is not defined.

Proposition 3.6.1: *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 $(M, w) \models B_i\varphi$ if and only if $(M, w'') \models \varphi$ for all $w'' \in \cup_{w' \in \mathcal{K}_i(w)}MP(\mathcal{P}_i(w'))$.*

This 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 3.7.3 below). It is easy to verify that in a finite model MP$(\mathcal{P}_i(w))$ is always defined.[5]

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

We now consider knowledge and belief together. This combination has been investigated in the literature [Kraus and Lehmann 1988; Voorbraak 1992]. In particular, Kraus and Lehmann [1988] 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. $B_i\varphi \Rightarrow K_iB_i\varphi$

KB2. $K_i\varphi \Rightarrow B_i\varphi$

[5] We note, however, that this finite model property is no longer true in the first-order case. In this case, the two models of beliefs are not equivalent; see [Friedman, Halpern, and Koller 1996].

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

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

Theorem 3.6.3: *AX$^{\mathrm{KB}}$ (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., $\mathcal{M}^{\mathrm{CONS}}$, $\mathcal{M}^{\mathrm{CONS,NORM}}$).*

As a corollary, we can show that there is a close relationship between our framework and that of [Kraus and Lehmann 1988]. Let KL be the logic of Kraus and Lehmann:

Corollary 3.6.4: *For any $\varphi \in \mathcal{L}^{KB}$, $KL \models \varphi$ if and only if $\mathcal{M}^{\mathrm{CONS,NORM}} \models \varphi$.*

We now relate to three other notions of beliefs in the literature—those of Moses and Shoham [1993], Voorbraak [1992], and Lamarre and Shoham [1994].

Moses and Shoham [1993] 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 $B_i^{\alpha}\varphi$ to be an abbreviation for $K_i(\alpha \Rightarrow \varphi)$.[6] The following result relates our notion of belief to that of Moses and Shoham.

Lemma 3.6.5: *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 $\varphi \in \mathcal{L}^{KB}$ that includes only the modalities K_i and B_i, $(M,w) \models \varphi$ if and only if $(M,w) \models \varphi^*$, where φ^* is the result of recursively replacing each subformula of the form $B_i\psi$ in φ by $K_i(\alpha \Rightarrow \psi^*)$.*

Voorbraak [1992] 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 $K_i\varphi \Rightarrow B_i\varphi$, and suggests $B_i\varphi \Rightarrow B_iK_i\varphi$. The difference lies in

[6]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.

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 [1994] 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 $B^\alpha\varphi$ to read "given evidence α, φ holds in the most plausible α-worlds". Their formal account of $B^\alpha\varphi$ is exactly $\alpha \rightarrow_i \varphi$ in our notation. Unlike 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.

3.7 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 [Fagin and Halpern 1994] 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 [Lewis 1973] and [Burgess 1981] (see Section 2.8), while the axioms for CONS and SDP (and also UNIF) correspond directly to the axioms suggested in [Fagin and Halpern 1994] 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 [Halpern and Moses 1992] and for conditionals [Friedman and Halpern 1996b].

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 3.2); the component for reasoning about knowledge consists of K2–K5 and RK2 (from Section 3.2); the component for reasoning about conditionals consists of the standard axioms and rules for conditional logic C1–C4, RC1, and RC2 described in Section 2.8. Let AX consist of K1–K5, C1–C4, RK1, RK2, RC1, and RC2.

Theorem 3.7.1: *AX is a sound and complete axiomatization for $\mathcal{L}^{KC}$ with respect to $\mathcal{M}$.*

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, described in Section 2.8. CONS and SDP correspond to the following axioms, respectively;

C9. $K_i\varphi \Rightarrow N_i\varphi$

C10. $(\varphi \rightarrow_i \psi) \Rightarrow K_i(\varphi \rightarrow_i \psi)$

It is interesting to note that the axioms for CONS and UNIF are derived from the axioms defined in [Fagin and Halpern 1994] by replacing $w(\varphi) = 1$ (the probability of φ is 1) by $N_i\varphi$, 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 3.7.2: *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 $\mathcal{A}$.*

We now consider the complexity of the validity problem. Our results are based on a combination of results for complexity of epistemic logics [Halpern and Moses 1992] and conditional logics [Friedman and Halpern 1996b]. Again, the technical details are much in the spirit of those in [Fagin and Halpern 1994].

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 $\mathit{Sub}(\varphi)$ be the set of subformulas in φ. It is easy to see that an upper bound on $|\mathit{Sub}(\varphi)|$ is the number of symbols in φ.

Theorem 3.7.3: *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 $2^{|\mathit{Sub}(\varphi)|}$ worlds.*

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 preferential ordering $\prec_{(w,i)}$ on $W_{(w,i)}$ that induces $\mathrm{Pl}_{(w,i)}$ using the construction of Theorem 2.4.2. Recall that a preferential 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 preferential orderings $\preceq_{(w,i)}$ 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 [Friedman and Halpern 1996b], we immediately get the following lemma:

Lemma 3.7.4: *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 $|\mathrm{Sub}(\varphi)|N$ worlds.*

Combining this with Theorem 3.7.3, 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 [Halpern and Moses 1992; Friedman and Halpern 1996b]). However, if there is only one agent and we assume CONS and either UNIF or SDP, then we can get polynomial-sized models.

Theorem 3.7.5: *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 $|\mathrm{Sub}(\varphi)|^3$ worlds.*

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 $|\varphi|$.

Theorem 3.7.6: *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.*

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 3.7.7: *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.*

3.8 Discussion

In this chapter we have proposed a framework that combines knowledge and plausibility (and hence beliefs), and investigated a number of properties of the framework, such as complete axiomatization for various sublanguages and various properties of the relationships between the modal operators.

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, 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 work [Friedman and Halpern 1997a], 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., [Boutilier 1992]), but also to approaches based on κ-rankings (e.g., [Goldszmidt and Pearl 1992]), probably measures (e.g., [Fagin and Halpern 1994]), or any other measure of uncertainty. Finally, plausibility measures have the advantage of greater expressive power than other approaches.

As we shall see in the next chapters, once we introduce time into the framework, plausibility allows us to construct a natural mechanism for minimal change in beliefs.

Chapter 4

Adding Time

4.1 Introduction

In the previous chapter, 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 chapter, 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.

4.2 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, [Kraus and Lehmann 1988]). We introduce more structure into the description by adopting the framework of Halpern and Fagin [1989] 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.[1]

This definition of knowledge has proved useful in many applications in distributed systems and AI (see [Fagin, Halpern, Moses, and Vardi 1995] 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 [Gabbay, Pnueli, Shelah, and Stavi 1980]). These include $\bigcirc\varphi$ 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

- $(\mathcal{I}, r, m) \models \bigcirc\varphi$ if $(\mathcal{I}, r, m+1) \models \varphi$.[2]

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

[1] It is straightforward to extend these definitions to deal with continuous time. This is done, for example, in [Brafman, Latombe, Moses, and Shoham 1994].

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

4.3 Example: Circuit Diagnosis Revisited

We now show how the framework can be used to extend the example of Section 3.4 to incorporate time, allowing the agent to perform a sequence of tests.

We want to model diagnosis as 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 3.4 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 3.4. 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 4.3.1: *If there is some $f \in \mathit{Bel}(\mathcal{I}_{diag,1}, r, m)$ that is consistent with the new observation $o_{(r,m+1)}$, then $\mathit{Bel}(\mathcal{I}_{diag,1}, r, m+1)$ consists of all the failure sets in $\mathit{Bel}(\mathcal{I}_{diag,1}, r, m)$ 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 $o_{(r,m+1)}$.*

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 4.3.1 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 $(\mathcal{I}, r', m') \models \varphi$ 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 $(\mathcal{I}_{diag,1}, r, m) \models B_i \neg \bigcirc \varphi$.

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 4.3.2: *$Bel(\mathcal{I}_{diag,2}, r, m+1)$ 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 $o_{(r,m+1)}$.*

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 [1987, Section 5].

4.4 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 3.7, 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 3.7 and the following axioms and rule the describe the properties of $\bigcirc$:.

T1. $\bigcirc\varphi \wedge \bigcirc(\varphi \Rightarrow \psi) \Rightarrow \bigcirc\psi$

T2. $\bigcirc\varphi \equiv \neg\bigcirc\neg\varphi$

RT1. From φ infer $\bigcirc\varphi$.

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

Theorem 4.4.1: *The axiom system AX^T is a sound and complete axiomatization of $\mathcal{L}^{KCT}$ with respect to $\mathcal{C}$.*

We can also prove a result analogous to Theorem 3.7.2 that describes a complete axiomatization for the classes of systems satisfying some of the assumptions we examined in Chapter 3.

Theorem 4.4.2: *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 $\mathcal{A}$.*

4.5 Prior Plausibilities

The formal framework of knowledge and plausibility 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.[3]

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 [De Finetti 1972; Ramsey 1931], 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:

$$\Pr(A|E) \le \Pr(B|E) \text{ if and only if } \Pr(A \cap E) \le \Pr(B \cap E).$$

We want the analogous property for plausibility:

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

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 Chapter 6 and [Friedman and Halpern 1995] 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

[3] 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 $\Pr'(E) = 1$ [Kullback and Leibler 1951]. Indeed, this holds true for other distance measures as well [Diaconis and Zabell 1982].

$A, B \subseteq W_1$, 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 [Friedman and Halpern 1995]).

This discussion suggests that we define $\mathrm{Pl}_{(r,m+1,i)}$ to be the result of conditioning $\mathrm{Pl}_{(r,m,i)}$ 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 [Halpern and Tuttle 1993] 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* [Halpern and Vardi 1989]. 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 ordering $\mathcal{P}_i(r,m)$ 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). This is process is shown in Figure 4.1. 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

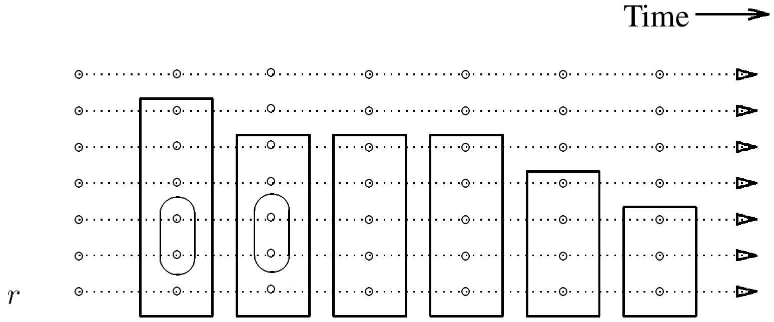

Figure 4.1: Schematic description of how the agent's knowledge evolves in time in synchronous systems with perfect recall. The boxes represent the set of points in $\mathcal{K}_i(r, m)$. Since the system is synchronous, at each time point, the agent consider possible points at the same time. Since the 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.

any conditional plausibility measure that results from conditioning on E.[4]

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 $\mathcal{R}_{(r,i)} \subseteq \mathcal{R}$. 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 4.2. The main complications are due to the transition back and forth between entities defined over runs and ones defined over points.

[4]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 $\mathrm{Pl}(\overline{A}) = \perp$, 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.

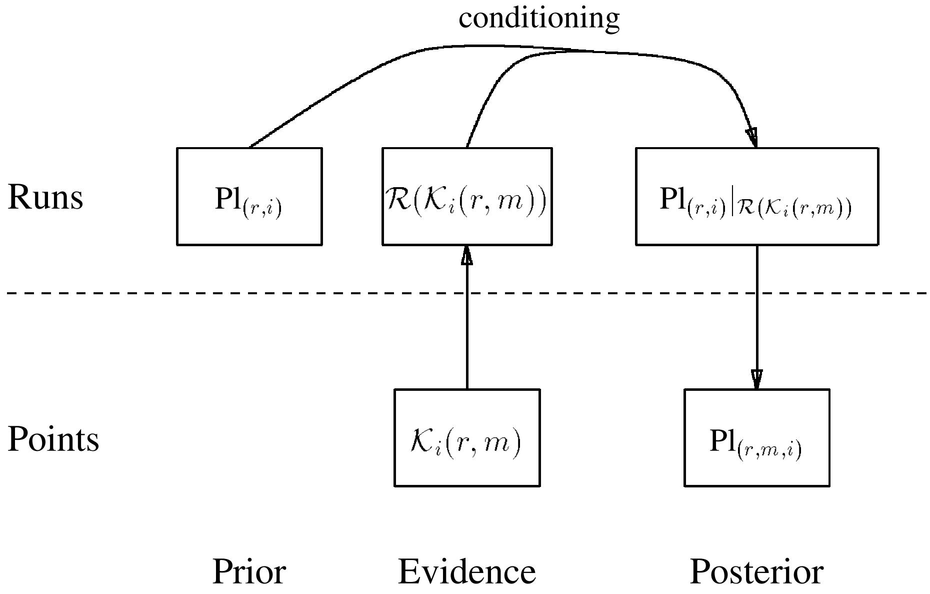

Figure 4.2: Schematic description of the entities involved in the definition of priors. Note some are defined over runs and some over points.

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 4.5.1: *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* $\mathrm{Pl}_{(r,m+1)}(A) \le \mathrm{Pl}_{(r,m+1)}(B)$ *if and only if* $\mathrm{Pl}_{(r,m)}(\mathit{prev}(A)) \le \mathrm{Pl}_{(r,m)}(\mathit{prev}(B))$*, for all runs r, times m, and sets $A, B \in W_{(r,m+1)}$.*

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 4.5.2: It is easy to verify that the two systems we consider in Section 4.3 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 orderings, we have reduced the question of where the plausibility ordering 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.[5] 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 ($\neg\varphi$) of measure 0. The plausibility approach has no problem with this: the conditioning process described above still makes perfect sense.

4.6 Conditioning as Minimal Change of Belief

In this section we examine the properties of conditioning as an approach to minimal change of beliefs. We show that in some cases we can reason about the agent's beliefs in the language. Then we relate our approach others in the literature.

We start by examining an alternative justification for qualitative plausibility measures. Recall that qualitative plausibility measures are necessary in order to get the KLM properties of conditionals. Should we restrict ourselves to qualitative measures if we care only about an agent's beliefs? As we now show, the answer is yes.

Recall that one of the standard properties of beliefs is that it closed under logical implications. This property was captured by the axiom K2. We now show that in a system satisfying PRIOR, agents' beliefs are closed under logical implications if and only if their priors are all qualitative.

Theorem 4.6.1: *Let $\mathcal{I}$ be a synchronous system satisfying and PRIOR. The axiom K2 is valid in $\mathcal{I}$ with respect to the B_i modal operator if and only if the prior $\mathrm{Pl}_{(r,i)}$ satisfies A2 for all runs r and agents i.*

In view of this result, we focus from now on only on systems with qualitative prior.

Next, we examine how changes in beliefs are determined by the prior. Using Proposition 4.5.1, 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 to his knowledge at (r, m) if, for all $(r', m) \in \mathcal{K}_i(r, m)$, we have $(r', m+1) \models \varphi$ 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.

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

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 Chapter 5.) In such systems, we can characterize within the agent's belief change process in the language.

Proposition 4.6.2: *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 $(\mathcal{I}, r, m) \models \bigcirc(\varphi \wedge \psi) {\rightarrow_i} \bigcirc\xi$.*

Corollary 4.6.3: *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 $(\mathcal{I}, r, m) \models K_i(\bigcirc\varphi \Rightarrow (\bigcirc\varphi {\rightarrow_i} \bigcirc\psi))$. Moreover, if $\mathcal{I}$ also satisfies SDP, then $(\mathcal{I}, r, m+1) \models B_i\psi$ if and only if $(\mathcal{I}, r, m) \models \bigcirc\varphi {\rightarrow_i} \bigcirc\psi$.*

We now use this result to relate our approach to other approaches for modeling conditionals in the literature. Boutilier [1992], Goldszmidt and Pearl [1992], and Lamarre and Shoham [1994] give conditional statements similar semantics (using a preferential ordering), but $\varphi {\rightarrow} \psi$ 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.[6] 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 $\varphi \equiv \bigcirc\varphi$. These two assumptions lead to a characterization of belief change.

Corollary 4.6.4: *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 $(\mathcal{I}, r, m) \models \varphi {\rightarrow_i} \psi$.*

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 $\varphi {\rightarrow_i} \psi$) and "after learning φ, ψ is believed" (which is how it is interpreted in [Boutilier 1992; Goldszmidt and Pearl 1992; Lamarre and Shoham 1994]), the two readings are in general quite different. For one thing, notice that Corollary 4.6.4 assumes that φ and ψ are propositional formulas. This is a necessary assumption. If φ and ψ contain modal formulas, then $\varphi {\rightarrow} \psi$ 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 $(\mathcal{I}, r, m) \models \varphi {\rightarrow_i} B_i\psi$, regardless of whether $B_i\psi$ is believed at $(r, m+1)$. In [Friedman and Halpern 1994a], we examine conditionals of the form $\varphi > \psi$ 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

[6] 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.

would like agents to be persistent in their assessment, which is exactly what conditioning captures. In the presence of several assumptions, we get an 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 [1997]. They consider a logic of knowledge, belief and time, and attempt to capture properties of "minimal change" of beliefs. Their language is slightly different than ours. Instead of introducing a temporal modality, they define a different belief and knowledge modality for each time step: $B^t\varphi$ 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 3.6 to characterize the connection between knowledge and belief) that also includes two additional axioms that can be written in our language as

BT1. $B_i \bigcirc B_i\varphi \Rightarrow B_i\varphi$

BT2. $B_i\varphi \Rightarrow B_i \bigcirc B_i\varphi$

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. The minimal assumption is as follows. We say that 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 an agent can make at each point).

Lemma 4.6.5: *Let $\mathcal{I}$ be a synchronous static system satisfying PRIOR, RANK, SDP, and perfect recall that has finite branching. Then $(\mathcal{I}, r, m) \models B_i\varphi \equiv B_i \bigcirc B_i\varphi$ for all propositional formulas φ.*

Are these conditions necessary to characterize BT1 and BT2? The answer is no. First, the proof of Lemma 4.6.5 applies to systems with infinite branching, if the agents' prior satisfies an infinitary version of A2. As shown in [Friedman, Halpern, and Koller 1996], this infinitary version is satisfied by κ rankings and well-founded preference orderings. Thus, any system whose prior is a well-founded ranking and with static propositions satisfies BT1 and BT2. Additionally, these axioms do not necessarily characterize ranking, 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 and ranking. In particular, any system with a well-founded ranked prior and static propositions satisfies BT1 and BT2. As we shall see in the next section, similar requirements are made by the AGM formulation of belief revision [Alchourrón, Gärdenfors, and Makinson 1985].

4.7 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 4.7.1: *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 $\mathcal{I}$.*

Thus by constructing priors that satisfy various properties, we can ensure that the resulting system also satisfies them. In particular, Proposition 4.7.1 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 ordering (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 [1982] 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 $\mathrm{Pr}(A) = 0$ implies that $\mathrm{Pr}'(A) = 0$, 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 $\mathrm{Pr}''(A \times X) = \mathrm{Pr}(A)$ (so that Pr$''$ can be viewed as an extension of Pr) and $\mathrm{Pr}'(A) = \mathrm{Pr}''(A \times X | E)$.

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. $N_i \bigcirc \varphi \Rightarrow \bigcirc N_i \varphi$

Proposition 4.7.2: *If $\mathcal{I}$ is a synchronous and coherent system, then COH is valid in $\mathcal{I}$.*

There is a sense in which the converse to Proposition 4.7.2 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.[7]

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

Proposition 4.7.3: *If $\mathcal{I}$ is a synchronous system satisfying perfect recall and PRIOR, then $\mathcal{I}$ is coherent.*

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 $\varphi \in \mathcal{L}^{KCT}$ 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 $(\varphi \rightarrow_i \psi) \Rightarrow \bigcirc B_i \psi$ is temporally linear, while $K_i(\bigcirc\varphi \rightarrow_i \bigcirc\psi) \Rightarrow \bigcirc B_i \psi$ 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 4.7.4: *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 $\varphi \in \mathcal{L}^{KCT}$, we have $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', f(r), m) \models \varphi$.*[8]

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 Chapter 5 these postulates can be captured by systems with the appropriate prior plausibility.

[7] We remark that COH is analogous to the axiom $K_i \bigcirc \varphi \Rightarrow \bigcirc K_i \varphi$ that characterizes perfect recall in synchronous systems [Fagin, Halpern, Moses, and Vardi 1995]. Roughly speaking, this is because coherence ensures that the agent does not forget what he ruled out as implausible.

[8] 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).

Can we extend Theorem 4.7.4 to the full language? We conjecture that Theorem 4.7.4 actually holds for all $\varphi \in \mathcal{L}^{KCT}$, 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 C5, we do not get any new properties by assuming PRIOR.

Note that the construction described by Theorem 4.7.4 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 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 4.7.5: *If $\mathcal{I}$ is a synchronous system that satisfies perfect recall and either PRIOR and UNIF, or SDP, then $\mathcal{I}$ satisfies PERSIST.*

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 4.7.4.

Theorem 4.7.6: *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 $\varphi \in \mathcal{L}^{KCT}$, $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', f(r), m) \models \varphi$.*

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 4.5.1 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 $m+1$.[9] 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

[9]We could, of course, redefine PRIOR so as to guarantee that Proposition 4.5.1 holds, but this leads to other complications.

	c	d
c	(-1,-1)	(-10,0)
d	(-10,0)	(-5,-5)

Figure 4.3: The prisoner's dilemma: each player can either defect (d) or cooperate (c).

a probabilistic setting in [Piccione and Rubinstein 1996], in 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.

4.8 Example: Repeated Games

Prisoner's dilemma is a well-known game where each player can either defect or cooperate. The motivating story is about two robbers how are caught by the police near the scene of a crime. They are then interrogated in separate rooms, and each is given the choice of *defecting*, that is, telling the police the facts of the robbery, or *cooperating*, that is, sticking with the cover story that they are innocent bystander. They are told that the outcome will be as follows. If they both maintain consistent cover story, then they will be booked for a minor offense (payoff of -1 for both). If one provides the police evidence about the other, then he will go free (payoff 0) and the other will be convicted for a long jail term (payoff -10). Finally, if both confess, then they will both be convicted for a lesser jail term (payoff of -5). It is easy to see that each player maximizes his payoff by defecting no matter what the other player does, but the utility when both cooperate is higher than when both defect.

In general, we get this phenomena if taking u_j to denote player j's utility assignment, then $u_1(c,d) < u_1(d,d) < u_1(c,c) < u_1(d,c)$ (where (c,d) denotes the situation where player 1 cooperates and player 2 defects, and so on); u_2 is symmetric, i.e., $u_2(x,y) = u_1(y,x)$, so $u_2(d,c) < u_2(d,d) < u_2(c,c) < u_2(c,d)$. If we assume that a *rational* player is one that chooses a dominating strategy if there is one, then it is clear that rational players playing a one-shot prisoner's dilemma both defect.

There is a standard backwards induction argument that shows that in any finitely-repeated prisoner's dilemma, if rationality is initially *common knowledge* (so that the players both know that the players both know ... that each of them is rational), then the players must play the unique Nash equilibrium solution, which is to always defect. This argument goes as follows. Consider the last game of the sequence. Since it is the last game, agents' choices only effect the outcome of this game. Thus, this game can be treated as a one-shot game, and thus, since both are rational players, they will both defect in that last move. Moreover, since there is common knowledge of rationality, there is also common knowledge that both players will defect in the last game. Now assume, by induction, that it is common knowledge that both agent will defect in the last k games. This implies that the agents' choices in the k from last last game would not effect the way the last k games are played. Hence, both players will also defect in that game. By induction we conclude that both agent will always defect.

This seems somewhat paradoxical. How do we account for the fact that rational players do much worse in repeated prisoner's dilemma than supposedly irrational players who cooperate? There has been a great deal of effort in the game-theoretic literature to construct models of prisoner's dilemma where rational players can cooperate (see [Kreps, Milgrom, Roberts, and Wilson 1982] for one of the best-known examples).

More recently, there has been intense scrutiny of the assumption of common knowledge of rationality. Indeed, it has been argued variously (a) that common knowledge of rationality is an inconsistent assumption [Bicchieri 1989; Reny 1992], (b) that it is consistent and it indeed implies the backwards induction solution [Aumann 1993], and (c) that, while consistent, it does not necessarily imply the backwards induction solution [Ben-Porath 1992; Stalnaker 1992]. While a comparison of these arguments is beyond the scope of this example, we note that the subtleties typically arise when knowledge is treated as "believed to hold with probability 1" (as is often the case in the game theory literature). Under this definition, it is not clear how a player should update her beliefs if the other player cooperates on the first move. The problem is that if we take the standard Bayesian approach of conditioning, then we are conditioning on a measure 0 event.

As we mentioned above, if we use our definition of belief, we avoid the problem of conditioning on measure 0 events. As we now show, under our definitions, while initial common *belief* of rationality is consistent with the backwards induction argument, it does not imply it. On the other hand, if we assume perfect recall (so that the players do not forget what they have seen), then common *knowledge* of rationality does imply the backwards induction argument (using our definition of knowledge as truth in all accessible worlds). Thus, the distinction between knowledge and belief in our framework plays a crucial role here.

We construct a system that distinguishes the knowledge and belief of the players during the game. This construction is similar in spirit that of Ben-Porath [1992] and Stalnaker [1992]. We assume two players are playing n iterations of prisoners dilemma. We model this game as a a two agent system, where each player is an agent. Recall that a strategy for player i is a function that returns a move given a sequence of moves (intuitively the ones that have been made up to that time). Since the moves in our case are cooperate and defect, we can view a strategy as a function from a string of c's and d's to $\{c, d\}$. (We allow the empty string as an argument; this represents the initial move.) Each player has a *type* that determines the strategy played by the player and a assessment of the plausibility of the the opponent's types. This ranking describes the player's prior beliefs about his opponent. Formally, we consider a set T of types. With each $t \in T$ we associate a strategy s_t and a plausibility measure over over T. We assume that a player's type does not change throughout a run; let $type(r, i)$ denote player i's type in run r. A player's local state at time m contains the player's type and the moves taken up to time m. This means that players know their own type and have perfect recall of the game. The set $\mathcal{R}$ of runs consists of all the runs where both players observe the same history and act according to their strategies. We define $\mathcal{P}_i(r, m)$ according to player i's type: $\mathrm{Pl}_{(r,m,i)}(A) \geq \mathrm{Pl}_{(r,m,i)}(B)$ if $\mathrm{Pl}_{type(r,i)}(\{type(r', j) : \exists m'((r', m') \in A\})) \geq \mathrm{Pl}_{type(r,i)}(\{type(r', j) : \exists m'((r', m') \in B)\})$ where j is i's opponent. It is easy to verify that this system satisfies PRIOR, REF and SDP.

Let $\mathcal{S}$ be the set of possible strategies in the game, and for each $s \in \mathcal{S}$ let the proposition $p_{i,s}$ denote that player i plays strategy s. Using these propositions we can analyze the player's beliefs about the opponent's strategy. Let $S_i(r, m)$ denote the set of strategies player i believes her opponent may be using. Formally, $S_i(r, m) = \{s \in \mathcal{S} : (r, m) \models \neg B_i \neg p_{j,s}\}$, where j is i's opponent. These beliefs change from step to step: If the opponent's move at step m is consistent with i's beliefs (i.e., if there are strategies in $\mathcal{S}_i(r, m)$ consistent with the observed move), then $\mathcal{S}_i(r, m+1)$ is just the subset of $\mathcal{S}_i(r, m)$ consisting of all the strategies consistent with the observed move. If the opponent's move is inconsistent with i's beliefs, then $\mathcal{S}_i(r, m+1)$ is determined by i's prior ranking, but there is no necessary connection between $\mathcal{S}_i(r, m)$ and $\mathcal{S}_i(r, m+1)$. Intuitively, if the opponent's move is inconsistent with i's beliefs, then i is surprised; this case essentially corresponds to the occurrence of an event of measure 0.

Now we can define rationality. Of course, we expect that a rational player would not use a strategy that is dominated by another one, assuming that the other player is using a strategy in $\mathcal{S}(r, 0)$. However, this condition is not sufficient, since the player's beliefs may change during the game and a strategy that was not dominated initially might be dominated at time m with respect to the player's beliefs at m. We expect a rational player to always use a rational strategy. This motivates the following definition: If x is a string of moves and s is a strategy, then s^x is the strategy such that $s^x(y) = s(xy)$ (where xy is the concatenation of the moves in x and y). Let x_j be the sequence of moves that have been performed by player j up to the point (r, m) and let t be $type(r, i)$. We say that player i is *locally rational* at the point (r, m) if $s_t^{x_j}$ is not dominated by any strategy when played against strategies s^{x_i} for $s \in \mathcal{S}_i(r, m)$; i.e., there is no strategy that has at least as good a payoff as $s_t^{x_j}$ when played against s^{x_i} for each $s \in \mathcal{S}_i(r, m)$, and does better that $s_t^{x_j}$ when played against at least one strategy s^{x_i} for $s \in \mathcal{S}_i(r, m)$. We say that player i is *rational* at (r, m) if player i is locally rational at (r', m') for all $m' \geq m$ and for all r' such that $r(m) \sim_i r'(m)$.[10]

We now show that there are runs such that both players cooperate until the last $k+2$ rounds and it is common belief that both players are rational throughout the run, where k is the minimal natural number such that $k(u_1(c, c) - u_1(d, d)) > u_1(d, c) + u_1(c, d) - 2u_1(d, d)$. In particular, for appropriate choices of utilities, there are runs such that both players cooperate in all but the last two rounds and it is common belief that both players are rational throughout the run.

We start by defining three families of strategies s_l^h, where $h \in \{1, 2, 3\}$ and $l \leq n$. In these definitions, x and y represent (possible empty) strings of moves; c^k is the string composed of k c's, and $|x|$ is the length of the string x.

$$s_l^1(x) \quad = \quad \begin{cases} c & \text{if } x = c^j \text{ and either } 0 \leq j < \min(l, n-k-2) \text{ or } l < j < n-1 \text{, or} \\ & \text{if } x = c^j d c^{j'},\ 0 \leq j < \min(l, n-k-2) \text{, and } |x| < n-2 \\ d & \text{otherwise} \end{cases}$$

[10]This definition is similar to Stalnaker's definition [Stalnaker 1992], but has the advantage that it can be represented in our language without introducing counterfactuals. Indeed, it is straightforward (although tedious) to write a formula *rational*$_i$ in $\mathcal{L}^{KPT}$ such that $(\mathcal{I}, r, m) \models$ *rational*$_i$ if player i is rational at the point (r, m) in the interpreted plausibility system $\mathcal{I}$.

$$s_l^2(x) = \begin{cases} c & \text{if } x = c^j \text{ and } 0 \le j < \min(l, n-k-2) \text{ or} \\ & \text{if } x = c^j d c^{j'}, 0 \le j < \min(l, n-k-2), \text{ and } |x| < n-2 \\ d & \text{otherwise} \end{cases}$$

$$s_l^3(x) = \begin{cases} c & \text{if } x = yc^j \text{ and } |y| \le l \\ d & \text{otherwise} \end{cases}$$

Let t_l^h be types whose associated strategy is s_l^h; in the case that h is 1 or 2, we define the plausibility measure $\mathrm{Pl}_{t_l^h}$ below in such a way as to make a player of type t_l^h rational. We do not bother to define the plausibility measure for for t_l^3 since it is not rational. We define $\mathrm{Pl}_{t_l^h}$ to be a ranking, and thus it suffices to describe only the plausibility of singleton events. The plausibility for t_l^1 for agent i is such that if j is i's opponent, then

$$\top = \mathrm{Pl}_{t_l^1}(\{t_0^1\}) > \ldots \mathrm{Pl}_{t_l^1}(\{t_{l-1}^1\}) > \mathrm{Pl}_{t_l^1}(\{t_l^2\}) > \mathrm{Pl}_{t_l^1}(\{t_{l+1}^3\})$$

We require that all other types are considered less plausible than t_{l+1}^3. The plausibility measure associated with t_l^2 is similar, except that it does not include the type t_{l+1}^3:

$$\top = \mathrm{Pl}_{t_l^2}(\{t_0^1\}) > \ldots \mathrm{Pl}_{t_l^2}(\{t_{l-1}^2\}) > \mathrm{Pl}_{t_l^2}(\{t_l^2\})$$

Again, we require that all other types for j are considered less plausible than t_l^2.

Thus, if player 1 is of type t_l^1 then he starts out believing that player 2 is of type t_0^1. If player 2's first move is c, then player 1 is surprised. He revises his beliefs so that he now believes that player 2 is of type t_1^1. If he sees another c, he is again surprised, and now believes her type is t_2^1, and so on. The initial part of the plausibility ordering explains how revision works if a sequence of moves of the form $c^j d^k$ is seen. It does not explain what happens if a different sequence is seen. The missing details in the description of $\mathrm{Pl}_{t_l^h}$ can easily be arranged to ensure that if player 1 observed a sequence that does not correspond to any type in the beginning of the ordering (as defined above), then player 1 believes that player 2 will play d for the rest of the run.

It is easy to see that if r is a run where both players are of type t_{n-k-2}^1, then they will cooperate for the first $n-k-2$ rounds, and defect for the last $k+2$ rounds. Moreover, we can show that it is common belief that they are rational throughout the run. After m rounds where $m < n-k-2$, both believe that the other is of type t_m^1. Since a player of type t_m^1 believes that the other player is of type t_m^2, and a player of type t_m^2 believes that the other is also of type t_m^2, at this point common belief of rationality holds. After m rounds where $n-k-2 \le m \le n$ both players believe that the other is of type t_{n-k-2}^2. Since a player of type t_{n-k-2}^2 believes that the other is also of type t_{n-k-2}^2, we again have common belief of rationality. Thus, in all rounds a common belief of rationality holds.

We note that the only incentive for player 1 to cooperate in this example is the belief that it will surprise player 2, leading her to believe that he irrational. More precisely, at time $m < l$ player 1 of type t_l^1 believes that player 2 is of type t_m^1 and that after he will cooperate she will believe him to be of type t_{m+1}^3. Thus, cooperation is maintained because both players surprise each other at every round.

4.9 Discussion

In this chapter we have extended our framework to include time. The key development in this chapter is the notion of conditioning, which capture the intuition of minimal change. As we shall see in the next chapter, we can represent the standard approaches to minimal change—belief revision and belief update—in terms of conditioning.

Chapter 5

Capturing Revision and Update

5.1 Introduction

The framework we developed in the previous chapters allows us to represent a broad spectrum of notions of belief change. In this chapter, we show how belief revision and update can be represented. Doing this allows us to compare the assumptions implicit in each method and to understand the principles underlying them.

We are certainly not the first to provide semantic models for belief revision and update. For example, [Alchourrón, Gärdenfors, and Makinson 1985; Grove 1988; Gärdenfors and Makinson 1988; Rott 1991; Boutilier 1992; de Rijke 1992] deal with revision, and [Katsuno and Mendelzon 1991a; del Val and Shoham 1992] deal with update. In fact, there are several works in the literature that capture both using the same machinery [Katsuno and Satoh 1991; Goldszmidt and Pearl 1992], and others that simulate belief revision using belief update [Grahne, Mendelzon, and Rieter 1992; del Val and Shoham 1994]. Our approach is different from most in that we do not construct a specific framework to capture one or both of these belief change paradigms. Instead, we start from a natural framework to model how an agent's knowledge changes over time and add to it machinery that captures a defeasible notion of belief.

We believe that our representation offers a number of advantages, and gives insight into both revision and update. For one thing, we show that both revision and update can be viewed as proceeding by revision of initial prior plausibilities. Thus, our representation emphasizes the role of conditioning as a way of understanding *minimal change*. Moreover, it shows that that the major differences between revision and update can be understood as corresponding to differences in initial beliefs. For example, revision places full belief on the assumption that the propositions used to describe the world are *static*, and do not change their truth value over time.[1] By way of contrast,

[1] In the literature, belief revision has been described (by Katsuno and Mendelzon [1991a], for example) as a process of changing beliefs about a *static world*, but this is slightly misleading. In fact, what is important for revision is not that the world is static, but that the propositions used to describe the world are static. For example, "At time 0 the block is on the table" is a static proposition, while "The block is on the table" is not, since it implicitly references the current state of affairs. The assumption that the the propositions are static is not unique to belief revision. Bayesian updating, for

update allows for the possibility that propositions change their truth value over time. However, the family of prior plausibilities corresponding to update have the property that they prefer sequences of events where abnormal events occur as late as possible. The fact that time appears explicitly in our framework allows us to make this precise.

Finally, our representation makes it clear how the intuitions of revision and update can be applied in settings where the postulates that characterize them are not sound. For example, we consider situations where there may be irreversible changes (such as death, or breaking a glass vase), and where the agent may perform actions beyond just making observations. As we show, our framework allows us to extend belief revision and update naturally to handle such situations.

The rest of the chapter is organized as follows. In Section 5.2, we give a brief review of belief revision and belief update. In Section 5.3, we define a specific class of structures that embody assumptions that are common to both update and revision. In Section 5.4, we describe additional assumptions that are required to capture revision. In Section 5.5, we describe the assumptions that are required to capture update. In Section 5.6, we reexamine the differences and similarities between belief revision and update. In Section 5.7 we consider possible extensions to the setup of revision and update, and discuss how these extensions can be handled in our framework. Finally, in Section 5.8, we conclude with a discussion of related and future work.

5.2 Review of Revision and Update

We now present a brief review of belief revision and update.

Belief revision attempts to describe how a rational agent incorporates new beliefs. As we said earlier, the main intuition is that as few changes as possible should be made. Thus, when something is learned that is consistent with earlier beliefs, it is just added to the set of beliefs. The more interesting situation is when the agent learns something inconsistent with her current beliefs. She must then discard some of her old beliefs in order to incorporate the new belief and remain consistent. The question is which ones?

The most widely accepted notion of belief revision is defined by the AGM theory [Alchourrón, Gärdenfors, and Makinson 1985; Gärdenfors 1988]. This theory was originally developed in philosophy of science, where one attempts to understand when a scientist changes her beliefs (e.g., theory of physical laws) in a rational manner. In this context, it seems reasonable to assume that the world is *static*; that is, the laws of physics do not change while the scientist is performing experiments.

Formally, this theory assumes a logical language $\mathcal{L}_e$ over a set Φ_e of primitive propositions with a consequence relation $\vdash_{\mathcal{L}_e}$ that contains the propositional calculus and satisfies the deduction theorem. The agent's epistemic state is represented as a *belief set*—that is, a set of formulas in $\mathcal{L}_e$ closed under deduction. There is also assumed to be a revision operator $\circ$ that takes a belief set A and a formula φ and returns a new belief set $A \circ \varphi$, intuitively, the result of revising A by φ. The following AGM postulates are an attempt to characterize the intuition of "minimal change":

example, makes similar assumptions.

(R1) $A \circ \varphi$ is a belief set

(R2) $\varphi \in A \circ \varphi$

(R3) $A \circ \varphi \subseteq Cl(A \cup \{\varphi\})$[2]

(R4) If $\neg\varphi \notin A$ then $Cl(A \cup \{\varphi\}) \subseteq A \circ \varphi$

(R5) $A \circ \varphi = Cl(\mathit{false})$ if and only if $\vdash_L \neg\varphi$

(R6) If $\vdash_L \varphi \Leftrightarrow \psi$ then $A \circ \varphi = A \circ \psi$

(R7) $A \circ (\varphi \wedge \psi) \subseteq Cl(A \circ \varphi \cup \{\psi\})$

(R8) If $\neg\psi \notin A \circ \varphi$ then $Cl(A \circ \varphi \cup \{\psi\}) \subseteq A \circ (\varphi \wedge \psi)$.

The essence of these postulates is the following. After a revision by φ the belief set should include φ (postulates R1 and R2). If the new belief is consistent with the belief set, then the revision should not remove any of the old beliefs and should not add any new beliefs except these implied by the combination of the old beliefs with the new belief (postulates R3 and R4). This condition is called *persistence*. The next two conditions discuss the coherence of beliefs. Postulate R5 states that the agent is capable of incorporating any consistent belief and postulate R6 states that the syntactic form of the new belief does not affect the revision process. The last two postulates enforce a certain coherency on the outcome of revisions by related beliefs. Basically, they state that if ψ is consistent with $A \circ \varphi$ then $A \circ (\varphi \wedge \psi)$ is just $A \circ \varphi \circ \psi$.

The notion of *belief update* originated in the database community [Keller and Winslett 1985; Winslett 1988]. The problem is how a knowledge base should change when something is learned about the world. For example, suppose that a transaction adds to the knowledge base the fact "Table 7 is in Office 2", which contradicts the previous belief that "Table 7 is in Office 1". What else should change? The intuition that update attempts to capture is that such a transaction describes a change that has occurred in the world. Thus, in our example, update might reason that the table was moved from office 1 to office 2. This example shows that, unlike revision, update does not assume that the world is static.

Katsuno and Mendelzon [1991a] suggest a set of postulates that an update operator should satisfy. The update postulates are expressed in terms of formulas, not belief sets. That is, an update operator $\diamond$ maps a pair of formulas, one describing the agent's current beliefs and the other describing the new observation, to a new formula that describes the agent's updated beliefs. This is not unreasonable, since we can identify a formula φ with the belief set $Cl(\varphi)$. Indeed, if Φ is finite (which is what Katsuno and Mendelzon assume) every belief set A can be associated with some formula φ_A such that $Cl(\varphi_A) = A$, and every formula φ corresponds to a belief set $Cl(\varphi)$. Thus, any update operator induces an operator that maps a belief state and an observation to a new belief state. We slightly abuse notation and use the same symbol to denote both type of mappings. We say that a belief set A is *complete* if, for every $\varphi \in \mathcal{L}_e$, either $\varphi \in A$ or $\neg\varphi \in A$. A formula μ is *complete* if $Cl(\mu)$ is complete.

The KM postulates are:

[2] $Cl(A) = \{\varphi | A \vdash_{\mathcal{L}_e} \varphi\}$ is the deductive closure of a set of formulas A.

(U1) $\vdash_{\mathcal{L}_e} \mu \diamond \varphi \Rightarrow \varphi$

(U2) If $\vdash_{\mathcal{L}_e} \mu \Rightarrow \varphi$, then $\vdash_{\mathcal{L}_e} \mu \diamond \varphi \Leftrightarrow \mu$

(U3) $\vdash_{\mathcal{L}_e} \neg\mu \diamond \varphi$ if and only if $\vdash_{\mathcal{L}_e} \neg\mu$ or $\vdash_{\mathcal{L}_e} \neg\varphi$

(U4) If $\vdash_{\mathcal{L}_e} \mu_1 \Leftrightarrow \mu_2$ and $\vdash_{\mathcal{L}_e} \varphi_1 \Leftrightarrow \varphi_2$ then $\vdash_{\mathcal{L}_e} \mu_1 \diamond \varphi_1 \Leftrightarrow \mu_2 \diamond \varphi_2$

(U5) $\vdash_{\mathcal{L}_e} (\mu \diamond \varphi) \wedge \psi \Rightarrow \mu \diamond (\varphi \wedge \psi)$

(U6) If $\vdash_{\mathcal{L}_e} \mu \diamond \varphi_1 \Rightarrow \varphi_2$ and $\vdash_{\mathcal{L}_e} \mu \diamond \varphi_2 \Rightarrow \varphi_1$, then $\vdash_{\mathcal{L}_e} \mu \diamond \varphi_1 \Leftrightarrow \mu \diamond \varphi_2$

(U7) If μ is complete then $\vdash_{\mathcal{L}_e} (\mu \diamond \varphi_1) \wedge (\mu \diamond \varphi_2) \Rightarrow \mu \diamond (\varphi_1 \vee \varphi_2)$

(U8) $\vdash_{\mathcal{L}_e} (\mu_1 \vee \mu_2) \diamond \varphi \Leftrightarrow (\mu_1 \diamond \varphi) \vee (\mu_2 \diamond \varphi)$.

The essence of these postulates is as following. After learning φ, the agent believes φ (postulate U1, which is analogous to R2). If φ is already believed, then updating by φ does not change the agent's beliefs (postulate U2, which is a weaker version of R3 and R4). The next two postulates (U3 and U4) deal with coherence of the belief change process. They are analogous to R5 and R6, respectively, with minor differences. Postulates U5 and U6 deal with observations that are related to each other. U5 states that beliefs after learning φ that are consistent with ψ are also believed after learning $\varphi \wedge \psi$. U6 states that if φ_2 is believed after learning φ_1 and φ_1 is believed after learning φ_2, then learning either φ_1 or φ_2 leads to the same belief set. Finally, U7 and U8 deal with decomposition properties of the update operation. U7 states that if μ is essentially a truth assignment to $\mathcal{L}$, then if ψ is believed after learning φ_1 and is also believed after learning φ_2 then it is believed after learning $\varphi_1 \vee \varphi_2$. U8 states that the update of the knowledge base can be computed by independent updates on each sub-part of the knowledge. That is if $\mu = \mu_1 \vee \mu_2$, then we can apply update to each of μ_1 and μ_2, and then combine the results.

5.3 Belief Change Systems

We want to model belief change—particularly belief revision and belief update—in the framework of systems. To do so, we consider a particular class of systems that we call *belief change systems*. In belief change systems, the agent makes observations about an external environment. Just as is (implicitly) assumed in both revision and update, we assume that these observations are described by formulas in some logical language. We then make other assumptions regarding the plausibility measure used by the agent. We formalize our assumptions as conditions BCS1–BCS6, described below, and say that a system $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ is a *belief change system* if it satisfies these conditions. We denote by $\mathcal{C}^{BCS}$ the set of belief change systems.

Assumption BCS1 restricts our attention to single-agent systems that are well behaved, in the sense that the agent's plausibility assessment is consistent with her knowledge and is a function of her local state.

BCS1 $\mathcal{I}$ involves only one agent and satisfies SDP and CONS.

From now on we will use $r_a(m)$ to denote the agent's state (i.e., $r_1(m)$). We do not explicitly mention the agent's name when it is clear from the context. For example, we use $\mathrm{Pl}_{(r,m)}$ to denote $\mathrm{Pl}_{(a,r,m)}$.

Assumption BCS2 formalizes the intuition that our language includes propositions for reasoning about the environment, whose truth depends only on the environment state.

> **BCS2** The language $\mathcal{L}$ includes a propositional sublanguage $\mathcal{L}_e$ over a set $\mathbf{\Phi}_e$ of primitive propositions. $\mathcal{L}_e$ contains the usual propositional connectives and comes equipped with a consequence relation $\vdash_{\mathcal{L}_e}$. The interpretation $\pi(r, m)$ assigns truth to propositions in $\mathbf{\Phi}_e$ in such a way that
>
> (a) $\pi(r, m)$ is consistent with $\vdash_{\mathcal{L}_e}$, that is, $\{p : p \in \mathbf{\Phi}_e, \pi(r, m)(p) = \textbf{true}\} \cup \{\neg p : p \in \mathbf{\Phi}_e, \pi(r, m)(p) = \textbf{false}\}$ is $\vdash_{\mathcal{L}_e}$ consistent, and
>
> (b) $\pi(r, m)(p)$ depends only on $r_e(m)$ for propositions in $\mathbf{\Phi}_e$; that is, $\pi(r, m)(p) = \pi(r', m')(p)$ whenever $r_e(m) = r'_e(m')$.

Part (b) of BCS2 implies that we can evaluate formulas in $\mathcal{L}_e$ with respect to environment states; that is, if $\varphi \in \mathcal{L}_e$ and $r_e(m) = r'_e(m')$, then $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}, r', m') \models \varphi$. Since the environment is all that is relevant for formulas in $\mathcal{L}_e$, if $\varphi \in \mathcal{L}_e$, we write $s_e \models \varphi$ if $(\mathcal{I}, r, m) \models \varphi$ for some point (r, m) such that $r_e(m) = s_e$.

BCS3 is concerned with the form of the agent's local state. The functional form of the revision and update operators suggests that all that matters regarding how an agent changes her beliefs is the agent's belief state and what is learned. This implies that the agent's local state at time $m+1$ should be a function of her local state of time m and the observation made at time m. We in fact make the stronger assumption here that the agent's state consists of the sequence of observations made by the agent. This means that the agent remembers all her past observations. Note that this surely implies that the agent's local state at time $m + 1$ is determined by her state at time m and the observation made at time m. We make a further assumption (that is implicitly made in all the papers on belief revision and belief update), and that is that the observations made by the agent can be described by formulas in $\mathcal{L}_e$. This is quite a strong assumption on the expressive power of $\mathcal{L}_e$. Although this assumption is quite strong, it is standard in the literature. These assumptions are formally in BCS3:

> **BCS3** For all $r \in R$ and for all m, we have $r_a(m) = \langle o_{(r,1)}, \ldots, o_{(r,m)} \rangle$ where $o_{(r,k)} \in \mathcal{L}_e$ for $1 \leq k \leq m$.

Intuitively, $o_{(r,k)}$ is the observation the agent makes immediately after the transition from time $k - 1$ to time k in run r. Thus, it represent what the agent observes about the new state of the system at time k. Note that BCS3 implies that the agent's state at time 0 is the empty sequence in all runs. Moreover, it implies that $r_a(m + 1) = r_a(m) \cdot o_{(r,m+1)}$, where $\cdot$ is the append operation on sequences. That is, the agent's state at $(r, m + 1)$ is the result of appending to her previous state the latest observation she has made about the system. It is not too hard to show that belief change system are synchronous and satisfy perfect recall. (We remark that the agents' local states are modeled in a similar way in the model of knowledge bases presented in [Fagin, Halpern, Moses, and Vardi 1995].)

Clearly we want to reason in our language about the observations the agent makes. Thus, we assume that the language includes propositions that describe the observations made by the agent.

BCS4 The language $\mathcal{L}$ includes a set Φ_{obs} of primitive propositions disjoint from Φ_e such that $\Phi_{obs} = \{learn(\varphi) : \varphi \in \mathcal{L}_e\}$. Moreover, $\pi(r, m)(learn(\varphi)) = \mathbf{true}$ if and only if $o_{(r,m)} = \varphi$ for all runs r and times m.

In a system satisfying BCS1–BCS4, we can talk about belief change. The agent's state encodes observations, and we have propositions that allow us to talk about what is observed. The next assumption is somewhat more geared to situations where observations are always "accepted", so that after the agent observes φ, she believes φ. While this is not a necessary assumption, it is made by both belief revision and belief update, and is a fairly standard intuition. We capture this assumption here by assuming that observations are reliable: the agent observes φ only if the current state of the environment satisfies φ. While this is not the only way of capturing the assumption that observations are accepted, it is perhaps the simplest. In addition, as we shall see, this assumption is consistent with both revision and update, in the sense that we can capture both in systems satisfying it.

BCS5 $(\mathcal{I}, r, m) \models o_{(r,m)}$ for all runs r and times m.

Note that BCS5 implies that the agent never observes *false*. Moreover, it implies that after observing φ, the agent knows that φ is true.

Finally, we assume that belief change proceeds by conditioning. While there are certainly other assumptions that can be made, as we have tried to argue, conditioning is a principled approach that captures the intuitions of minimal change, given the observations. And, as we shall see, conditioning (as captured by PRIOR) is consistent with both revision and update.

BCS6 $\mathcal{I}$ satisfies PRIOR.

Many interesting systems can be viewed as BCS's:

Example 5.3.1: Consider the systems $\mathcal{I}_{diag,1}$ and $\mathcal{I}_{diag,2}$ of Section 4.3 Are these systems BCSs? Not quite, since π_{diag} is not defined on primitive propositions of the form $learn(\varphi)$, but we can easily embed both systems in a BCS. Let $\mathcal{L}_{diag}$ the propositional language defined over Φ_{diag}, and let Φ^+_{diag} consist of Φ_{diag} together with all the primitive propositions of the form $learn(\varphi)$ for $\varphi \in \mathcal{L}_{diag}$. Let π^+_{diag} be the obvious extension of π_{diag} to Φ^+_{diag}, defined so that BCS4 holds. Then in it is easy to see that $(\mathcal{R}_{diag}, \pi^+_{diag}, \mathcal{P}_{diag,i})$ is a BCS: We take the Φ_e of BCS2 to be Φ_{diag}, and define $\vdash_{\mathcal{L}_{diag}}$ so that it enforces the relationships determined by the circuit layout. Thus, for example, if c_1 is an AND gate with input lines l_1 and l_2 and output line l_3, then we would have $\vdash_{\mathcal{L}_{diag}} \neg f_1 \Rightarrow (h_3 \Leftrightarrow h_1 \wedge h_2)$. It is then easy to see that BCS3–BCS6 hold by our construction. ∎

These definitions set the background for our presentation of belief revision and belief update.

5.4 Capturing Revision

Revision can be captured by restricting to BCSs that satisfy several additional assumptions. Before describing these assumptions, we briefly review a well-known representation of revision that will help motivate them.

While there are several representation theorems for belief revision, the clearest is perhaps the following [Grove 1988; Katsuno and Mendelzon 1991b]: We associate with each belief set A a set W_A of possible worlds that consists of those worlds where A is true. Thus, an agent whose belief set is A believes that one of the worlds in W_A is the real world. An agent that performs belief revision behaves as though in each belief state A she has a *ranking*, i.e., a total preorder, over all possible worlds such that the minimal (i.e., most plausible) worlds in the ranking are exactly those in W_A. When revising by φ, the agent chooses the minimal worlds satisfying φ in the ranking and constructs a belief set from them. It is easy to see that this procedure for belief revision satisfies the AGM postulates. Moreover, in [Grove 1988; Katsuno and Mendelzon 1991b], it is shown that any belief revision operator can be described in terms of such a ranking.

This representation suggests how we can capture belief revision in our framework. We define $\mathcal{C}^R \subseteq \mathcal{C}^{BCS}$ to be the set of belief change systems $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ that satisfy the conditions REV1–REV4 that we define below.

Revision assumes that the world does not change during the revision process. Formally this implies that propositions in Φ_e do not change their truth value along a run, i.e., $(\mathcal{I}, r, m) \models p$ if and only if $(\mathcal{I}, r, m+1) \models p$ for all $p \in \Phi_e$. This says that the state of the world is the same with respect to the properties that the agent reasons about (i.e., the propositions in Φ_e).

REV1 $\pi(r, m)(p) = \pi(r, 0)(p)$ for all $p \in \Phi_e$ and points (r, m).

The representation of [Grove 1988; Katsuno and Mendelzon 1991a] requires the agent to totally order possible worlds. We put a similar requirement on the agent's plausibility assessment. Recall that BCS6 says that the agent's plausibility is induced by a prior Pl_a; REV2 strengthens this assumption.

REV2 The prior Pl_a of BCS6 is ranked; that is, for all $A, B \subseteq \mathcal{R}$, either $\mathrm{Pl}_a(A) \leq \mathrm{Pl}_a(B)$ or $\mathrm{Pl}_a(B) \leq \mathrm{Pl}_a(A)$, and $\mathrm{Pl}(A \cup B) = \max(\mathrm{Pl}(A), \mathrm{Pl}(B))$.

The representation of [Grove 1988; Katsuno and Mendelzon 1991a] also requires that the agent considers all truth assignments possible. We need a similar condition, except that we want not only that all truth assignments be considered possible, but that they have nontrivial plausibility as well.

To make this precise, it is helpful to introduce some notation that will be useful for our later definitions as well. Given a system $\mathcal{I}$ and two sequences $\varphi_1, \ldots, \varphi_k$ and $o_1, \ldots, o_{k'}$ of formulas in $\mathcal{L}_e$, let $\mathcal{R}[\varphi_1, \ldots, \varphi_k; o_1, \ldots, o_{k'}]$ consist of all runs r where for each i with $1 \leq i \leq k$, the formula φ_i is true that (r, i) and the agent observes $o_1, \ldots, o_{k'}$. That is, $\mathcal{R}[\varphi_0, \ldots, \varphi_k; o_1, \ldots, o_{k'}] = \{r \in \mathcal{I} : (\mathcal{I}, r, i) \models \varphi_i, i = 0, \ldots, k, \text{ and } r_a(k') = \langle o_1, \ldots, o'_k \rangle\}$. We allow either sequence of formulas to be empty, so, for example, $\mathcal{R}[\varphi; \cdot]$ consists of all runs for which φ is true at the initial state. (Note that if REV1 holds, this means that φ is true in all subsequent states as well.) We use the notation $\mathcal{R}[\varphi_1, \ldots, \varphi_m]$ as an abbreviation for $\mathcal{R}[\varphi_1, \ldots, \varphi_m; \cdot]$.

REV3 If $\varphi \in \mathcal{L}_e$ is consistent, then $\mathrm{Pl}(\mathcal{R}[\varphi]) > \perp$.

It might seem that REV1–REV3 capture all of the assumptions made by the representation of [Grove 1988; Katsuno and Mendelzon 1991a]. However, there is another assumption implicit in the way revision is performed in these representations that we must make explicit in our representation, because of the way we have distinguished observing φ (captured by the formula *learn*(φ)) from φ itself. Intuitively, when the agent observes φ, she updates her plausibility assessment by conditioning on φ. This is essentially what we can think of the earlier representations as doing. However, in our representation, the agent does *not* condition on φ, but on the fact that she has observed φ. Although we do require that φ must be true if the agent observes it (BCS5), the agent may in general gain extra information by observing φ.

To understand this issue, consider the following example. Suppose that $\mathcal{R}$ is such that the agent observes p_1 at time (r, m) only if p_2 and q are also true at (r, m), and she observes $p_1 \wedge p_2$ at (r, m) only if q is false. It is easy to construct a BCS satisfying REV1–REV3 that also satisfies these requirements. In this system, after observing p_1, the agent believes p_2 and q. According to AGM's postulate R7 (and also KM's postulate U5) the agent must believe q after observing $p_1 \wedge p_2$. However, in $\mathcal{R}$, the agent believes (indeed knows) $\neg q$ after observing $p_1 \wedge p_2$. Thus, revision and update both are implicitly assuming that the observation of φ does not provide such additional knowledge. The following assumption ensures that this is the case for revision (a more general version will be required for update; see Section 5.5).

REV4 $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi; o_1, \ldots, o_m])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o_1 \wedge \ldots \wedge o_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi \wedge o_1 \wedge \ldots \wedge o_m])$.

This assumption captures the intuition that observing $o_1, \ldots, o_k$ provides no more information than just the fact that $o_1 \wedge \ldots \wedge o_m$ is true. That is, the agent compares the plausibility of φ and ψ in the same way after conditioning by the observations $o_1, \ldots, o_m$ as after conditioning by the fact that $o_1 \wedge \ldots \wedge o_m$ is true. It easily follows from REV4 and PRIOR that the agent believes ψ after observing $o_1 \wedge \ldots \wedge o_m$ exactly if $o_1 \wedge \ldots \wedge o_m \wedge \psi$ was initially considered more plausible than $o_1 \wedge \ldots \wedge o_m \wedge \neg\psi$. Thus, the agent believes ψ after observing $o_1 \wedge \ldots \wedge o_m$ exactly if initially, she believed ψ conditional on $o_1 \wedge \ldots \wedge o_m$: the observations provide no extra information beyond the fact that each of the o_i's are true.

REV4 is quite a strong assumption. Not only does it say that observations do not give the agent any additional observation (beyond the fact that they are true), it also says that all consistent observations can be made (since if $\varphi \wedge o$ is consistent, we must have $\mathrm{Pl}_a(\mathcal{R}[\varphi; o]) = \mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o]) > \perp$, by REV3 and REV4. We might instead consider using a weaker version that says that, provided an observation can be made, it gives no additional information. Formally, this would be captured as

REV4′ If $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) > 0$, then $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi; o_1, \ldots, o_m])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o_1 \wedge \ldots \wedge o_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi \wedge o_1 \wedge \ldots \wedge o_m])$.

The following examples suggests that REV4$'$ may be more reasonable in practice than REV4. We used REV4 only because it comes closer to the spirit of the requirement of revision that all observations are possible.

Example 5.4.1: Consider the system $\mathcal{I}_{diag,1}$ described in Section 4.3. As discussed in Example 5.3.1, this system can be viewed as a BCS. Is it a revision system? It is easy to see that $\mathcal{I}_{diag,1}$ satisfies REV2 and REV3. It clearly does not satisfy REV1, since propositions that describe input/output lines can change their values from one point to the next. However, as we are about to show, a slight variant of $\mathcal{I}_{diag,1}$ does satisfy REV1. A more fundamental problem is that $\mathcal{I}_{diag,1}$ does not satisfy REV4. This is inherent in our assumption that the agent never directly observes faults, so that, for example, we have $\mathrm{Pl}_{diag,1}(\mathcal{R}[\cdot; f_1]) = \perp$, while $\mathrm{Pl}_{diag,1}(\mathcal{R}[f_1] > \perp$. It does, however, satisfy REV4$'$.

To see how to modify $\mathcal{I}_{diag,1}$ so as to satisfy REV1, recall that in the diagnosis task, the agent is mainly interested in her beliefs about faults. Since faults are static in $\mathcal{I}_{diag,1}$, we can satisfy REV1 if we ignore all propositions except $f_1, \ldots, f_n$. Let $\Phi'_{diag} = \{f_1, \ldots, f_n\}$ and let $\mathcal{L}'_{diag}$ be the propositional language over Φ'_{diag}. For every observation o made by the agent regarding the value of the lines, there corresponds a formula in $\mathcal{L}'_{diag}$ that characterizes all the fault sets that are consistent with o. Thus, for every run r in $\mathcal{I}_{diag,1}$, we can construct a run r' where the agent's local state is a sequence of formulas in $\mathcal{L}'_{diag}$. Let $\mathcal{I}'_{diag}$ be the system consisting of all such runs r'. We can clearly put a plausibility assignment on these runs so that $\mathcal{I}_{diag,1}$ and $\mathcal{I}'_{diag}$ are isomorphic in an obvious sense. In particular, the agent has the same beliefs about formulas in $\mathcal{L}'_{diag}$ at corresponding points in the two systems. More precisely, if $\varphi \in \mathcal{L}'_{diag}$, then $(\mathcal{I}'_{diag}, r, m) \models \varphi$ if and only if $(\mathcal{I}_{diag,1}, r, m) \models \varphi$ for all points (r, m) in $\mathcal{I}_{diag,1}$.

It is easy to verify that $\mathcal{I}'_{diag}$ satisfies REV1–REV3 and REV4$'$, although it still does not satisfy REV4. ∎

We want to show that a revision operator corresponds to a system in $\mathcal{C}^R$ and vice versa. To do so, we need to examine the beliefs of the agent at each point (r, m). First we note that if the $(r, m) \sim_a (r', m')$ then $(\mathcal{I}, r, m) \models B\varphi$ if and only if $(\mathcal{I}, r', m') \models B\varphi$; this is a consequence of the requirement that the agent's plausibility assessment is a function of her local state. Thus, we think of the agent's beliefs as a function of her local state. We use the notation $(\mathcal{I}, s_a) \models B\varphi$ as shorthand for $(\mathcal{I}, r, m) \models B\varphi$ for some (r, m) such that $r_a(m) = s_a$. Let s_a be some local state of the agent. We define the agent's *belief state* at s_a as

$$\mathrm{Bel}(\mathcal{I}, s_a) = \{\varphi \in \mathcal{L}_e : (\mathcal{I}, s_a) \models B\varphi\}.$$

Since the agent's state is a sequence of observations, the agent's state after observing φ is simply $s_a \cdot \varphi$, where $\cdot$ is the append operation. Thus, $\mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi)$ is the belief state after observing φ. We adopt the convention that if the agent can never attain the local state s_a in $\mathcal{I}$, then $\mathrm{Bel}(\mathcal{I}, s_a) = \mathcal{L}_e$. With these definitions, we can compare the agent's belief state before and after observing φ, that is $\mathrm{Bel}(\mathcal{I}, s_a)$ and $\mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi)$.

We start by showing that every AGM revision operator can be represented in $\mathcal{C}^R$.

Theorem 5.4.2: *Let $\circ$ be an AGM revision operator and let $K \subseteq \mathcal{L}_e$ be a consistent belief state. Then there is a system $\mathcal{I}_{\circ,K} \in \mathcal{C}^R$ such that Bel$(\mathcal{I}_{\circ,K}, \langle\rangle) = K$ and*

$$Bel(\mathcal{I}_{\circ,K}, \langle\rangle) \circ \varphi = Bel(\mathcal{I}_{\circ,K}, \langle\varphi\rangle)$$

for all $\varphi \in \mathcal{L}_e$.

Thus, Theorem 5.4.2 says that we can represent a revision operator $\circ$ in the sense that we have a family of systems $\mathcal{I}_{\circ,K} \in \mathcal{C}^R$, one for each consistent belief state K, such that K is the agent's initial belief state in $\mathcal{I}_{\circ,K}$, and for each formula φ in $\mathcal{L}_e$, the agent's belief state after learning φ is $K \circ \varphi$. Notice that we restrict attention to consistent belief states K. The AGM postulates allow the agent to "escape" from an inconsistent state, so that $K \circ \varphi$ may be consistent even if K is inconsistent. We might thus hope to extend the theorem so that it also applies to the inconsistent belief state, but this is impossible in our framework. If *false* $\in$ Bel$(\mathcal{I}_{\circ,K}, s_a)$ for some state s_a, and $r_a(m) = s_a$, then $\text{Pl}_{(r,m)}(W_{(r,m)}) = \perp$. Since we update by conditioning, we must have $\text{Pl}_{(r,m+1)}(W_{(r,m+1)}) = \perp$, so the agent's belief state will remain inconsistent no matter what she learns. Although we could modify our framework to allow the agent to escape from inconsistent states, we actually consider this to be a defect in the AGM postulates, not in our framework. To see why, suppose that the agent's belief set is inconsistent at s_a, and $r_a(m) = s_a$. Thus, the agent considers all states in $W_{(r,m)}$ to be completely implausible (since $\text{Pl}_{(r,m)}(W_{(r,m)}) = \perp$). On the other hand, to escape inconsistency, she must have a plausibility ordering over the worlds in $W_{(r,m)}$. These two requirements seem somewhat inconsistent.

Not surprisingly, this inconsistency creates problems for other semantic representations in the literature. For example, Boutilier's representation theorem [1992] states that for every revision operator $\circ$ and belief state K, there is a ranking R such that $\psi \in K \circ \varphi$ if and only if ψ is believed in the minimal φ-worlds according to R. If we examine this theorem, we note that he does not state that the minimal (i.e., most preferred) worlds in R correspond to the belief state K (in the sense that the minimal worlds are precisely those where the formulas in K hold); this would be the analogue of our requiring that Bel$(\mathcal{I}_{\circ,K}, \langle\rangle) = K$. In fact, if K is $\vdash_{\mathcal{L}_e}$-consistent, the minimal worlds do correspond to K. However, if K is inconsistent, they cannot, since any nonempty ranking induces a consistent set of beliefs. We could state a weaker version of Theorem 5.4.2 that would correspond exactly to Boutilier's theorem. We presented the stronger result (that does not apply to inconsistent belief states) to bring out what we believe to be a problem with the AGM postulates. See [Friedman and Halpern 1996a] for further discussion of this issue.

Theorem 5.4.2 shows that, in a precise sense, we can map AGM revision operations to $\mathcal{C}^R$. What about the other direction? The next theorem shows that the first belief change step in systems in $\mathcal{C}^R$ satisfies the AGM postulates.

Theorem 5.4.3: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$. Then there is an AGM revision operator $\circ_{\mathcal{I}}$ such that*

$$Bel(\mathcal{I}, \langle\rangle) \circ_{\mathcal{I}} \varphi = Bel(\mathcal{I}, \langle\varphi\rangle)$$

for all $\varphi \in \mathcal{L}_e$.

We remark that if we used REV4$'$ instead of REV4, then we would be able to prove this result only for those formulas φ that are observable (i.e., for which $\mathrm{Pl}(\mathcal{R}[\varphi]) > \bot$).

Both Theorems 5.4.2 and 5.4.3 apply to one-step revision, starting from the initial (empty) state. What happens once we allow iterated revision? In our framework, observations are taken to be known, so if the agent makes an inconsistent sequence of observations, then her belief state will be inconsistent, and (as we observed above) will remain inconsistent from then on, no matter what she observes. This creates a problem if we try to get analogues to Theorems 5.4.2 and 5.4.3 for iterated revision. As the following theorem demonstrates, we can already see the problem if we consider one-step revisions from a state other than the initial state.

Theorem 5.4.4: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ and let $s_a = \langle \varphi_1, \ldots, \varphi_k \rangle$ be a local state in $\mathcal{I}$. Then there is an AGM revision operator $\circ_{\mathcal{I},s_a}$ such that*

$$\mathit{Bel}(\mathcal{I}, s_a) \circ_{\mathcal{I},s_a} \varphi = \mathit{Bel}(\mathcal{I}, s_a \cdot \varphi)$$

for all formulas $\varphi \in \mathcal{L}_e$ such that $\varphi_1 \wedge \ldots \varphi_k \wedge \varphi$ is consistent.

We cannot do better than this. If $\varphi_1 \wedge \ldots \varphi_k \wedge \varphi$ is inconsistent then, because of our requirements that all observations must be true of the current state of the environment (BCS5) and that propositions are static (REV1), there cannot be any global state in $\mathcal{I}$ where the agent's local state in $s_a \cdot \varphi$. Thus, $\mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi)$ is inconsistent, contradicting R5.

There is another problem with trying to get an analogue of Theorem 5.4.3 for iterated revision, a problem that seems inherent in the AGM framework. Our framework makes a clear distinction between the agent's *epistemic state* at a point (r, m) in $\mathcal{I}$, which we can identify with her local state $s_a = r_a(m)$, and the agent's *belief state* at (r, m), $\mathrm{Bel}(\mathcal{I}, s_a)$, which is the set of formulas she believes. In a system in $\mathcal{C}^R$, the agent's belief state does not in general determine how the agent's beliefs will be revised; her epistemic state does. On the other hand, the AGM postulates assume that revision is a function of the agent's belief state. Now suppose we have a system $\mathcal{I}$ and two points (r, m) and (r, m') on some run $r \in \mathcal{I}$ such that (1) the agent's belief set is the same at (r, m) and (r, m'), that is $\mathrm{Bel}(\mathcal{I}, r_a(m)) = \mathrm{Bel}(\mathcal{I}, r_a(m'))$, (2) the agent observes φ at both (r, m) and (r, m'), (3) $\mathrm{Bel}(\mathcal{I}, r_a(m+1)) \neq \mathrm{Bel}(\mathcal{I}, r_a(m'+1)$. It is not hard to construct such a system $\mathcal{I}$. However, there cannot be an analogue of Theorem 5.4.3 for $\mathcal{I}$, even if we restrict to consistent sequences of observations. For suppose there were a revision operator $\circ$ such $\mathrm{Bel}(\mathcal{I}, \langle\rangle)) \circ \varphi_1 \circ \cdots \circ \varphi_k = \mathrm{Bel}(\mathcal{I}, \langle \varphi_1, \ldots, \varphi_k \rangle)$ for all $\varphi_1, \ldots, \varphi_k$ such that $\varphi_1 \wedge \ldots \wedge \varphi_k$ is consistent. Then we would have $\mathrm{Bel}(\mathcal{I}, r_a(m+1)) = \mathrm{Bel}(\mathcal{I}, r_a(m)) \circ \varphi = \mathrm{Bel}(\mathcal{I}, r_a(m')) \circ \varphi = \mathrm{Bel}(\mathcal{I}, r_a(m'+1))$, contradicting our assumption.

The culprit here is the assumption that revision depends only on the agent's belief state. To see why this is an unreasonable assumption, consider a situation where at time 0 the agent believes both p and q, but her belief in q is stronger than her belief in p (i.e., the plausibility of q is greater than that of p). We can well imagine that after observing $\neg p \vee \neg q$ at time 1, she would believe $\neg p$ and q. However, if she first observed p at time 1 and then $\neg p \vee \neg q$ at time 2, she would believe p and $\neg q$, because, as a result of observing p, she would assign p greater plausibility than q. Note, however, that the AGM postulates dictate that after an observation that is already believed, the agent does not change her beliefs. Thus, the AGM setup would force the agent to have the same beliefs after learning $\neg p \vee \neg q$ in both situations.

There has been a great deal of work on the problem of *iterated belief revision* [Boutilier 1993; Darwiche and Pearl 1994; Freund and Lehmann 1994; Lehmann 1995; Levi 1988; Williams 1994]). Much of the recent work moves away from the assumption that belief revision depends solely on the agent's belief state. For example the approaches of Boutilier [1993] and Darwiche and Pearl [1994] define revision operators that map (rankings $\times$ formulas) to rankings. Because our framework makes such a clear distinction between epistemic states and belief states, it gives us a natural way of basically maintaining the spirit of the AGM postulates while assuming that revision is a function of epistemic states. Rather than taking $\circ$ to be a function from (belief states $\times$ formulas) to belief states, we take it $\circ$ to be a function from (epistemic states $\times$ formulas) to epistemic states. We can easily modify the AGM postulates to deal with such revision operators on epistemic states. We start by assuming that there is a set of epistemic states and a function $\mathrm{Bel}(\cdot)$ that maps epistemic states to belief states. We then have analogues to each of the AGM postulates, obtained by replacing each belief set by the beliefs in the corresponding epistemic state. For example, we have

(R1′) $E \circ \varphi$ is an epistemic state

(R2′) $\varphi \in \mathrm{Bel}(E \circ \varphi)$

(R3′) $\mathrm{Bel}(E \circ \varphi) \subseteq Cl(\mathrm{Bel}(E) \cup \{\varphi\})$

and so on, with the obvious transformation.[3]

We can get strong representation theorems if we work at the level of epistemic states. Given a language $\mathcal{L}_e$ (with an associated consequence relation $\vdash_{\mathcal{L}_e}$), let $\mathcal{E}_{\mathcal{L}_e}$ consist of all finite sequences of formulas in $\mathcal{L}_e$. Note that we allow $\mathcal{E}_{\mathcal{L}_e}$ to include sequences of formulas whose conjunction is inconsistent. We define revision in $\mathcal{E}_{\mathcal{L}_e}$ in the obvious way: if $E \in \mathcal{E}_{\mathcal{L}_e}$, then $E \circ \varphi = E \cdot \varphi$.

Theorem 5.4.5: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ whose local states are $\mathcal{E}_{\mathcal{L}_e}$. There is a function $Bel_{\mathcal{I}}$ that maps epistemic states to belief states such that*

- *if s_a is a local state of the agent in $\mathcal{I}$, then $Bel(\mathcal{I}, s_a) = Bel_{\mathcal{I}}(s_a)$, and*
- *$(\circ, Bel_{\mathcal{I}})$ satisfies R1′–R8′.*

Notice that, by definition, we have $\mathrm{Bel}_{\mathcal{I}}(\mathcal{I}, \langle\rangle \circ_{\mathcal{I}} \varphi_1 \circ_{\mathcal{I}} \ldots \circ_{\mathcal{I}} \varphi_k) = \mathrm{Bel}_{\mathcal{I}}(\mathcal{I}, \langle \varphi_1, \ldots, \varphi_k \rangle)$, so, at the level of epistemic states, we get an analogue to Theorem 5.4.3. We remark that to ensure that R5′ holds for $(\circ, \mathrm{Bel}_{\mathcal{I}})$, we need to define $\mathrm{Bel}_{\mathcal{I}}(E)$ appropriately for sequences $E \in \mathcal{E}_{\mathcal{I}}$ whose conjunction is inconsistent.

Theorem 5.4.5 shows that any system in $\mathcal{C}^R$ corresponds to a revision operator over epistemic states that satisfies the generalized AGM postulates. We would hope that the converse also holds. Unfortunately, this is not quite the case. There are revision operators on epistemic states that satisfy

[3] The only problematic postulate is R6. The question is whether R6′ should be "If $\vdash_{\mathcal{L}_e} \varphi \Leftrightarrow \psi$ then $\mathrm{Bel}(E \circ \varphi) = \mathrm{Bel}(E \circ \psi)$" or "If $\vdash_{\mathcal{L}_e} \varphi \Leftrightarrow \psi$ then $E \circ \varphi = E \circ \psi$". Dealing with either version is straightforward. For definiteness, we adopt the first alternative here.

the generalized AGM postulates but do not correspond to a system in $\mathcal{C}^R$. This is because systems in $\mathcal{C}^R$ satisfy an additional postulate:

(R9′) If $\not\vdash_{\mathcal{L}_e} \varphi \wedge \psi$ then $\mathrm{Bel}(E \circ \varphi \circ \psi) = \mathrm{Bel}(E \circ \varphi \wedge \psi)$.

We start by showing that R9′ is sound in $\mathcal{C}^R$, by proving the following strengthening of Theorem 5.4.5.

Proposition 5.4.6: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ whose local states are $\mathcal{E}_{\mathcal{L}_e}$. There is a function $Bel_{\mathcal{I}}$ that maps epistemic states to belief states such that*

- *if s_a is a local state of the agent in $\mathcal{I}$, then $Bel(\mathcal{I}, s_a) = Bel_{\mathcal{I}}(s_a)$, and*
- *$(\circ, Bel_{\mathcal{I}})$ satisfies R1′–R9′.*

We can prove the converse to Proposition 5.4.6: a revision system on epistemic states that satisfies the generalized AGM postulates *and R9′* does correspond to a system in $\mathcal{C}^R$.

Theorem 5.4.7: *Given a function $Bel_{\mathcal{L}_e}$ mapping epistemic states in $\mathcal{E}_{\mathcal{L}_e}$ to belief sets over $\mathcal{L}_e$ such that $Bel_{\mathcal{L}_e}(\langle\rangle)$ is consistent and $(Bel_{\mathcal{L}_e}, \circ)$ satisfies R1′–R9′, there is a system $\mathcal{I} \in \mathcal{C}^R$ whose local states are in $\mathcal{E}_{\mathcal{L}_e}$ such that $Bel_{\mathcal{L}_e}(s_a) = Bel(s_a)$ for each local state $s_a$in $\mathcal{I}$.*

Notice that, by definition, for the system $\mathcal{I}$ of Theorem 5.4.7, we have $\mathrm{Bel}(\langle\rangle \circ \varphi_1 \circ \ldots \circ \varphi_k) = \mathrm{Bel}(\langle\varphi_1, \ldots, \varphi_k\rangle)$ as long as $\varphi_1 \wedge \ldots \wedge \varphi_k$ is consistent.

5.5 Capturing Update

Update tries to capture the intuition that there is a preference for runs where all the observations made are true, and where changes from one point to the next along the run are minimized. To capture the notion of "minimal change from world to world", we use a *distance function* d on worlds.[4] Given two worlds w and w', $d(w, w')$ measures the distance between them. Distances might be incomparable, so we require that d map pairs of worlds into a *partially ordered* domain with a unique minimal element 0 and that $d(w, w') = 0$ if and only if $w = w'$.

We start with some preliminary definitions. Let $\mathcal{I}$ be a BCS, and let $s_0, \ldots, s_n$ be a set of environment states in $\mathcal{I}$. We define $[s_0, \ldots, s_n]$ as the set of runs where $r_e(i) = s_i$ for all $0 \le i \le n$. Thus, $[s_0, \ldots, s_n]$ describes a set of runs that share a common prefix of environment states. A prior plausibility space $\mathcal{P}_a = (\mathcal{R}, \mathrm{Pl}_a)$ is *consistent* with a distance measure d if the following holds:

> $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$ if and only if there is some $j < n$ such that $s_k = s'_k$ for all $0 \le k \le j$, $s_{j+1} \neq s'_{j+1}$, and $d(s_j, s_{j+1}) < d(s_j, s'_{j+1})$.

[4] Katsuno and Mendelzon identify a "world" with a truth assignment to the primitive propositions. For us, this is just the environment state.

Intuitively, we compare events of the form $[s_0, \ldots, s_n]$ using a lexicographic ordering based on d. Notice that this ordering focuses on the *first* point of difference. Runs with the a smaller change at this point is preferred, even if later there are abnormal changes. This point is emphasized in the borrowed car example below.

The requirement of consistency with a distance measure requires that the prior on run prefixes is lexicographic. Additionally, we say that Pl_a is *prefix-defined* if the plausibility of an event is uniquely defined by the plausibility of run-prefixes that are contained in it, so that

> $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_m])$ if and only if for all $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_m] - \mathcal{R}[\varphi_0, \ldots, \varphi_m]$ there is some $[s'_0, \ldots, s'_m] \subseteq \mathcal{R}[\varphi_0, \ldots, \varphi_m]$ such that $\mathrm{Pl}_a([s'_0, \ldots, s'_m]) > \mathrm{Pl}_a([s_0, \ldots, s_m])$.

Roughly speaking, this requirement states that we compare events by properties of dominance. This property is similar to one satisfied by the plausibility measures that we get from preference ordering using the construction of Theorem 2.4.2.

We define the set $\mathcal{C}^U$ to consist of BCSs $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ that satisfy the following four requirements UPD1–UPD4. UPD1 says that there are only finitely many possible truth assignments, and that there is a one-to-one map between environment states and truth assignments.

> **UPD1** The set Φ_e of propositions (of BCS2) is finite and π is such that for all environment states s, s', if $s \neq s'$, then there is a formula $\varphi \in \mathcal{L}_e$ such that $s \models \varphi$ and $s' \models \neg\varphi$.

UPD2–UPD4 are analogues to REV2–REV4. Like REV2, UPD2 puts constraints on the form of the prior, but now we consider lexicographic priors of the form described above.

> **UPD2** The prior of BCS6 is prefix defined and consistent with some distance measure.

Recall that REV3 requires only that that all truth assignments initially have nontrivial plausibility. In the case of revision, the truth assignment does not change over time, since we are dealing with static propositions. In the case of update, the truth assignment may change over time, so UPD3 requires that all consistent sequences of truth assignments have nontrivial plausibility.

> **UPD3** If $\varphi_i \in \mathcal{L}_e$, $i = 0, \ldots, k$, are consistent formulas, then $\mathrm{Pl}(\mathcal{R}[\varphi_0, \ldots, \varphi_k]) > \perp$.

Finally, like REV4, UPD4 requires that the agent gain no information from her observations beyond the fact that they are true.

> **UPD4** $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_{k+1}; o_1, \ldots, o_k]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_{m+1}; o_1, \ldots, o_m])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_m \wedge o_m, \varphi_{m+1}]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \psi_1 \wedge o_1, \ldots, \psi_m \wedge o_m, \psi_{m+1}])$

We remark that in the presence of REV1, UPD4 is equivalent to REV4. We might consider generalized versions of UPD4, where the two sequences of formulas can have arbitrary relative lengths; this version suffices for our purposes. We can also define an analogue UPD4$'$ in the spirit of REV4$'$, which applies only if $\mathrm{Pl}(\mathcal{R}[\varphi_0, \ldots, \varphi_{m+1}; o_1, \ldots, o_m]) > \perp$.

We now show that $\mathcal{C}^U$ corresponds to Katsuno and Mendelzon's notion of update. Recall that Katsuno and Mendelzon define an update operator as mapping a pair of formulas (μ, φ), where μ describes the agent's beliefs and φ describes the observation, to a new formula $\mu \diamond \varphi$ that describes the agent's new beliefs. However, as we discussed in Section 5.2, when Φ_e is finite, we can also treat $\diamond$ mapping a belief state and a formula to a new belief state. Also recall that $\mathrm{Bel}(\mathcal{I}, s_a)$ is the agent's belief set when her local state is s_a.

Theorem 5.5.1: *A belief change operator $\diamond$ satisfies U1–U8 if and only if there is a system $\mathcal{I} \in \mathcal{C}^U$ such that*

$$Bel(\mathcal{I}, s_a) \diamond \psi = Bel(\mathcal{I}, s_a \cdot \psi)$$

for all epistemic states s_a and formulas $\psi \in \mathcal{L}_e$.

This result immediately generalizes to sequences of updates:

Corollary 5.5.2: *A belief change operator $\diamond$ satisfies U1–U8 if and only if there is a system $\mathcal{I}_\diamond \in \mathcal{C}^U$ such that for all $\psi_1, \ldots, \psi_k \in \mathcal{L}_e$, we have*

$$Bel(\mathcal{I}_\diamond, s_a) \diamond \psi_1 \diamond \ldots \diamond \psi_k = Bel(\mathcal{I}_\diamond, s_a \cdot \psi_1 \cdot \ldots \cdot \psi_k).$$

These results show that for update, unlike revision, the systems we consider are such that the belief state does determine the result of the update, i.e., if $\mathrm{Bel}(\mathcal{I}, s_a) = \mathrm{Bel}(\mathcal{I}, s'_a)$, then for any φ we get that $\mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi) = \mathrm{Bel}(\mathcal{I}, s'_a \cdot \varphi)$. Roughly speaking, the reason is that the distance measure that determines the prior does not change over time. While this allows us to get an elegant representation theorem, it also causes problems for the applicability of update, as we shall see below.

Note that, since the world is allowed to change, there is no problem if we update by a sequence $\psi_1, \ldots, \psi_k$ of consistent formulas such that $\psi_1 \wedge \ldots \wedge \psi_k$ is inconsistent. There is no requirement that the formulas $\psi_1, \ldots, \psi_k$ be true simultaneously. All that matters is that ψ_i is true at time i. Also note that an update by an inconsistent formula does not pose a problem for our framework. The KM postulates states once the agent learns an inconsistent formula, she believes *false* from then on.

How reasonable is the notion of update? As the discussion of UPD2 above suggests, it has a preference for deferring abnormal events. This makes it quite similar to Shoham's *chronological ignorance* [1988], and it suffers from some of the same problems. Consider the following story, that we call the *borrowed-car example*.[5] At time 1, the agent parks her car in front of her house with a full fuel tank. At time 2, she is in her house. At time 3, she returns outside to find the car still parked where she left it. Since the agent does not observe the car while she is inside the house, there is no reason for her to revise her beliefs regarding the car's location. Since she finds it parked at time 3, she still has no reason to change her beliefs. Now, what should the agent believe when, at time 4, she notices that the fuel tank is no longer full? The agent may want to consider

[5]This example is based on Kautz's stolen car story [1986], and is due to Boutilier, who independently observed this problem [private communication, 1993].

a number of possible explanations for her time-4 observation, depending on what she considers to be the most likely sequence(s) of events between time 1 and time 4. For example, if she has had previous gas leaks, then she may consider leakage to be the most plausible explanation. On the other hand, if her spouse also has the car keys, she may consider it possible that he used the car in her absence. Update, however, prefers to defer abnormalities, so it will conclude that the fuel must have disappeared, for inexplicable reasons, between times 3 and 4. To see this, note that runs where the car has been taken on a ride have an abnormality at time 2, while runs where the car did not move at time 2 but the fuel suddenly disappeared, have their first abnormality at time 4, and thus are preferred!

We emphasize that the counterintuitive conclusion drawn in this example is not an artifact of our representation, but inherent in the definition of update. We can formalize the example using propositions such as *car-in-lot, fuel-tank-full*, etc. The observation of $\neg$*fuel-tank-full* at time 4 must be explained by some means, and an update operator will explain it in terms of a change that occurred in states consistent with the beliefs at time 3 (i.e., *car-in-lot, fuel-tank-full*). The exact change assumed will depend on the distance function embodied by the update operator. The key point is that update will not go back and revise the earlier beliefs about what happened between times 1 and 2.

So are there any situations where update is reasonable? Of course, this depends on how we interpret "reasonable". We briefly consider one approach here.

In a world w, the agent has some beliefs that are described by, say, the formula φ. These beliefs may or may not be *correct* (where we say a belief φ is correct in a world w if φ is true of w). Suppose something happens and the world changes to w'. As a result of the agent's observations, she has some new beliefs, described by φ'. Again, there is no reason to believe that φ' is correct. Indeed, it may be quite unreasonable to expect φ' to be correct, even if φ is correct. Consider the borrowed-car example. Suppose that while the agent was sitting inside the house, the car was, in fact, taken for a ride. Nevertheless, the most reasonable belief for the agent to hold when she observes that the car is still in the parked after she leaves the house is that it was there all along.

The problem here is that the information the agent obtains at times 2 and 3 is insufficient to determine what happened. We cannot expect all the agent's beliefs to be correct at this point. On the other hand, if she does obtain sufficient information about the change and her beliefs were initially correct, then it seems reasonable to expect that her new beliefs will be correct. But what counts as *sufficient* information?

We say that φ provides *sufficient information* about the change from w to w' if there is no world w'' satisfying φ such that $d(w, w'') < d(w, w')$. In other words, φ is sufficient information if, after observing φ in world w, the agent will consider the real world (w') one of the most likely worlds. Note that this definition is monotonic, in that if φ is sufficient information about the change, then so is any formula ψ that implies φ (as long as it holds at w'). Moreover, this definition depends on the agent's distance function d. What constitutes sufficient information for one agent might not for another. We would hope that the function d is realistic in the sense that the worlds judged closest according to d really are the most likely to occur.

We can now show that update has the property that if the agent has correct beliefs and receives sufficient information about a change, then she will continue to have correct beliefs.

Theorem 5.5.3: *Let $\mathcal{I} \in \mathcal{C}^U$. If the agent's beliefs at (r, m) are correct and $o_{(r,m)}$ provides sufficient information about the change from $r_e(m)$ to $r_e(m+1)$, then the agent's beliefs at $(r, m+1)$ are correct.*

As we observed earlier, we cannot expect the agent to always have correct beliefs. Nevertheless, we might hope that if the agent does (eventually) receive sufficiently detailed information, then she should realize that her beliefs were incorrect. But this is precisely what does *not* happen in the borrowed-car example. Intuitively, once the agent observes that the fuel tank is not full, this should be sufficient information to eliminate the possibility that the car remained in the parking lot. However, it is not. Roughly speaking, this is because update focuses only on the current state of the world, and thus cannot go back and revise beliefs about the past.

The problem here is again due to the fact that belief update is determined only by the agent's belief state and not her epistemic state. Thus, update can only take into account the agent's current beliefs and not other information, such as the sequence of observations that led to these beliefs. In our example, if we limit our attention to beliefs about the car's whereabouts and the fuel tank, then since the agent has the same belief state at time 1 and 3, she must change her beliefs in the same manner at both times. This implies that the observation the fuel tank is not full at time 4 cannot be sufficient information about the past, since a fuel leak might be the most plausible explanation of missing fuel at time 2.

Our discussion of update shows that update is guaranteed to be safe only in situations where there is always enough information to characterize the change that has occurred. While this may be a plausible assumption in database applications, it seems somewhat less reasonable in AI examples, particularly in cases involving reasoning about action.[6]

5.6 Synthesis

In previous sections we analyzed belief revision and belief update separately. We provided representation theorems for both notions and discussed issues specific to each notion. In this section, we try to identify some common themes and points of difference.

Katsuno and Mendelzon [1991a] focused on the following three differences between revision and update:

1. Revision deals with static propositions, while update allows propositions that are not static.

2. Revision and update treat inconsistent belief states differently. Revision allows the agent to "recover" from an inconsistent state after observing a consistent formula. Update dictates that once the agent has inconsistent beliefs, she will continue to have inconsistent beliefs. As we noted above, it seems that revision's ability to recover from an inconsistent belief set leads to several technical oddities in iterated revision.

[6]Similar observations were independently made by Boutilier [1994b], although his representation is quite different from ours.

Table 5.1: A summary of the restrictions we had to impose to capture revision and update.

Restriction on	Revision	Update
Environment changes	No change	All possible sequences
Initial plausibility	Total preorder	Lexicographic
Belief change	Conditioning	Conditioning

3. Revision considers only total preorders, while update allows partial preorders.

We remark that while the restriction to static propositions may seem to be a serious limitation of belief revision, we can always convert a dynamic proposition to a static one by adding time stamps. That is, we can replace a proposition p by a family of propositions p^m that stand for "p is true at time m". Thus, it is possible to use revision to reason about a changing world. (Of course, it would then be necessary to capture connections between propositions of the form p^m, but specific revision operators could certainly do this.)

In any case, our framework suggests a different approach to categorizing the differences between revision and update (and other approaches to belief change): focusing on the restrictions that have to be added to basic BCSs to obtain systems in $\mathcal{C}^R$ and $\mathcal{C}^U$, respectively. In particular, we should focus on three aspects of a system:

- How does the environment state change?
- How does the agent form her initial beliefs? What regularities appear in the agent's beliefs at the initial state?
- How does the agent change her beliefs?

Table 5.1 summarizes the answers to these questions for revision and update. The table highlights the different restrictions imposed by each. Revision puts a severe restriction on changes of the environment (i.e., the environment cannot change during revision), and a rather mild restriction on the agent's prior beliefs (i.e., they must form a total preorder). On the other hand, update allows all sequences of environment states, but requires the agent's prior beliefs to have a specific form. These formal properties match the intuitive description of revision and update given in [Alchourrón, Gärdenfors, and Makinson 1985; Katsuno and Mendelzon 1991b]. However, the explicit representation of time in our framework allows us to make these intuitions precise. Moreover, our framework makes explicit other assumptions made by revision and update. For example, the lexicographic nature of update is not immediately evident from the presentation in [Katsuno and Mendelzon 1991b].

The key point to notice in this table is that belief change in both revision and update is done by conditioning. This observation, and the naturalness of conditioning as a notion of change, support our claim that conditioning should be adopted as semantic foundations for minimal change.

5.7 Extensions

In the preceding sections, we introduced several assumptions that were needed to capture revision and update. Of course, there are other ways of capturing these notions that require somewhat different assumptions. Nevertheless, these assumptions give insight into the underlying choices made, either explicitly or implicitly, in the definition of revision and update. In addition, thinking in terms of such restrictions makes it straightforward to extend the intuitions of revision and update beyond the context where they were originally applied. In this section, we consider a number of such extensions, to illustrate our point.

Knowledge. In many domains of interest, the agent knows that some sequences of observations are impossible. We already saw in the circuit-diagnosis problem that observing failures was impossible. In the context of update, we know that we cannot observe a person die and then be alive, despite the fact that both being dead and being alive are consistent states.

We can easily maintain what we regard as the defining properties of revision and update, as discussed in the previous section: no change in the environment state and a ranked prior in the case of revision, and a lexicographic prior in the case of update, with belief change proceeding by conditioning in both cases. We simply drop REV3 and replace REV4 by REV4′ (resp., drop UPD3 and replace UPD4 and UPD4′). We remark that this change affects the postulates. For example, consider update. Suppose that the agent considers the possibility that Mr. Bond is dead. If she then observes Mr. Bond alive and well then, according to update, she must account for the new observation by some change from the worlds she previously considered possible. However, there is no transition from worlds in which Mr. Bond is dead that can account for the new observation. Thus, once the agent knows that certain transitions are impossible, some observations (e.g., observing that Mr. Bond is alive) require her to remove from consideration some of the worlds that she previously considered possible. As a consequence, postulate U8 does not hold, since the agent's new beliefs are not determined by a pointwise update at each of the worlds she previously considered possible.

Language of Beliefs. In our analysis of revision and update, we focused on the agent's beliefs about the current state of the environment. Often we are also interested in how the agent changes her beliefs about other types of statements, such as beliefs about future states of the environment, beliefs about other agents' beliefs, and introspective beliefs about her own beliefs. Again, it is straightforward in our framework to deal with an enriched language that lets us express such statements. For example, in [Friedman and Halpern 1994a] we examine *Ramsey conditionals*. These are formulas of the form $\varphi > \psi$, which can be read as saying "after learning φ, the agent believes ψ". This formula can be expressed as $\mathit{learn}(\varphi) \Rightarrow B\psi$ in the language $\mathcal{L}^{KPT}$. As is well known, if belief sets include Ramsey conditionals (and not just propositional formulas), then the AGM postulates become inconsistent (at least, provided we have at least three mutually exclusive consistent formulas in the language) [Gärdenfors 1986]. Similar inconsistency results arise when one tries to add other forms of introspective beliefs [Fuhrmann 1989]. In our setting, it is easy to see why the problem arises. Even if we allow belief sets to include nonpropositional formulas, it still seems quite clear that we want to distinguish the propositional formulas from formulas that talk explicitly about an agent's beliefs. For example, it is not clear that we should allow an observation of a formula such as $\varphi > \psi$. What would it mean to observe such a formula? It clearly seems quite different from observing a propositional formula. Nor does it make sense to extend an assumption such as REV1 to arbitrary formulas. While it may be reasonable to restrict to static propositions if we are viewing these as making statements about a relatively stable environment, it seems far less reasonable to assume that formulas that talk about an agent's beliefs will be static, especially when we are trying to model belief change!

Of course, if we allow only propositional formulas to be learned (or observed), and restrict REV1 to propositional formulas, then it is easy to see that all of our results still hold, even if the full language is quite rich; we avoid the triviality result completely.

Observations One of the strongest assumptions made by revision and update involves the treatment of observations. This assumption seems unreasonable in most domains. REV4 and UPD4 essentially assume that the observation that the agent makes is chosen randomly among all formulas consistent with the current state of the world. Thus, for example, if both φ and ψ describe state of the world that satisfy o_1, then we do not change are relative assessment of these two events after observing o_1. Formally, REV4 requires that $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o_1]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi \wedge o_1])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi; o_1])$. However, this seems somewhat unreasonable in practice. For example, if the agent is standing outdoors, then she is more likely to observe that the sun is shining than that the light is on indoors, even if both are true. Intuitively, the observation that the sun is shining, should increase the plausibility of situations where the agent is outdoors. However, REV4 requires her to maintain the same relative plausibility to both events. Moreover, it may be impossible for the agent to observe that the light is on indoors, but this is not permitted according to REV4 or UPD4.

In many domains it is useful to reason about hidden quantities that simply cannot be observed. For example, the event that component c_i is faulty in Example 5.4.1 is a basic event in our description of the problem, yet it cannot be observed. Similarly, the event where a patient has a disease X or the opponent is planning to capture the queen are useful in reasoning about medical diagnosis and game strategy, yet are not directly observable in practice. Thus, the requirement that all formulas in the language can be observed seems quite unnatural. We note that explicitly modeling sensory input is a standard practice in control theory and stochastic processes (e.g., in hidden Markov chains). In these fields, one models the probability of an observation in various situations. Making an observation increases the probability of situations where that observation is likely to be observation and decreases the probability of situations where it is unlikely. Again, it is straightforward to consider a more detailed model of the observation process in our framework; see Section 6.7.

Actions Our definition of belief change systems essentially assumes that the agent is *passive.* The situation is more complex when the agent can influence the environment. The agent's choice of action interacts with her beliefs. It is clear that after performing an action, the agent should change her beliefs.[7] Moreover, the information content of observations depends on the action the agent has just performed. For example, the agent might consider hearing a loud noise to be surprising. However, it would be expected after the agent pulls the trigger of her gun.

This list of possible extensions is clearly not exhaustive; there are many others that we may want to consider. Nevertheless, these are extensions that seem to be of interest. The main points we want to make here are (1) it is easy to accommodate these extensions in our framework while still maintaining the main characteristics of revision and update, and (2) it is difficult to deal with such extensions if we focus on postulates.

5.8 Conclusion

We have shown how the framework introduced in previous chapters can be used to capture belief revision and update. Modeling revision and update in the framework also gives us a great deal of further insight into their properties, and emphasizes the role of conditioning as a way of capturing minimal change.

Of course, revision and update are but two points in a wide spectrum of possible types of belief change. Our ultimate goal is to use this framework to understand the whole spectrum better and to help us design belief change operations that overcome some of the difficulties we have observed with revision and update. In particular, we want belief change operations that can handle dynamic propositions, while still being able to revise information about the past. We discuss one such proposal in the next chapter.

Finally, we note that our approach is quite different from the traditional approach to belief change [Alchourrón, Gärdenfors, and Makinson 1985; Gärdenfors 1988; Katsuno and Mendelzon 1991a]. Traditionally, belief change was considered as an abstract process. Our framework, on the other hand, models the agent and the environment she is situated in, and how both change in time. This allows us to model concrete agents in concrete settings (for example, our treatment of diagnostic systems), and to reason about the beliefs and knowledge of such agents. We can then investigate what plausibility ordering induces beliefs that match our intuitions. By gaining a better understanding of such concrete situations, we can better investigate more abstract notions of belief change.

[7] Indeed, an alternative interpretation of the update postulates is that they describe how the agent should update her beliefs after doing the action "achieve φ" [Goldszmidt and Pearl 1992; del Val and Shoham 1992; del Val and Shoham 1993]. However, as these works show, the update postulates are problematic under this interpretation.

Chapter 6

Markovian Belief Change

6.1 Introduction

In the previous chapter we examined AGM revision and KM update, two approaches that have received a fair amount of attention in the literature. Recall that revision treats a surprising observation as an indication that previous beliefs were wrong, while update treats a surprising observation as an indication that the external environment has changed. In general, we would expect that an agent making an observation may want both to revise some earlier beliefs and to assume that some change has occurred. To see this, consider again the borrowed car example of Section 5.5. In practice, the agent may want to consider a number of possible explanations for her observation that the gas tank is not full, depending on what she considers to be the most likely sequence(s) of events between time 1 and time 4. For example, if she has had previous gas leaks, then she may consider a gas leak to be the most plausible explanation. On the other hand, if her spouse also has the car keys, she may consider it possible that he used the car in her absence. Is there a reasonable approach that lets us capture these considerations in a natural manner? In this chapter, we show that there is and, indeed, we can get one by applying straightforward ideas from probability theory in a qualitative setting.

To understand our approach, it is helpful to review what a probabilist would do. The standard probabilistic approach would be to put a probability on the runs of the system. This is the agent's prior probability, and captures his initial beliefs about the relative likelihood of runs. As the agent receives information, she changes her beliefs using conditioning.

In the previous chapter we described the classes of prior plausibility measures that give us belief revision and belief update. As we argued, these are fairly special, and do not capture many typical situations. In general, we would like to specify a prior plausibility measure over runs that captures our initial assessment of the relative plausibility of runs. One obvious problem with this approach is that, even if there are only two possible states, there are uncountably many possible runs. How can an agent describe a prior probability over such a complex space?

Again, we turn for intuition to the probabilistic settings. In these settings, the standard solution to this problem is to assume that state transitions are independent of when they occur, that is, that the probability of the system going from state s to state s' is independent of the sequence of transitions that brought the system to state s. This *Markov assumption* significantly reduces the complexity of the problem. All that is necessary is to describe the probability of state transitions. Moreover, the Markov assumption has been shown to be widely applicable in practice [Kemeny and Snell 1960; Howard 1971].

As we show now, we can use (a qualitative analogue of) the Markov assumption to simplify the description of the prior plausibility. Making a (qualitative) Markov assumption gives us a well-behaved notion of belief change, without making the occasionally unreasonable assumptions made by belief revision and update. In particular, it allows a user to weigh the relative plausibility that a given observation is due to a change in the world or due to an inaccuracy in previous beliefs. In the borrowed car example, this means that the agent can decide the relative plausibility of a gas leak and his wife's taking the car, without making the commitment to one or the other, as required by update and revision.

6.2 The Markov Assumption in Probabilistic Reasoning

To reason about a space W probabilistically, we need a probability measure on W. Formally, that means we have a probability space $(W, \mathcal{F}, \mathrm{Pr})$, where $\mathcal{F}$ is an *algebra* of *measurable* subsets of W (that is, a set of subsets closed under union and complementation, to which we assign probability) function mapping each *event* (i.e., a subset of W) in $\mathcal{F}$ to a number in [0, 1], satisfying the well-known Kolmogorov axioms ($\mathrm{Pr}(\emptyset) = 0$, $\mathrm{Pr}(W) = 1$, and $\mathrm{Pr}(A \cup B) = \mathrm{Pr}(A) + \mathrm{Pr}(B)$ if A and B are disjoint).[1]

Probability theory also dictates how we should change our beliefs. If the agent acquires evidence E, her beliefs after learning E are specified by the *conditional probability* $\mathrm{Pr}(\cdot \mid E)$. Note that by using conditioning, we are implicitly assuming that the information E is correct (since we assign $\overline{E}$—the complement of E—probability 0), and that discovering E is telling us only that $\overline{E}$ is impossible; the relative probability of subsets of E is the same before and after conditioning.

We want apply probabilistic reasoning to dynamic systems. To do so, we need to identify the appropriate space W and the events of interest (i.e., $\mathcal{F}$). We take the space W to be a set of sequences of states from some set $\mathcal{S}$. Intuitively, we can think of W as a set of sequences of the form $\{\{r_e(m) : m \geq 1\} : r \in \mathcal{R}\}$; that is, the environment component of some set of runs. There are various events that will be of interest to us; we always assume without comment that the events of interest are in $\mathcal{F}$. One type of event of interest is that the system takes on a particular state at time m. We use $S_m = s$ to denote the event in which the system is in state s at time m.[2] A *time-m* event is a Boolean combination of events of the form $S_i = s$, for $i \leq m$. We are particularly interested in time-m events of the form $(S_0 = s_0) \cap \cdots \cap (S_m = s_m)$, which we abbreviate $[s_0, \ldots, s_m]$; Recall that this is the set of all runs in $\mathcal{R}$ with initial prefix $s_0, \ldots, s_m$. We call such an event an *m-prefix*. Note that any time-m event is a union of m-prefixes.

[1] For ease of exposition, we do not make the requirement that probability distributions be countably additive here.

[2] We can view S_n as a *random variable* over the set $\mathcal{R}$ of runs, that is, a function with domain $\mathcal{R}$ (and, in this case, range $\mathcal{S}$).

As discussed in the introduction, describing a distribution Pr on runs can be difficult. Even when $\mathcal{S}$ contains only two states, W is uncountable. In the probabilistic literature, this difficulty is often dealt with by making assumptions of *conditional independence*. We say that an event A is *conditionally independent* of event C given event B, denoted $\mathit{Ind}_{\Pr}(A; C \mid B)$ if either $\Pr(B, C) = 0$ or $\Pr(A \mid B, C) = \Pr(A \mid B)$. In temporal domains, a useful assumption of independence is the *Markov assumption*.

Definition 6.2.1: A *Markov chain* [Kemeny and Snell 1960] over $S_0, S_1, \ldots$ is a measure Pr that satisfies

- $\mathit{Ind}_{\Pr}(S_{m+1} = s_{m+1}; E \mid S_m = s_m)$, where E is any time-m event,
- $\Pr(S_{m+1} = s' \mid S_m = s) = \Pr(S_{n+1} = s' \mid S_n = s)$.

We say that Pr is a *Markovian* measure if it is a Markov chain over $S_0, S_1, \ldots$. ▮

The first requirement says that the probability of $S_{m+1} = s_{m+1}$ is independent of preceding states given the value of S_m: The probability of going from state $S_m = s_m$ to $S_{m+1} = s_{m+1}$ is independent of *how* the system reached $S_m = s_m$. The second requirement says that the *transition probabilities*, i.e., the probabilities of making a transition from state s to state s', do not depend on the time of the transition. Intuitively, the Markov assumption requires that states in $\mathcal{S}$ summarize all relevant aspects of past history that can affect the behavior of the system in the future. A standard mathematical argument shows that any stochastic process can modeled as a Markov process if we "enrich" the set of states. In practice, many systems can be modeled in a natural way so as to make the Markov assumption applicable.

If we specify the probability of the initial state and the probabilities of subsequent transitions, then there is a unique Markovian measure that is consistent with these transition probabilities. We now make this argument precise. The *transition* probability $p_{s,s'}$ denotes the probability of making a transition from s to s', that is, $\Pr(S_{m+1} = s' | S_m = s)$. To simplify the discussion below, we use a similar notation to describe the probability of the initial state by using a dummy state that we can think of as the unique state of the system before time 0. We denote by $p_{s_\vdash, s}$ the probability that $S_0 = s$, where $s_\vdash$ is some state such that $s_\vdash \notin \mathcal{S}$. If we specify *proper* transition probabilities, that is, $\sum_{s' \in \mathcal{S}} p_{s,s'} = 1$ for all $s \in \mathcal{S}^* = \mathcal{S} \cup \{s_\vdash\}$, then it is easy to show that there is a unique Markovian measure Pr on the algebra generated by events of the form $S_m = s$ such that $\Pr(S_0 = s) = p_{s_\vdash, s}$ and $\Pr(S_{m+1} = s' \mid S_m = s) = p_{s,s'}$. We can define Pr on the m-prefixes by induction on m. The case of $m = 0$ is given by $\Pr([S_0 = s]) = p_{s_\vdash, s}$. For the inductive step, assuming that we have defined $\Pr([s_0, \ldots, s_{m-1}])$, we have

$$\begin{aligned} \Pr([s_0, \ldots, s_m]) &= \Pr(S_m = s_m \mid [s_0, \ldots, s_{m-1}]) \times \Pr([s_0, \ldots, s_{m-1}])) \\ &= p_{s_{m-1}, s_m} \times \Pr([s_0, \ldots, s_{m-1}]. \end{aligned}$$

An easy induction argument now shows

$$\Pr([s_0, \ldots, s_m]) = p_{s_\vdash, s_0} \times p_{s_0, s_1} \times \cdots \times p_{s_{m-1}, s_m}.$$

Since a time-m event is a union of m-prefixes, this shows that Pr is determined for all time-m events. If $\mathcal{S}$ is finite, this gives us a compact representation of the distribution over $\mathcal{R}$.

6.3 Markovian Plausibility Measures

As we said in the introduction, we would like a notion of belief change that allows us both to revise our previous beliefs about the world and to allow for a change in the world occurring. Moreover, we need to address the question of representing the plausibility measure on runs. Can we get measures with reasonable belief-change properties that can be represented in a natural and compact manner? In the probabilistic framework, the Markov assumption provides a solution to both problems. As we now show, it is also useful in the plausibilistic setting.

As we have seen, the Markov assumption is an assumption about conditional independencies. Can we define conditional independence in plausibility measures? It is straightforward to generalize the definition of conditional independence to plausibility measures: A is independent of C given B, denoted $\mathit{Ind}_{\mathrm{Pl}}(A; C \mid B)$ if either $\mathrm{Pl}(B, C) = \perp$ or $\mathrm{Pl}(A \mid B, C) = \mathrm{Pl}(A \mid B)$. Consequently, the definition of Markovian probability measures also generalizes to plausibility measures. A conditional plausibility space $(W, \mathcal{F}, \mathrm{Pl})$ is *Markovian* if it satisfies the same conditions as in Definition 6.2.1, with Pr replaced by Pl. Given a Markovian plausibility space $(W, \mathcal{F}, \mathrm{Pl})$, we define the *transition plausibilities* analogously to the transition plausibilities: that is, $t_{s_\vdash, s} = \mathrm{Pl}(S_0 = s)$ and $t_{s,s'} = \mathrm{Pl}(S_{m+1} = s' \mid S_m = s)$.

In the probabilistic setting, the Markov assumption has many implications that can be exploited. In particular, we can easily show the existence and uniqueness of a Markovian prior with a given set of transition probabilities. Can we get a similar result for Markovian plausibility spaces? In general, the answer is no. To get this property, and the other desirable properties of Markovian plausibility spaces, we need to put more structure on plausibility measures.

In showing that there is a unique Markovian measure determined by the transition probabilities, we made use of two important properties of probability: The first is that $\mathrm{Pr}(A, B)$ is determined as $\mathrm{Pr}(B \mid A) \times \mathrm{Pr}(A)$. (We used this to calculate $\mathrm{Pr}([s_0, \ldots, s_m])$.) The second is that $\mathrm{Pr}(A \cup B) = \mathrm{Pr}(A) + \mathrm{Pr}(B)$, if A and B are disjoint. (This was used to get the probability of an arbitrary time-m event from the probability of the time-m prefixes.) We would like to have analogues of $+$ and $\times$ for plausibility.

To get an analogue of $+$, we need to assume that the plausibility of the union of two disjoint sets A and B is a function of the plausibility of A and the plausibility of B. We call a conditional plausibility measure Pl *decomposable* if there is a function $\bigoplus$ that is commutative, monotonic (so that if $d \le d'$, then $d \bigoplus e \le d' \bigoplus e$), and additive (so that $d \bigoplus \perp = d$ and $d \bigoplus \top = \top$) such that $\mathrm{Pl}(A \cup B \mid C) = \mathrm{Pl}(A \mid C) \bigoplus \mathrm{Pl}(B \mid C)$ if A and B are disjoint. In [Friedman and Halpern 1995], natural conditions are provided that guarantee that a plausibility measure is decomposable. Note that probability measures are decomposable, with $\bigoplus$ being $+$, possibility measures are decomposable with $\bigoplus$ being max, and κ-rankings are decomposable with $\bigoplus$ being min. By way of contrast, Dempster-Shafer belief functions are, in general, not decomposable.

We next want to get an analogue of $\times$, to make sure that conditioning acts appropriately. More precisely, we want there to be a function $\bigotimes$ that is commutative, associative, strictly monotonic (so that if $d > d'$ and $e \neq \perp$, then $d \bigotimes e > d' \bigotimes e$), and multiplicative (so that $d \bigotimes \perp = \perp$ and $d \bigotimes \top = d$) such that $\mathrm{Pl}(A \mid B, C) \bigotimes \mathrm{Pl}(B \mid C) = \mathrm{Pl}(A, B \mid C)$.

An *algebraic domain* is a tuple $\mathcal{D} = (D, \bigoplus, \bigotimes)$ where D is a partially-ordered pointed domain, and $\bigoplus$ and $\bigotimes$ are binary operations on D satisfying the requirements we described above, such that

- $\bigotimes$ distributes over $\bigoplus$,
- D is *rich* with respect to $\bigotimes$, so that if $x \geq y$ then there is unique a z such that $x \bigotimes z = y$.

Note that the assumption of richness implies that $\bigotimes$ is invertible, that is, there is an analogue of division.

A *structured* plausibility space $PL = (W, \mathcal{F}, \mathrm{Pl})$ is one for which there exist $\bigotimes$ and $\bigoplus$ such that $\mathcal{D} = (D, \bigoplus, \bigotimes)$, is an algebraic domain, and Pl is such that $\mathrm{Pl}(A \cup B \mid E) = \mathrm{Pl}(A \mid E) \bigoplus \mathrm{Pl}(B \mid E)$ for disjoint A and B, and $\mathrm{Pl}(A \mid B, C) \bigotimes \mathrm{Pl}(B \mid C) = \mathrm{Pl}(A, B \mid C)$. In this case we say that *PL* is *based* on the algebraic domain $\mathcal{D}$. (We remark that Darwiche [1992] and Weydert [1994b] consider notions similar to structured plausibility spaces; we refer the reader to [Friedman and Halpern 1995] for a more detailed discussion and comparison.)

From now on, we consider only plausibility spaces that are based on algebraic domains. In particular, we use the term *Markovian plausibility measure* to denote a structured plausibility measure that satisfies the Markov assumption.

Algebraic domains are typically used in quantitative notions of uncertainty. For example, $\mathcal{D}_p = ([0, 1], +, \times)$ is used for probability and $\mathcal{D}_\kappa = (\mathbb{N}^*, \min, +)$ is used for κ-rankings.[3] The standard examples of algebraic domains in the literature are totally ordered. However, it is not hard to construct partially-ordered algebraic domains. In fact, we now give a general technique of constructing partially-ordered domains. Let $\mathcal{D} = (D, \bigoplus, \bigotimes)$ be an algebraic domain. Given an integer k, we define the kth power of $\mathcal{D}$ to be the domain $\mathcal{D}^k = (D^k, \bigoplus_k, \bigotimes_k)$ where where $\bigoplus_k$ and $\bigotimes_k$ apply $\bigoplus$ and $\bigotimes$, respectively, in a pointwise manner, and $\langle x_1, \ldots, x_k\rangle \preceq \langle y_1, \ldots, y_k\rangle$ if $x_i \leq y_i$ for $1 \leq i \leq k$. We start by showing that this construction results in an algebraic domain.

Proposition 6.3.1: *If $\mathcal{D}$ be an algebraic domain, then $\mathcal{D}^k$ is also an algebraic domain.*

Moreover, $\mathcal{D}^k$ is clearly a partially-ordered algebraic domain. For example, if $x, y \in D$ are such that $x > y$, then $\langle x, y\rangle$ is incomparable to $\langle y, x\rangle$ in $\mathcal{D}^2$. This construction results in a universal domain for default reasoning.

Theorem 6.3.2: *Let $(W, \mathcal{F}, \mathrm{Pl})$ be a qualitative plausibility space such that $\mathcal{F}$ is finite. Then there is an integer k and a structured plausibility space $(W, \mathcal{F}, \mathrm{Pl}')$ based on $\mathcal{D}^k_\kappa$ such that* Pl′ *is default-isomorphic to* Pl.

In the probabilistic setting, Markovian priors are useful in part because they can be described in a compact way. Similar arguments show that this is the case in the plausibilistic setting as well. We start with a definition. We say that transition plausibilities are *proper* if $\bigoplus_{s' \in \mathcal{S}} t_{s,s'} = \top_D$ for all $s \in \mathcal{S}^*$.

[3] For possibility, $\mathcal{D}_{poss} = ([0, 1], \max, \min)$ is often used; this is not quite an algebraic domain, since min is not strictly monotonic.

Theorem 6.3.3: *Let $\mathcal{D} = (D, \bigoplus, \bigotimes)$ be a algebraic domain, let $\mathcal{S}$ be a finite set of system states, and let $\{t_{s,s'} : s \in \mathcal{S}^*, s' \in \mathcal{S}\}$ be a set of proper transition plausibilities. Then there is a unique structured Markovian plausibility space $(W, \mathcal{F}, \mathrm{Pl})$ based on $\mathcal{D}$ such that $\mathrm{Pl}(S_0 = s) = t_{s_\vdash,s}$ and $\mathrm{Pl}(S_{m+1} = s' \mid S_m = s) = t_{s,s'}$ for all states $s, s' \in \mathcal{S}$ and for all times m.*

Proof: We need to show that once we fix the algebraic domain and transition plausibilities, the plausibility of all events in the algebra generated from events of the form $S_m = s$ are determined.

We start with events of the form $[s_0, \ldots, s_m]$. As in the probabilistic case, using the Markovian assumption and the fact that $\bigotimes$ determines conditioning, we get that

$$\mathrm{Pl}([s_0, \ldots, s_m]) = t_{s_\vdash,s_0} \bigotimes t_{s_0,s_1} \bigotimes \cdots \bigotimes t_{s_{m-1},s_m}.$$

Now let A be some time-m event. Since A is the disjoint union of a finite number of events of the form $[s_0, \ldots, s_m]$, we get that

$$\mathrm{Pl}(A) = \bigoplus_{[s_0,\ldots,s_m] \subseteq A} \mathrm{Pl}([s_0, \ldots, s_m]).$$

Finally, we examine conditional plausibilities. Let A and B be two time-m events. Clearly, $A \cap B$ is also an time-m event. Thus, both $\mathrm{Pl}(A \cap B)$ and $\mathrm{Pl}(B)$ are determined by the transition plausibilities. Since $A \cap B \subseteq B$, we have that $\mathrm{Pl}(A \cap B) \leq \mathrm{Pl}(B)$. Moreover, since $\bigotimes$ is monotonically increasing and D is rich with respect to $\bigotimes$, there is a unique plausibility value $d \in D$ such that $d \bigotimes \mathrm{Pl}(B) = \mathrm{Pl}(A \cap B)$. Thus, we get that $\mathrm{Pl}(A \mid B) = d$. ∎

We note that this result applies only when there is are finitely many states. If there are infinitely many states, then there might be many Markovian plausibility spaces that are consistent with the transition plausibilities. This is not surprising, since our definition of decomposable measures does not put any restrictions on infinite unions. We believe that if we require infinite unions to satisfy fairly simple limit properties, we could state an analogous theorem for a countable number of system states. We defer this extension to future work.

This result states that once we specify a algebraic domain and a set of transition plausibilities, there is a unique Markovian plausibility measure with these transition plausibilities. Since we want to capture belief, we are particularly interested in qualitative plausibility spaces. Thus, it is of interest to identify what choices of algebraic domain and transition plausibilities ensure that the (unique) corresponding Markovian plausibility measure is qualitative. As we now show, if the algebraic domain is totally ordered is is sufficient (and essentially necessary) to require that $\bigoplus$ be max. On the other hand, if the algebraic domain is not totally ordered, then a weak condition ensures that the corresponding plausibility measure is *not* qualitative.

We start with the totally ordered case. Using Proposition 2.8.7 we have that if $\bigoplus$ is max, the corresponding Markovian measure is qualitative.

Proposition 6.3.4: *Let $(D, \bigoplus, \bigotimes)$ be an algebraic domain such that $\leq_D$ is totally ordered and $d \bigoplus d' = \max(d, d')$. Then the plausibility space of Theorem 6.3.3 is qualitative.*

It turns out that this condition is essentially necessary to ensure qualitativeness for totally ordered domains.

Proposition 6.3.5: *Let $(W, \mathcal{F}, \mathrm{Pl})$ be a structured qualitative plausibility measure based on $\mathcal{D} = (D, \bigoplus, \bigotimes)$, and let $d, d' \in D$ be such that $d \geq d'$. If there are pairwise disjoint sets $A, B, C \in \mathcal{F}$ such that $\mathrm{Pl}(A) = \mathrm{Pl}(B) = d$ and $\mathrm{Pl}(C) = d'$, then $d \bigoplus d' = d$.*

Turning to the partially ordered case, we mainly get negative results. These results show that for many seemingly reasonable choices of transition plausibilities the corresponding Markovian measure is not qualitative. To show this, we start with a simple property of structured qualitative plausibility measures:

Lemma 6.3.6: *If $(W, \mathcal{F}, \mathrm{Pl})$ is a qualitative decomposable plausibility space, and $A, B, C, D \in \mathcal{F}$ are pairwise disjoint sets such that $\mathrm{Pl}(A)$ is incomparable to $\mathrm{Pl}(B)$ (i.e., $\mathrm{Pl}(A) \not\leq \mathrm{Pl}(B)$ and $\mathrm{Pl}(B) \not\leq \mathrm{Pl}(A)$) and $\mathrm{Pl}(A) = \mathrm{Pl}(C)$, then $\mathrm{Pl}(B) \neq \mathrm{Pl}(D)$.*

This lemma can be used to show that a qualitative Markovian plausibility measure whose transition plausibilities are only partially ordered must satisfy some strong constraints. We start with a definition. Let $\{t_{s,s'}\}$ be a set of transition plausibilities. We say that a state s' is *reachable* from a state s if there is a sequence $s = s_1, \ldots, s_k = s'$ such that $t_{s_i,s_{i+i}} > \perp$ for all $1 \leq i < k$.

Theorem 6.3.7: *Under the conditions of Theorem 6.3.3, the unique Markovian plausibility measure of Theorem 6.3.3 is* not *qualitative if there is a state $s \in \mathcal{S}$ such that*

- *s is reachable from $s_\vdash$, and*
- *there are two sequences $s_1, \ldots, s_k$ and $s'_1, \ldots, s'_{k'}$ such that $s_1 = s'_1 = s_k = s'_{k'} = s$, neither sequence is a subsequence of the other, and $\bigotimes_{1 \leq i < k} t_{s_i,s_{i+1}}$ is incomparable to $\bigotimes_{1 \leq i < k'} t_{s'_i,s'_{i+1}}$.*

This theorem states that if there is a state s that is reachable from $s_\vdash$ and also from itself, then we have to ensure that all possible ways of reaching s from s have comparable plausibilities. As the following example shows, this constraint is easy to violate.

Example 6.3.8: Consider a Markovian plausibility measure Pl over some algebraic domain $(D, \bigoplus, \bigotimes)$ such that $t_{s_\vdash,s}$, $t_{s,s}$, $t_{u,u}$, $t_{s,u}$ and $t_{u,s}$ are all greater than $\perp$ and $t_{u,u}$ is incomparable to $t_{s,s}$. Clearly s is reachable from $s_\vdash$. Now consider the following two sequences:

$$s, u, u, s \tag{6.1}$$

$$s, s, u, s \tag{6.2}$$

Clearly, neither of these sequences is a subsequence of the other. Let $d = t_{s,u} \bigotimes t_{u,s}$. It is easy to see that the product of transition plausibilities in each sequence is, $d \bigotimes t_{u,u}$ in (6.1), and $d \bigotimes t_{s,s}$ in (6.2). Since $t_{s,s}$ and $t_{u,u}$ are incomparable and $d > \perp$, we have that $d \bigotimes t_{s,u}$ and $d \bigotimes t_{u,u}$ are also incomparable. Using Theorem 6.3.7, we get that Pl is not qualitative. ∎

In conclusion, the requirement that a plausibility measure is both Markovian and qualitative entails strong constraints on the transition plausibilities. Proposition 6.3.4 shows that if the algebraic domain is totally ordered, and $\bigoplus$ is max, then the resulting Markovian plausibility is qualitative. On the other hand, Theorem 6.3.7 shows that we cannot prove similar results for partially ordered algebraic domains. Moreover, this theorem shows that, typically, partially ordered transition plausibilities lead to Markovian plausibility measures that are not qualitative.

In light of these results, our main focus is on totally ordered algebraic domains where $\bigoplus$ is max (although we discuss one approach to dealing with partially ordered transition plausibilities in Section 6.6.) Clearly, any plausibility measure based on such a domains is a ranking. We call such domains *ranked* domains.

6.4 Markovian Belief Change

As in Chapter 5, we focus here on belief change systems (as defined in Section 5.3). Recall that in belief change systems, the environment states are associated with truth assignments to the primitive propositions in some set Φ_e and the agent's states are sequences of observations, where an observation is just a formula in $\mathcal{L}_e$. This implies that along each run, global states of the system are never repeated. Thus, while can define a Markovian prior over global system states, this prior would be of little use.

Instead, we define a Markovian prior on the possible evolutions of the environment. Thus, we let S_m denote the state of the environment at time m. Formally, the event $S_m = s$ consists of all the runs $r \in \mathcal{R}$ such that $r_e(m) = s$. We note that the system might not contain any such runs. Nonetheless, the event $S_m = s$ is well-defined.

Given an algebraic domain and a set of transition plausibilities $\{t_{s,s'} : s \in \mathcal{S}_e^*, s' \in \mathcal{S}_e\}$, we have a unique plausibility measure over the possible evolutions of the environment. This, however, does not determine the prior over runs in the system, since it does not determine the plausibility of what the agent observes. To get a prior on runs, we make the assumption that all possible observations that are consistent with the environment states are equally possible (we reexamine this assumption in Section 6.7). Formally, we capture this assumption using the same requirement as we had for update: UPD4. Recall that UPD4 requires that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0,\ldots,\varphi_{k+1};o_1,\ldots,o_k]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0,\ldots,\psi_{k+1};o_1,\ldots,o_k])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi_0,\varphi_1\wedge o_1,\ldots,\varphi_k\wedge o_k,\varphi_{k+1}]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0,\psi_1\wedge o_1,\ldots,\psi_k\wedge o_k,\psi_{k+1}])$.

We define $\mathcal{C}^M$ to consist of all BCSs that satisfy UPD4 and have a structured Markovian prior. Using Theorem 6.3.3, we can show that given a specification in the form of transition plausibilities, there exists a corresponding BCS in $\mathcal{C}^M$.

Proposition 6.4.1: *Let $\mathcal{S}_e$ be a finite set of environment states, let π_e be a mapping from states in $\mathcal{S}_e$ to truth assignments to Φ_e, let $\mathcal{D} = (D,\bigoplus,\bigotimes)$ be a ranked algebraic domain, and let $\{t_{s,s'} \in D : s \in \mathcal{S}_e^*, s' \in \mathcal{S}_e\}$ be a proper set of transition plausibilities. Then there exists a BCS $\mathcal{I} = (\mathcal{R},\pi,\mathcal{P}) \in \mathcal{C}^M$ with a Markovian prior Pl_a based on $\mathcal{D}$ such that $\mathrm{Pl}_a(S_0 = s) = t_{s_\vdash,s}$ and $\mathrm{Pl}_a(S_{m+1} = s' \mid S_m = s) = t_{s,s'}$ for all states s and s' and all times m.*

While our definition of $\mathcal{C}^M$ does not require the prior to be ranked, we do not have at this time an existence proof for algebraic domains that are not ranked. Note that according to Theorem 6.3.3, if the set of environment states is finite, then for any set of proper transition plausibilities, there is a unique Markovian plausibility measure. However, we cannot guarantee that this measure is qualitative. Moreover, even if this measure over environment states is qualitative, we have to construct a qualitative prior over runs. In the ranked case, this can be done by assigning each run the plausibility determined by the sequence of environment states. In the partially-ordered case, a similar construction violates qualitativeness, since it assigns two run-prefixes that have the same sequence of environment states an equal plausibility. As Lemma 6.3.6 shows, such a plausibility measure cannot be qualitative. At the current time we do not have a solution to this problem.

We now illustrate the Markovian belief change by examining how we deal with the borrowed car example of Section 5.5.

Example 6.4.2: Recall that in the original stolen car story [Kautz 1986], the car is parked at time 0 and at time 3 the owner returns to find it gone. In the variant, the car is parked with a full fuel tank at time 0, at time 2 the owner returns to find it parked where she left it, but at time 3 she observes that the fuel tank is empty.

We model these examples by assuming that Φ_e consists of two propositions: *Parked* denotes that the car is parked, and *Full* denotes that fuel tank is full. To simplify the example, we assume that $\neg \mathit{Full} \wedge \neg \mathit{Parked}$ is not satisfiable in $\mathcal{L}_e$. Thus, we consider three states: s_{pf}, $s_{\mathrm{p}\bar{\mathrm{f}}}$, and $s_{\bar{\mathrm{p}}\mathrm{f}}$ with the obvious truth assignments. For example, in s_{pf}, the car is parked with a full tank and in $s_{\mathrm{p}\bar{\mathrm{f}}}$ the car is parked with an empty tank. Since both variants assume that the car is initially parked with full fuel tank, we assume that $s_0 = s_{\mathrm{pf}}$ in all runs of the system.

In the original story, the evidence at time 3 is captured by $E_{\mathrm{stolen}} = \mathcal{R}[\mathit{Parked} \wedge \mathit{Full}, \mathit{true}, \mathit{true}, \neg \mathit{Parked}]$. Similarly, the evidence in the variant at time 2 is captured by $E^2_{\mathrm{borrowed}} = \mathcal{R}[\mathit{Parked} \wedge \mathit{Full}, \mathit{true}, \mathit{Parked}]$, and at time 3 by $E^3_{\mathrm{borrowed}} = \mathcal{R}[\mathit{Parked} \wedge \mathit{Full}, \mathit{true}, \mathit{Parked}, \mathit{Parked} \wedge \neg \mathit{Full}]$

We now examine one possible Markovian prior for this system. The story suggests that the most likely transitions are the ones where no change occurs. Suppose we attempt to capture this using κ-rankings. Recall that κ-rankings are based on the algebraic domain $(I\!N^*, \min, +)$. We could, for example, take $t_{s,s} = 0$ for all $s \in \mathcal{S}_e$. If we believe that the transition from s_{pf} to $s_{\mathrm{p}\bar{\mathrm{f}}}$ is less likely than the transition from s_{pf} to $s_{\bar{\mathrm{p}}\mathrm{f}}$ and from $s_{\bar{\mathrm{p}}\mathrm{f}}$ to $s_{\mathrm{p}\bar{\mathrm{f}}}$, we can take $t_{s_{\mathrm{pf}}, s_{\mathrm{p}\bar{\mathrm{f}}}} = 3$, $t_{s_{\mathrm{pf}}, s_{\bar{\mathrm{p}}\mathrm{f}}} = 1$, and $t_{s_{\bar{\mathrm{p}}\mathrm{f}}, s_{\mathrm{p}\bar{\mathrm{f}}}} = 1$.

Suppose the agent knows that the car is parked at time 0 but then observes that it is gone at time 3. It is easy to verify that there are exactly three 3-prefixes with plausibility 0 after we condition on E_{stolen}: $[s_{\mathrm{pf}}, s_{\bar{\mathrm{p}}\mathrm{f}}, s_{\bar{\mathrm{p}}\mathrm{f}}, s_{\bar{\mathrm{p}}\mathrm{f}}]$, $[s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\bar{\mathrm{p}}\mathrm{f}}, s_{\bar{\mathrm{p}}\mathrm{f}}]$ and $[s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\bar{\mathrm{p}}\mathrm{f}}]$. Thus, the agent believes that the car was stolen before time 3, but has no more specific beliefs as to when.

Suppose the agent knows that the car is parked and has a full tank at time 0, and then observes that it is still parked at time 2. This amounts to conditioning on E^2_{borrowed}. In this case, the most plausible time-2 prefix is $[s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{pf}}]$; the expected observation that the car is parked does not cause the agent to believe that any change occurred. What happens when she observes that the tank is empty, i.e., conditions on E^3_{borrowed}? There are two possible explanations: Either the gas leaked at some stage (so that there was a transition from s_{pf} to $s_{\mathrm{p\bar{f}}}$ before time 3) or the car was "borrowed" without the agent's knowledge (so that there was a transition from s_{pf} to $s_{\mathrm{\overline{p}f}}$ and then from $s_{\mathrm{\overline{p}f}}$ to $s_{\mathrm{p\bar{f}}}$). Applying the definition of $\bigotimes$ we conclude that $\mathrm{Pl}([s_{\mathrm{pf}}, s_{\mathrm{\overline{p}f}}, s_{\mathrm{p\bar{f}}}]) = 2$ and $\mathrm{Pl}([s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}]) = 3$. Thus, the agent considers the most likely explanation to be that the car was borrowed.

It is worth noticing how the agent's beliefs after conditioning on E^3_{borrowed} depend on details of the transition plausibilities and the $\bigotimes$ function. For the κ-ranking above, we have $t_{s_{\mathrm{pf}},s_{\mathrm{p\bar{f}}}} > t_{s_{\mathrm{pf}},s_{\mathrm{\overline{p}f}}} \bigotimes t_{s_{\mathrm{\overline{p}f}},s_{\mathrm{p\bar{f}}}}$; with this choice, the most likely explanation is $[s_{\mathrm{pf}}, s_{\mathrm{\overline{p}f}}, s_{\mathrm{p\bar{f}}}]$. If, instead, we had used a κ-ranking such that $t_{s_{\mathrm{pf}},s_{\mathrm{p\bar{f}}}} < t_{s_{\mathrm{pf}},s_{\mathrm{\overline{p}f}}} \bigotimes t_{s_{\mathrm{\overline{p}f}},s_{\mathrm{p\bar{f}}}}$ (for example, by taking $t_{s_{\mathrm{pf}},s_{\mathrm{p\bar{f}}}} = t_{s_{\mathrm{pf}},s_{\mathrm{\overline{p}f}}} = t_{s_{\mathrm{\overline{p}f}},s_{\mathrm{p\bar{f}}}} = 1$), then the most likely explanation would have been $\mathrm{Pl}([s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}])$. Finally, if we had used a κ-ranking such that $t_{s_{\mathrm{pf}},s_{\mathrm{p\bar{f}}}} = t_{s_{\mathrm{pf}},s_{\mathrm{\overline{p}f}}} \bigotimes t_{s_{\mathrm{\overline{p}f}},s_{\mathrm{p\bar{f}}}}$ (for example, by taking $t_{s_{\mathrm{pf}},s_{\mathrm{p\bar{f}}}} = 2$ and $t_{s_{\mathrm{pf}},s_{\mathrm{\overline{p}f}}} = t_{s_{\mathrm{\overline{p}f}},s_{\mathrm{p\bar{f}}}} = 1$), then the agent would have considered both explanations possible. ∎

As this example shows, using qualitative Markovian plausibilities, the agent can revise earlier beliefs as revision does (for example, the agent may revise her beliefs regarding whether the car was parked at time 1 once she sees that the fuel tank is no longer full), or she may think that a change occurred in the world that explains her current beliefs (the gas tank leaked). Of course, the agent might also consider both explanations to be possible.

It is interesting to compare the behavior of Markovian plausibility measures in this example to that of Katsuno and Mendelzon's update [1991a]. As Kautz [1986] noted, given E_{stolen}, we should believe that the car was stolen during the period we were gone, but should not have more specific beliefs. Markovian measures give us precisely this conclusion. Update, on the other hand, leads us to believe that the car was stolen just before we notice that it is gone; see Section 5.5. To see this, note that any observation that is implied by the agent's current beliefs does not change those beliefs (this is one of Katsuno and Mendelzon's postulates). Combined with the fact that update never revises beliefs about the past, we must conclude that the agent believes that the car was not stolen at time 1 or 2. In the second variant, the differences are even more significant. Using update, we conclude that there was a gas leak. Update cannot reach the conclusion that the car was borrowed, since that involves changing beliefs about the past: For the agent to consider it possible at time 3 that the car was borrowed, she has to change his belief that the car was parked at time 2. Moreover, update does not allow us to compare the relative plausibility of a gas leak to that of the car being borrowed. (See Section 5.5 for further discussion of update's behavior in this example.)

This discussion shows that Markovian priors are useful for representing plausibility measures compactly. Moreover, they give the user the right level of control: by setting the transition plausibilities, the user can encode her preference regarding various explanations. Markovian priors have computational advantages as well, as we now show.

6.5 Updating Beliefs in Markovian Systems

We now examine what computations are needed in order to perform belief change. We assume here that the agent is interested in keeping track of her beliefs about the current state of the system, that is, maintaining $\text{Bel}(\mathcal{I}, r, m)$. Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ be a BCS with a structured Markovian prior Pl_a. Recall that $\varphi \in \text{Bel}(\mathcal{I}, r, m)$ if $\text{Pl}_{(r,m)}([\![\varphi]\!]_{(r,m)}) > \text{Pl}_{(r,m)}([\![\neg\varphi]\!]_{(r,m)})$. From the definition of conditioning and UPD4, we have that if $\mathcal{I}$ is a BCS that satisfies UPD4, then $\varphi \in \text{Bel}(\mathcal{I}, r, m)$ if and only if

$$\text{Pl}_a(\mathcal{R}[\mathit{true}, o_{(r,1)}, \ldots, o_{(r,m-1)}, o_{(r,m)} \wedge \varphi]) > \text{Pl}_a(\mathcal{R}[\mathit{true}, o_{(r,1)}, \ldots, o_{(r,m-1)}, o_{(r,m)} \wedge \neg\varphi]). \tag{6.3}$$

Let $E_m = \mathcal{R}[\mathit{true}, o_{(r,1)}, \ldots, o_{(r,m)}]$ be the evidence we condition on at time m. It is easy to see that the plausibilities $\text{Pl}_a(S_m = s, E_m)$, $s \in \mathcal{S}$, suffice for determining whether the agent believes φ at time m. This follows from the observation that

$$\text{Pl}_a(\mathcal{R}[\mathit{true}, o_{(r,1)}, \ldots, o_{(r,m-1)}, o_{(r,m)} \wedge \varphi]) = \bigoplus\nolimits_{s \models \varphi} \text{Pl}_a(S_m = s, E_m),$$

and similarly for $\text{Pl}_a(\mathcal{R}[\mathit{true}, o_{(r,1)}, \ldots, o_{(r,m-1)}, o_{(r,m)} \wedge \neg\varphi])$. Given an algebraic domain and the transition plausibilities, the Markov assumption allows us to compute the plausibilities of every m-prefix that is consistent with the evidence; then using $\bigoplus$, we can compute whether φ is more plausible than its negation. When m is large, this procedure is unreasonable. We do not want to examine all m-prefixes in order to evaluate our beliefs about time m. Fortunately, the Markov assumption allows us to maintain our beliefs about the current state of the system in an efficient way.

As we show now, it is straightforward for an agent who maintains a record of the plausibilities $\text{Pl}_a(S_m = s, E_m)$, $s \in \mathcal{S}$, to update them when she makes a new observation. To see this, observe that if $s \not\models o_{(r,m+1)}$, then $(S_{m+1} = s) \cap E_{m+1} = \emptyset$. On the other hand, if $s \models o_{(r,m+1)}$, then $(S_{m+1} = s) \cap E_{m+1} = (S_{m+1} = s) \cap E_m$. Thus,

$$\text{Pl}_a(S_{m+1} = s, E_{m+1}) = \begin{cases} \text{Pl}_a(S_{m+1} = s, E_m) & \text{if } s \models o_{(r,m+1)}, \\ \perp & \text{if } s \not\models o_{(r,m+1)}. \end{cases}$$

In algebraic domains, we can compute $\text{Pl}_a(S_{m+1} = s, E_m)$ using much the same techniques as in probabilistic domains:

$$\text{Pl}_a(S_{m+1} = s, E_m) = \bigoplus\nolimits_{s' \in \mathcal{S}} \text{Pl}_a(S_{m+1} = s \mid S_m = s', E_m) \bigotimes \text{Pl}_a(S_m = s', E_m)$$

Using the Markov assumption, $\text{Pl}_a(S_{m+1} = s \mid S_m = s', E_m) = \text{Pl}_a(S_{m+1} = s \mid S_m = s') = t_{s',s}$. Thus, we can compute $\text{Pl}_a(S_{m+1} = s, E_{m+1})$ using $\text{Pl}_a(S_m = s', E_m)$ and the transition plausibilities $t_{s',s}$.

To summarize, there is a simple two-stage procedure for updating beliefs given new observations. In the first stage, we compute the plausibility of the current states using our beliefs about the previous state. In the second stage, we prune states that are inconsistent with the current observation. It important to note that the arguments we made here are identical to standard probabilistic arguments. Indeed, this procedure is the plausibilistic analogue of a standard algorithm in probabilistic reasoning. In the probabilistic version one maintains $\Pr(S_m \mid E_m)$. This value is computed by *renormalizing* the probabilities after the second step. A similar process can be defined for structured plausibilities, however it is not necessary for the evaluation of beliefs.

Again, it is interesting to compare this approach to Katsuno and Mendelzon's belief update. One of the assumptions made by update is that the agent maintains only her beliefs about the current state of the system. Roughly speaking, this amounts to tracking only the states of the system that have maximal plausibility given past observations. Thus, update can require less information to update beliefs. This, however, comes at the price of assuming that abnormalities are delayed as much as possible. As we saw in the case of the borrowed car example, this may lead to unintuitive conclusions. We conjecture that to avoid such conclusions, the agent must keep track of either information about past events or degrees of plausibilities of all possible states of the system at the current time.

6.6 Partially Specified Transitions

Up to now, we have implicitly assumed that when the user specifies transition plausibilities, she has some underlying algebraic domain in mind. This assumption is quite strong, since it assumes that the user knows the exact plausibility value of the transitions, and the functions $\bigotimes$ and $\bigoplus$ that relate them.

In this section, we focus on situations where the user specifies only a partial order on transitions. This is a natural form of knowledge that can be specified/assessed in a relatively straightforward manner. For example, in the stolen car domain, the user might specify that a transition where the car is stolen, is more plausible then a transition in which a missing car is returned. Moreover, the discussion in the last section focused mainly on ranked transition plausibilities. The techniques we describe here allow us to treat partial comparisons among transition plausibilities as uncertainty as to what is the "true" set of totally ordered transition plausibilities.

We start with a language for representing constraints on the relative plausibility of transitions. We then examine what beliefs must hold, given a sequence of observations, in all Markovian belief change systems where the transition plausibilities are consistent with these constraints. In particular, we examine a sound procedure for determining these beliefs that is based on the properties of algebraic domains.

We represent partial knowledge about transitions as follows. Given a set $\mathcal{S}_e$ of states, consider *transition variables* of the form $x_{s,s'}$, for $s \in \mathcal{S}_e^*$ and $s' \in \mathcal{S}_e$. We think of $x_{s,s'}$ as a variable representing the plausibility of the transition from s to s'. We consider constraints of three kinds on these plausibilities:

- Strict dominance: $x_{s,s'} < x_{u,u'}$
- Dominance: $x_{s,s'} \le x_{u,u'}$[4]
- Impossibility: $x_{s,s'} \le \perp$.

Assume that we are given a set C of constraints on these transition variables. These can be thought of as constraints on the relative plausibility of transitions, together with constraints saying that some

[4]Of course, we can define equality constraints $x = y$ to be an abbreviation for $x \le y$ and $y \le x$.

transitions are impossible, i.e., $x_{s,s'} = \bot$. A *plausibility assignment* to $x_{s,s'}$ is a function $a(\cdot)$ that maps transition variables to some set D of plausibility values. We say that a plausibility assignment a is *consistent* with C if all of the constraints in C are satisfied by the assigned plausibilities; that is, $a(\bot) = \bot_D$, $a(\top) = \top_D$, $a(y) \leq_D a(y')$ for all $y \leq y'$ in C, and $a(y) <_D a(y')$ for all $y < y'$ in C. We are particularly interested in plausibility assignments that lead to proper transition plausibilities. An assignment a is *proper* if $\bigoplus_{s'} a(x_{s,s'}) = \top$ for all $s \in \mathcal{S}_e^*$. Finally, we want to limit our attention to qualitative Markovian plausibilities. This requires us to examine transition plausibilities that generate qualitative plausibility measures. In light of the discussion in Section 6.3, we focus on plausibility assignments onto ranked algebraic domains.

We say that a set of constraints C is *safe* if there some ranked algebraic domain $\mathcal{D} = (D, \bigoplus, \bigotimes)$ and a consistent and proper plausibility assignment a whose range is D. Given a safe set of constraints we define two relations over transitions variables: $y \leq_C y'$ and $y <_C y'$. These hold if $a(y) \leq a(y')$ (resp. $a(y) < a(y')$) for all consistent and proper plausibility assignments a onto a ranked domain. In Appendix E, we provide a characterization of safe sets of constraints and a sound and complete set of inference rules for them.

If a set of constraint C is safe, then we can find a consistent qualitative Markovian plausibility measure.

Theorem 6.6.1: *Let C be a set of constraints on the transitions over a finite set $\mathcal{S}_e$ of states. If C is safe, then there is a qualitative structured Markovian plausibility space $(W, \mathcal{F}, \mathrm{Pl})$ consistent with C.*

Proof: Since C is safe, there is some ranked algebraic domain $\mathcal{D} = (D, \bigoplus, \bigotimes)$, and a consistent and proper plausibility assignment a onto $\mathcal{D}$. Applying Theorem 6.3.3, with the set of transition plausibilities $\{t_{s,s'} = a(x_{s,s'})\}$, we get that there is a Markovian plausibility space $(W, \mathcal{F}, \mathrm{Pl})$ based on $\mathcal{D}$ with these transition plausibilities. Thus, Pl is consistent with C. Moreover, using Proposition 6.3.4 we get that Pl is qualitative. ∎

Unlike the situation described in Section 6.3, where transition plausibilities were fully specified, in general, there are several distinct plausibility measures Pl consistent with C. This is due to the fact that we can choose various plausibility values as well as various operators $\bigotimes$ and $\bigoplus$ in the algebraic domain, while remaining consistent with C. To see this, consider the following variant of Example 6.4.2.

Example 6.6.2: Using the notation of Example 6.4.2, consider the constraints $x_{s_{\mathrm{pf}},s_{\mathrm{p\bar{f}}}} < x_{s_{\mathrm{pf}},s_{\mathrm{\bar{p}f}}} = x_{s_{\mathrm{\bar{p}f}},s_{\mathrm{p\bar{f}}}} < x_{s_{\mathrm{pf}},s_{\mathrm{pf}}} = x_{s_{\mathrm{\bar{p}f}},s_{\mathrm{\bar{p}f}}} = x_{s_{\mathrm{p\bar{f}}},s_{\mathrm{p\bar{f}}}}$. One way of satisfying these constraints is by using the standard κ-ranking described in Example 6.4.2, for which we have $t_{s_{\mathrm{pf}},s_{\mathrm{p\bar{f}}}} = 3$, $t_{s_{\mathrm{pf}},s_{\mathrm{\bar{p}f}}} = t_{s_{\mathrm{\bar{p}f}},s_{\mathrm{p\bar{f}}}} = 1$, and $t_{s,s} = 0$ for all states s. If Pl is the plausibility function generated by this κ-ranking, we have $\mathrm{Pl}([s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}]) = 3$ and $\mathrm{Pl}([s_{\mathrm{pf}}, s_{\mathrm{\bar{p}f}}, s_{\mathrm{p\bar{f}}}]) = 2$, since $\bigotimes$ is $+$ for κ-rankings.

However, there is another way of satisfying the same constraints, again using κ-rankings. Let Pl' be a Markovian plausibility measure with the following transition plausibilities: $t'_{s_{\mathrm{pf}},s_{\mathrm{p\bar{f}}}} = 3$,

$t'_{s_{\mathrm{pf}},s_{\overline{\mathrm{pf}}}} = t'_{s_{\overline{\mathrm{pf}}},s_{\mathrm{p\bar{f}}}} = 2$, and $t'_{s,s} = 0$ for all states s. It is easy to verify that Pl$'$ satisfies the constraints we described above. However, it is easy to check that we have Pl$'([s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}]) = 4$ and Pl$'([s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}]) = 3$. This means that Pl$([s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}]) >$ Pl$'([s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}])$, while Pl$([s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}]) <$ Pl$([s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}])$. Thus, although Pl and Pl$'$ satisfy the same constraints on transition plausibilities, they lead to different orderings of the 3-prefixes, and thus to different notions of belief change. ∎

As this example shows there might be several Markovian plausibility measures that are consistent with a set C of constraints. Thus, there is no unique Markovian belief change system that corresponds to C. Nevertheless, we might ask what beliefs the agent must have if her transition plausibilities are consistent with C. We address this question by examining the consequence of constraints on the relative order of transition plausibilities on the relative plausibility of m-prefixes. Since all time-m events are unions of m-prefixes, such constraints may allow us to determine that some time-m are more plausible than others.

We start with some notation. Let C be a set of constraints on transition variables, and let Pl be a Markovian measure consistent with these constraints. As we observed, Pl$([s_0, \ldots, s_m]) = t_{s_\vdash,s_0} \otimes \cdots \otimes t_{s_{m-1},s_m}$ once we fix the transition plausibilities and $\otimes$. Thus, the plausibility of a sequence is determined by the transitions involved. Moreover, we know that $\otimes$ is commutative and monotonic and multiplicative. Combining the constraints on the relative plausibility of transitions and the properties of $\otimes$, we can derive constraints on the relative plausibility of m-prefixes. We define two relations on m-prefixes:

- $[s_0, s_1, \ldots, s_m] \preceq_C [s'_0, \ldots, s'_m]$ if there is an i such that $x_{s_{i-1},s_i} \leq_C \perp$ (where we take s_{-1} and s'_{-1} to be $s_\vdash$), or if there is some permutation σ over $\{0, \ldots, m\}$ such that $x_{s_{i-1},s_i} \leq_C x_{s'_{\sigma(i)-1},s'_{\sigma(i)}}$, for all $i = 0, \ldots, m$.
- $[s_0, s_1, \ldots, s_m] \prec_C [s'_0, \ldots, s'_m]$ if there is an i such that $x_{s_{i-1},s_i} \leq_C \perp$ and $\perp <_C x_{s'_{j-1},s'_j}$ for all $j \leq m$, or if there is some permutation σ over $\{0, \ldots, m\}$ such that there is an i with $x_{s_{i-1},s_{i+1}} <_C x_{s'_{\sigma(i)-1},s'_{\sigma(i)}}$, and $x_{s_{i-1},s_i} \leq_C x_{s'_{\sigma(i)-1},s'_{\sigma(i)}}$ for all $i = 0, \ldots, m$.

We define $\approx_C$ using $\preceq_C$ in the standard manner. If $[s_0, \ldots, s_m] \approx_C [s'_0, \ldots, s'_m]$ we say these two m-prefixes are *equivalent*. Intuitively, $\preceq_C$ and $\prec_C$ capture what is forced by all Markov plausibility measures consistent with C, since these relations captures what is true for all choices of structured plausibility spaces.

Theorem 6.6.3: *Let C be a set of constraints and let* Pl *be a structured Markovian plausibility measure consistent with C. If $[s_0, \ldots, s_m] \preceq_C [s'_0, \ldots, s'_m]$, then* Pl$([s_0, \ldots, s_m]) \leq$ Pl$([s'_0, \ldots, s'_m])$. *Moreover, if $[s_0, \ldots, s_m] \prec_C [s'_0, \ldots, s'_m]$, then* Pl$([s_0, \ldots, s_m]) <$ Pl$([s'_0, \ldots, s'_m])$.

The converse to Theorem 6.6.3 does not hold in general. A Markovian plausibility measure consistent with C may introduce more comparisons between m-prefixes than those determined by the $\preceq_C$ and $\prec_C$ orderings.

Example 6.6.4: Consider again the setup of Example 6.4.2, using the constraints on transitions considered in Example 6.6.2. Applying the definition of $\preceq$, we find that $[s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}] \prec_C [s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{pf}}]$ and similarly $[s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}] \prec_C [s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{pf}}]$. It is easy to verify that the plausibility measure Pl described in Example 6.6.2 satisfies these constraints. However, $[s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}]$ is incomparable to $[s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}]$ according to $\preceq_C$, but, as we calculated in Example 6.6.2, we have $\mathrm{Pl}([s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}]) < \mathrm{Pl}([s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}])$. ∎

Theorem 6.6.3 shows that $\preceq_C$ and $\prec_C$ are sound with respect to the set of Markovian plausibility measures that are consistent with C. It would be interesting to show that these are all the constraints we deduce about this set of Markovian plausibility measures. We leave this question for future work.

Using Theorem 6.6.3, we can deduce the relative plausibilities of some m-prefixes. Based on these consequences, can use conclude what beliefs must hold in any Markovian measure consistent with C.

We start with a preliminary definition. Let E be some event. Define

$$\mathit{MAX}^m(E) =_{\mathrm{def}} \bigcup\{[s_0, \ldots, s_m] \subseteq E \mid \forall [s'_0, \ldots, s'_m] \subseteq E,\, [s_0, \ldots, s_m] \not\prec_C [s'_0, \ldots, s'_m]\}.$$

$\mathit{MAX}^m(E)$ is the union of m-prefixes in E that are maximal according to $\prec_C$. It easily follows from axiom A2 and Theorem 6.6.3 that in any qualitative Markovian measure Pl consistent with C, $\mathrm{Pl}(\mathit{MAX}^m(E)) > \mathrm{Pl}(E - \mathit{MAX}^m(E))$. In particular, if we take $E = \mathcal{R}[\mathit{true}, o_{(r,1)}, \ldots, o_{(r,m)}]$ and $\mathit{MAX}^m(E) \subseteq \mathcal{R}[\mathit{true}, o_{(r,1)}, \ldots, o_{(r,m-1)}, o_{(r,m)} \wedge \varphi]$, we get that $\mathrm{Pl}(\mathcal{R}[\mathit{true}, o_{(r,1)}, \ldots, o_{(r,m-1)}, o_{(r,m)} \wedge \varphi]) > \mathrm{Pl}(\mathcal{R}[\mathit{true}, o_{(r,1)}, \ldots, o_{(r,m-1)}, o_{(r,m)} \wedge \neg\varphi])$. As a consequence, we have that φ is believed at (r, m). We now make this statement precise:

Theorem 6.6.5: *Let C be a safe set of constraints over $\mathcal{S}_e$, let $\mathcal{I} \in \mathcal{C}^M$ be a BCS system with a structured Markovian prior plausibility that is consistent with C, and let $s_a = \langle o_1, \ldots, o_m\rangle$ be a local state. If $s_m \models \varphi$ for all $[s_0, \ldots, s_m] \subseteq \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$, then $\varphi \in \mathit{Bel}(\mathcal{I}, s_a)$.*

Thus, by examining the most plausible (according to $\prec_C$) m-prefixes, we get a sufficient condition for a formula φ to be believed. The converse to Theorem 6.6.5 does not hold: the agent might believe φ even if some of the m-prefixes in $\mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$ do not satisfy φ. However, if the m-prefixes in $\mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$ are equally plausible, then the converse does hold.

Proposition 6.6.6: *Let C be a safe set of constraints over $\mathcal{S}_e$, and let $\mathcal{I} \in \mathcal{C}^M$ be a ranked Markovian system with a prior plausibility that is consistent with C, and let $s_a = \langle o_1, \ldots, o_m\rangle$ be a local state. If all the m-prefixes in (i.e., that are subsets of) $\mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$ are equivalent according to $\approx_C$, then $\varphi \in \mathit{Bel}(\mathcal{I}, s_a)$ if and only if $s_m \models \varphi$ for all $[s_0, \ldots, s_m] \subseteq \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$.*

Example 6.6.7: We now examine the setup of Example 6.4.2 using partially specified transition plausibilities. To capture our intuition that changes are unlikely, we require that $x_{s,s} \le x_{s',s''}$ for all $s, s', s'' \in \{s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}, s_{\overline{\mathrm{pf}}}\}$. We also assume that $x_{s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}}$, $x_{s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}}$, and $x_{s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}}$ are all strictly less likely than $x_{s_{\mathrm{pf}}, s_{\mathrm{pf}}}$, but are not comparable to each other. Finally, we assume all other transitions are impossible.

Suppose we get the evidence E_{stolen}, that the car is parked at time 0 but gone at time 3. It is easy to verify that $\mathit{MAX}^3(E_{\mathrm{stolen}})$ consists of the three 3-prefixes $[s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}, s_{\overline{\mathrm{pf}}}, s_{\overline{\mathrm{pf}}}]$, $[s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}, s_{\overline{\mathrm{pf}}}]$, and $[s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}]$. Moreover, it easy to check that these three 3-prefixes are equivalent. From Proposition 6.6.6, it now follows that if Pl is a qualitative Markovian measure, then $\mathit{MAX}^3(E_{\mathrm{stolen}})$ characterizes the agent's beliefs. The agent believes that the car was stolen before time 3, but has no more specific beliefs as to when. This proves that all Markovian priors that are consistent with C generate the same beliefs as the κ-ranking described in Example 6.6.2, given this observation.

Suppose we instead get the evidence E^2_{borrowed}, that the car is parked and has a full tank at time 0, and is still parked at time 2. In this case, we have $\mathit{MAX}^2(E^2_{\mathrm{borrowed}}) = [s_{\mathrm{pf}}, s_{\mathrm{pf}}, s_{\mathrm{pf}}]$—the expected observation that the car is parked does not cause the agent to believe that any change occurred. What happens when she observes that the tank is empty, i.e., E^3_{borrowed}? As noted in Example 6.4.2, there are two possible explanations: Either the gas leaked or the car was "borrowed". Without providing more information, we should not expect the agent to consider one of these cases to be more plausible than the others. Indeed, it is easy to verify that $\mathit{MAX}^3(E^3_{\mathrm{borrowed}})$ contains all the 3-prefixes that correspond to these two explanations. Unlike the previous case, some of these 3-prefixes are not equivalent. Thus, different qualitative Markovian plausibility measures may lead to different beliefs at time 3, even if they are consistent with the specifications.

If we add the further constraint $x_{s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}} < x_{s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}}$ (so that a gas leak is more plausible than a theft), then the most likely 3-prefixes are the ones where there is a gas leak. However, if we specify that $x_{s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}}$ is less likely than both $x_{s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}}$ and $x_{s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}}$ (so that a gas leak is less plausible than the car being taken or returned), then $\mathit{MAX}^3(E^3_{\mathrm{borrowed}})$ will still contain both explanations. Even with this additional information, $\preceq$ cannot compare $[s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}, s_{\mathrm{p\bar{f}}}]$ to $[s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}, s_{\mathrm{p\bar{f}}}, s_{\mathrm{p\bar{f}}}]$, because although $t_{s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}} < t_{s_{\mathrm{pf}} \cdot s_{\overline{\mathrm{pf}}}}$ and $t_{s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}} < t_{s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}}$, it might be that $t_{s_{\mathrm{pf}}, s_{\mathrm{p\bar{f}}}} \otimes t_{s_{\mathrm{pf}}, s_{\mathrm{pf}}} \not< t_{s_{\mathrm{pf}}, s_{\overline{\mathrm{pf}}}} \otimes t_{s_{\overline{\mathrm{pf}}}, s_{\mathrm{p\bar{f}}}}$. Our specification, in general, does not guarantee that one explanation is preferred to the other. ∎

As this example shows, there are nontrivial consequences of restrictions on the agent's transition plausibility. Of course, given very few constraints on the transition plausibilities there is little we can infer about the beliefs the agent must hold.

These results suggest one way of using partial ordering over transition plausibilities. We simply consider all ranked plausibilities that are consistent with our partial comparisons. The results we presented here suggest one method of computing the agent's beliefs after conditioning by some evidence.

As the last example shows, the language of constraint cannot capture constraints that compare combinations of transitions. Of course, we can enrich the language of constraints over transition variables and construct analogue of $\prec_C$ for more expressive sets of constraints.

6.7 Modeling Observations

In our discussion of belief change in the previous sections we made the strong assumption that the agent's prior satisfies UPD4, which we also made in our analysis of update. A similar assumption (REV4) was made in the analysis of revision. Recall that this assumption implies that all the observations that are consistent with the environment state are "equally" plausible. As discussed in Section 5.7, this assumption is clearly unrealistic. In many situations there are observations that thrust themselves on us while others are very unlikely to occur. For example, a loud and sudden noise is unlikely to pass unobserved. Thus, a more realistic model would also model the plausibility of various observations in particular situations.

We now outline how to extend the construction of Markovian priors to relax this assumptions, using intuitions from probabilistic reasoning.

Consider the following variant of the stolen car example: the agent parks the car in front of her house at time 0; at time 1 the agent is working in the front yard and does not make an observation about the car; at time 2 the agent is inside the house; finally, at time 3 the agent observes that the car is gone. We assume that the agent's front yard has a tall hedge that separates it from the street. Thus, the plausibility of the car being stolen is the same whether the agent is in the front yard or inside the house. On the other hand, if the agent is in the front yard, she is more likely to hear the car been stolen (for example, if the thief has to break a window or force the lock).

The sequence of observations in this example is identical to the sequence of observations in the original stolen car example. However, different conclusions are warranted. Our assumptions imply that the plausibility of the environment states where the car is stolen at time 1 and these where it is stolen at time 2 are identical. However, the plausibility of this sequence of observations given the two sequences of environment states is different. Since at time 1 the agent was in the yard, it is implausible that she have not noticed that the car was being stolen. Thus, it seems unreasonable to conclude that it is equally plausible that the car was stolen at time 1 or 2. Instead, we would like to get that the run where the car was stolen at time 2 is more plausible.

To capture this example formally, we need to explicitly model the plausibility of different observations at different states. More precisely, we wish to model the conditional plausibility of observations given the state of the environment—$\mathrm{Pl}(O_1 = o_1, \ldots, O_m = o_m \mid [s_0, \ldots, s_m])$, where the random variable O_m denotes the observation made by the agent at time m. To simplify the following discussion, we extend our notation for run prefixes. We denote by $[s_0, \ldots, s_m; o_1, \ldots, o_k]$ the set of runs where $r_e(i) = s_i$ for $0 \leq i \leq m$ and $o_{(r,i)} = o_i$ for $1 \leq i \leq k$.

Such a model allows us to compute the posterior plausibility of different state formulas given the agent's state, i.e., a sequence of observations. To see this, let $E_m = [\cdot; o_{(1)}, \ldots, o_{(m)}]$ be the evidence encoded by the agent's state at time m, and let A and B be two events defined over $S_1, \ldots, S_m$. Recall that $\mathrm{Pl}(A \mid E_m) > \mathrm{Pl}(B \mid E_m)$ if and only if $\mathrm{Pl}(A \cap E_m) > \mathrm{Pl}(B \cap E_m)$. Using the definition of $\otimes$ we have that $\mathrm{Pl}(A \cap E_m) = \mathrm{Pl}(A) \otimes \mathrm{Pl}(E_m \mid A)$, and similarly $\mathrm{Pl}(B \cap E_m) = \mathrm{Pl}(B) \otimes \mathrm{Pl}(E_m \mid B)$. We conclude that

$$\mathrm{Pl}(A \mid E_m) > \mathrm{Pl}(B \mid E_m) \text{ if and only if } \mathrm{Pl}(E_m \mid A) \otimes \mathrm{Pl}(A) > \mathrm{Pl}(E_m \mid B) \otimes \mathrm{Pl}(B)$$

This establishes an analogue of Bayes rule for conditional plausibilities: A is more plausible than B given the evidence E_m if the prior plausibility of A combined with the plausibility of the evidence given A is more plausible than the prior of B combined with the plausibility of the evidence given B. Thus, the agent's beliefs after observing E_m depend not only on the plausibility of various runs of the environment, but also on the plausibility of observing this particular sequence of observations in these runs.

Note that the assumption UPD4 that we have made previously, essentially implies that $\mathrm{Pl}(E_m \mid A) = \mathrm{Pl}(E_m \mid B)$ for all events A and B that are consistent with E_m. Thus, when we assume UPD4, we can compare $\mathrm{Pl}(A|E_m)$ to $\mathrm{Pl}(B|E_m)$ based on $\mathrm{Pl}(A)$ and $\mathrm{Pl}(B)$. In the general case we also have to take into account the plausibility of the observations in each event.

Again, we are faced with a representation problem: there are uncountably many sequences of observations and environment states. Thus, we cannot hope to describe $\mathrm{Pl}([s_0, \ldots, s_m; o_1, \ldots, o_m])$ for each possible combination of environment states and observations. Instead we make independence assumptions that allow for a compact representation of the plausibility $\mathrm{Pl}([s_0, \ldots, s_m; o_1, \ldots, o_m])$. As before, we start with a probabilistic model of a *partially observable Markov chain* and generalize it to plausibility.

Definition 6.7.1: A *partially observable Markov chain* (or *hidden Markov chain*) over $S_0, S_1, \ldots$ and $O_1, O_2, \ldots$, is a plausibility measure Pl on that satisfies

- Pl is a Markov chain on $S_0, S_1, \ldots$, and
- $\mathit{Ind}_{\mathrm{Pl}}(O_m = o; E \mid S_m = s)$, where E is any time-$(m-1)$ event (i.e., an event described by $S_0, \ldots, S_{m-1}$ and $O_1, \ldots, O_{m-1}$), and
- $\mathrm{Pl}(O_m = o \mid S_m = s) = \mathrm{Pl}(O_n = o \mid S_n = s)$ for all m and n.

∎

The first requirement says that the marginal of Pl on $S_0, S_1, \ldots$ is Markovian. Thus, as before, we assume that future changes of the environment depend only on the current state and not on how it was reached. The second requirement says that the plausibility of $O_m = o$ is independent of preceding states and observations given that $S_m = s$. This implies that the observation at time m is independent of how the system reached state s and also of what the agent observed before time m. Finally, the last requirement says that the plausibility of observing o in state s does not depend on the current time.

Clearly, the assumptions we make about the observations might be too simplistic. However, it turns out that these assumptions allow us to model many domains in a useful way. One possible alternative is to assume that O_m depends on S_{m-1} and S_m. Intuitively, this assumes that observations are a function of the changes that took place between time $m-1$ and m. This variant is slightly more complex, but can be dealt with by straightforward modifications to the approach discussed here.

As with Markovian plausibility measures, we have that the plausibility of a run can be described using basic quantities. We define the *extended transition plausibilities* to consist of transition plausibilities of the form $t_{s,s'}$ and *observation plausibilities* of the form $t_{s,o}$, where $\mathrm{Pl}(O_m = o \mid S_m = s) = t_{s,o}$. As with Markovian plausibility measures, assuming that Pl is defined over an algebraic domain, we can define the plausibility of m-prefixes, in this case including observations, as a function of the extended transition plausibilities. The argument is a straightforward extension of the argument given for the Markovian case. Using the definition of partially observable Markovian measures, we get that

$$\begin{array}{rcl}\mathrm{Pl}([s_0,\ldots,s_m;o_1,\ldots,o_m]) & = & \mathrm{Pl}(O_m = o_m \mid S_m = s_m, [s_0,\ldots,s_{m-1};o_1,\ldots,o_{m-1}]) \otimes \\ & & \mathrm{Pl}(S_m = s_m \mid [s_0,\ldots,s_{m-1};o_1,\ldots,o_{m-1}]) \otimes \\ & & \mathrm{Pl}([s_0,\ldots,s_m;o_1,\ldots,o_{m-1}] \\ & = & t_{s_m,o_m} \otimes t_{s_{m-1},s_m} \otimes \mathrm{Pl}([s_0,\ldots,s_m;o_1,\ldots,o_{m-1}]\end{array}$$

By induction on m we conclude that

$$\mathrm{Pl}([s_0,\ldots,s_m;o_1,\ldots,o_m]) = t_{s_m,o_m} \otimes t_{s_{m-1},s_m} \otimes \cdots \otimes t_{s_1,o_1} \otimes t_{s_0,s_1} \otimes t_{s_\vdash,s_0}.$$

This suggests that, given a description of a partially observable Markovian process in terms of the transition plausibilities and observation plausibilities, we can construct a system with a partially observable Markovian prior. Again, we have to ensure that the observation plausibilities add up to $\top$. Additionally, we have to require that only observations that are consistent with the environment state are assigned positive plausibility. We say that a set of observation plausibilities are *proper* if $\bigoplus_o t_{s,o} = \top$ for all $s \in \mathcal{S}_e$ and if $t_{s,o} = \bot$ for all states s and observations o such that $s \not\models o$.

Proposition 6.7.2: *Let $\mathcal{S}_e$ be a finite set of environment state, let $(D, \bigoplus, \bigotimes)$ be a ranked algebraic domain, let $\{t_{s,s'} \in D : s \in \mathcal{S}_e^*, s' \in \mathcal{S}_e\}$ be a set of proper transition plausibilities, and let $\{t_{s,o} \in D : s \in \mathcal{S}_e, o \in \mathcal{L}_e\}$ be a set of proper observation plausibilities. Then there exists a BCS $\mathcal{I}$ with a structured prior Pl_a based on D such that Pl_a partially observable Markov chain, and $\mathrm{Pl}_a(S_0 = s) = t_{s_\vdash,s}$, $\mathrm{Pl}_a(S_{m+1} = s' \mid S_m = s) = t_{s,s'}$, and $\mathrm{Pl}(O_m = o \mid S_m = s) = t_{s,o}$ for all states s and s', all observations o, and all times m.*

Proof: The proof is similar to the proof of Proposition 6.4.1, and we do not repeat it. ∎

Using this machinery, we can describe a prior plausibility by specifying an algebraic domain, the initial plausibilities of states, the transition plausibilities for pairs of states, and the observation plausibilities for pairs of a state and an observation. This allows us to handle the variant of the stolen car example described in the beginning of this section.

Example 6.7.3: We now model the stolen car example when observations have different plausibilities. There are two propositions of interest: *Stolen*, which denotes that the car is stolen, and *Yard*, which denotes that the agent is in the front yard. The possible states are $s_{\overline{\mathrm{s}}\mathrm{y}}$, $s_{\overline{\mathrm{sy}}}$, s_{sy}, $s_{\mathrm{s}\overline{\mathrm{y}}}$ with the obvious truth assignment associated with each state. We assume that the system $\mathcal{R}$ is such that the agent knows whether she is in the yard. As in the previous examples we use κ-rankings. We define transition plausibilities that capture the intuition that the car is more likely to stay parked than disappear: $t_{s,s'} = 0$ if $s \models$ *Stolen* and $s' \models$ *Parked*, or if $s \models \neg$*Stolen* and $s' \models \neg$*Stolen*, and $t_{s,s'} = 2$ if $s \models \neg$*Stolen* and $s' \models$ *Stolen*. Now we also have to specify the plausibility of the possible observations at each state. According to the example, it likely that the agent observes that the car is stolen when she is in the front yard. Thus, we set $\mathrm{Pl}(O_m = \mathit{Stolen} \mid S_m = s_{\mathrm{sy}}) = 0$, and $\mathrm{Pl}(O_m = \mathit{true} \mid S_m = s_{\mathrm{sy}}) = 1$ (recall that *true* is the uninformative observation). Similarly, it is unlikely that the agent observes that the car is stolen (if it is stolen) when she in inside the house. Thus, we set $\mathrm{Pl}(O_m = \mathit{Stolen} \mid S_m = s_{\mathrm{s\bar{y}}}) = \mathrm{Pl}(O_m = \neg\mathit{Stolen} \mid S_m = s_{\overline{\mathrm{sy}}}) = 2$, $\mathrm{Pl}(O_m = \mathit{true} \mid S_m = s_{\mathrm{s\bar{y}}}) = \mathrm{Pl}(O_m = \mathit{true} \mid S_m = s_{\overline{\mathrm{sy}}}) = 0$. Finally, when the agent is in the front yard and the car is there, she may or may not notice it: $\mathrm{Pl}(O_m = \neg\mathit{Stolen} \mid S_m = s_{\mathrm{\bar{s}y}}) = \mathrm{Pl}(O_m = \mathit{true} \mid S_m = s_{\mathrm{\bar{s}y}}) = 0$.

With these transition and observation plausibilities, it is easy to verify that the agent believes that the car was parked at time 2 and was gone between times 2 and 3. To see this, we compute the plausibility of each of the three possible runs: $\mathrm{Pl}([s_{\mathrm{\bar{s}y}}, s_{\mathrm{sy}}, s_{\mathrm{s\bar{y}}}, s_{\mathrm{sy}}; \mathit{true}, \mathit{true}, \mathit{Stolen}]) = 3$, $\mathrm{Pl}([s_{\mathrm{\bar{s}y}}, s_{\mathrm{\bar{s}y}}, s_{\mathrm{s\bar{y}}}, s_{\mathrm{sy}}; \mathit{true}, \mathit{true}, \mathit{Stolen}]) = 2$, and $\mathrm{Pl}([s_{\mathrm{\bar{s}y}}, s_{\mathrm{\bar{s}y}}, s_{\overline{\mathrm{sy}}}, s_{\mathrm{sy}}; \mathit{true}, \mathit{true}, \mathit{Stolen}]) = 2$. In particular, $\mathrm{Pl}(\mathcal{R}[\mathit{true}, \neg\mathit{Stolen}; \mathit{true}, \mathit{true}, \mathit{Stolen}]) > \mathrm{Pl}(\mathcal{R}[\mathit{true}, \mathit{Stolen}; \mathit{true}, \mathit{true}, \mathit{Stolen}])$. ∎

As this example shows, this model allows us to capture situations that we could not capture using systems that satisfy UPD4.

We note that it is straightforward to generalize the results of Sections 6.5 and 6.6 to partially observable Markovian priors. As in Section 6.5, we need only to maintain the plausibility of events of the form $(S_m = s) \cap E_m$ in order to compute the agent's beliefs at time m. The analysis of Section 6.6 can be generalized to include constraints on relative plausibilities of observations. Since the extensions of the results of both sections is straightforward, we do not repeat these results here.

In our discussion so far, we required that $t_{s,o} = \perp$ if $s \not\models o$. This restriction is implied by BCS5, the requirement that observations are always true of the environment state. However, if we remove that requirement, we can also deal with "misleading" observations that are not necessarily true of the environment state. This allows us to reason about whether we believe φ after we observe it. For example, suppose we observe, via a radio report, that Martians have landed on earth. We can then compare the plausibility of receiving a misleading observation (namely, the plausibility that this radio report is forged) against the plausibility that indeed Martian have landed on earth and that the radio would report this event.

Similarly, our analysis so far did not rely on the propositional nature of observations. Similar analysis applies when we consider an arbitrary set of "sensory input values" that the agent might receive. In particular, it applies to agents that play board games, where at each round each agent observes the opponent's move, and to a vision system, where the agent observes an array of pixel values at each time step.

Of course, we might want to use the propositional structure of observations to further decompose the description of observation plausibilities. For example, we might assume that the plausibility of observing p is independent of the plausibility of observing q. This assumption can be captured by requiring that $\mathrm{Pl}(O_m = p \wedge q | S_m = s) = \mathrm{Pl}(O_m = p | S_m = s) \bigotimes \mathrm{Pl}(O_m = q | S_m = s)$. This is one possible way of providing a compact description of the observation plausibilities. Clearly, this idea needs to be fleshed out carefully; we defer this to future work.

6.8 Characterizing Markovian Belief Change

Most approaches to belief change in the literature start with a collection of postulates, argue that they are reasonable, and prove some consequences of these postulates. Occasionally, a semantic model for the postulates is provided and a representation theorem is proved (of the form that every semantic model corresponds to some belief revision process, and that every belief revision process can be captured by some semantic model).

Our formalization of Markovian belief change is quite different. We have not discussed postulates at all. There is a good reason for this: Markovian belief change does not satisfy any postulates of the standard sort beyond those of any other notion of belief change.

As we saw in Chapter 5, the AGM revision postulates relate the agent's beliefs before and after revision. It is a straightforward exercise to state all the revision and update postulates, other than U8, in our language. (U8 requires some more sophistication; see [Friedman and Halpern 1994a].) For example, the revision postulate R2 can be stated in our language as $\mathit{learn}(\varphi) \Rightarrow B\varphi$, and R3 can be stated in our language as $\bigcirc\mathit{learn}(\varphi) \Rightarrow (\bigcirc B\psi \Rightarrow B(\varphi \Rightarrow \psi))$.

Let $\mathcal{C}^{BCS}$ consist of all BCSs, let $\mathcal{C}^{UPD4}$ consist of all BCSs that satisfy UPD4, and let $\mathcal{C}^{POM}$ consist of all BCSs with a structured partially observable Markovian prior. Recall that $\mathcal{C}^{M}$ consists of all BCSs that satisfy UPD4 and have a structured Markovian prior. Since $\mathcal{C}^{POM} \subseteq \mathcal{C}^{BCS}$, any formula valid in $\mathcal{C}^{BCS}$ must also be valid in $\mathcal{C}^{POM}$. Similarly, since $\mathcal{C}^{M} \subseteq \mathcal{C}^{UPD4}$, any formula valid in $\mathcal{C}^{UPD4}$ must also be valid in $\mathcal{C}^{M}$. As we now show, the converse holds as well.

We start by showing that we can assume, without loss of generality, that any "reasonable" BCS has a structured prior. We then show that any BCS with a structured prior can be simulated by one with a partially observable Markovian prior. This will suffice to show that if a formula is consistent in $\mathcal{C}^{BCS}$ then it is also consistent in $\mathcal{C}^{POM}$.

The basis for the first step is an extension of Theorem 6.3.2.

Theorem 6.8.1: *Let $(W, \mathcal{F}, \mathrm{Pl})$ be a qualitative plausibility space such that $\mathcal{F}$ is countable. Then there is an algebraic domain $\mathcal{D}$ and a structured plausibility space $(W, \mathcal{F}, \mathrm{Pl}')$ based on $\mathcal{D}$ such that Pl' is default-isomorphic to Pl.*

Using this result we can show that any "reasonable" BCS can be assumed to have a structured prior.

Theorem 6.8.2: *Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ be a BCS with a countable number of global states. Then there exists a system $\mathcal{I}' = (\mathcal{R}, \pi, \mathcal{P}')$ with a structured prior Pl'_a such that $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', r, m) \models \varphi$ for all runs r, times m, and formulas φ.*

Finally, we show that any BCS can be simulated with a BCS that satisfies the partially observable Markov assumption. Note that any partially observable Markovian prior is also, by definition, a Markovian prior.

Theorem 6.8.3: *Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ be a BCS. Then there is a systems $\mathcal{I}' = (\mathcal{R}', \pi', \mathcal{P}')$ and a one-to-one function f from $\mathcal{R}$ to $\mathcal{R}'$ such that $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', f(r), m) \models \varphi$ for all runs r, times m, and formulas φ. Moreover, if the prior of $\mathcal{I}$ is structured, then $\mathcal{I}' \in \mathcal{C}^{POM}$, and if $\mathcal{I}$ also satisfies UPD4, then $\mathcal{I}' \in \mathcal{C}^{M}$.*

Combining these results, we get the following theorem:

Theorem 6.8.4: *Let $\varphi \in \mathcal{L}^{KPT}$.*

- *φ is valid in $\mathcal{C}^{BCS}$ if and only if φ is valid in $\mathcal{C}^{POM}$.*
- *φ is valid in $\mathcal{C}^{UPD4}$ if and only if φ is valid in $\mathcal{C}^{M}$.*

Thus, there are no special postulates for Markovian belief change over and above the postulates that hold for any approach to belief change.

6.9 Related Work

We now briefly compare our approach to others in the literature.

One of the goals of Markovian belief change is to be able to combine aspects of revision and update. Recent work of Boutilier [1995, 1996] is similarly motivated. Essentially, Boutilier proposes conditioning on sequences of length two, using κ-rankings. While he does not pursue the probabilistic analogy, his discussion describes the belief change operation as a combination of beliefs about the state of the system, i.e., Pl$(S_1 = s)$ and beliefs about the likelihood of transitions, i.e., Pl$(S_2 = s' \mid S_1 = s)$. Boutilier proposes a two-stage procedure for updating belief which is similar to the one we outlined in Section 6.4. It is important to note that we derived this procedure using standard probabilistic arguments, something that Boutilier was not able to do in his framework. Our work, which was done independently of Boutilier's, can be viewed as extending his framework. We have arbitrary sequences of states, not just one-step transitions. In addition, because our approach is based on plausibility, not κ-rankings, we can treat the case where we have partially-ordered plausibilities.

Another novel aspect of our work is the examination of observations. As discussed in Section 6.7 it is unreasonable to assume that all consistent observations are equally likely. However, this assumption is implicitly made in virtually all works in the belief change literature.[5]

Finally, we note that the Markov assumption has been used extensively in the literature on probabilistic reasoning about action. Papers on this topic typically represent situations using *dynamic belief networks* [Dean and Kanazawa 1989]. Dynamic belief networks are essentially Markov chains with additional structure on states: A state is assumed to be characterized by the values of a number of variables. The probability of a transition is described by a belief network. Belief networks allow us to express more independence assumptions than just those characterizing

[5]The one exception that we know of is [Boutilier 1996]. Our results were developed independently of Boutilier's.

Markovian probabilities. For example, using a belief network, we can state that the value of variable X_2 at time $n+1$ is independent of the value of X_1 at time n given the value of X_2 at time n. Darwiche [1992] showed how qualitative Bayesian networks could be captured in his framework; it is straightforward to add such structure to the plausibility framework as well, once we restrict to structured plausibility spaces.

6.10 Discussion

There are three key developments in this chapter. First, we demonstrate how the Markov assumption can be implemented in a qualitative setting that allows for a natural notion of belief. While similar intuitions regarding independence may have guided previous work on belief change (e.g., [Katsuno and Mendelzon 1991a]), we are the first to make an explicit Markovian assumption. We believe that this approach provides a natural means of constructing a plausibility assessment over sequences given an assessment of the plausibility of transitions. Moreover, as we have seen, this assumption also leads to computational benefits similar to these found in the probabilistic setting.

Second, we examined what conclusions can be drawn given only an ordering on the transition plausibilities. Starting with a specification of constraints on the relative plausibility of transitions, we describe properties of the belief change operation. Since all we consider are comparisons between plausibilities of transitions, our conclusions are not always that strong. Of course, it is reasonable to consider richer forms of constraints, that might also constrain the behavior of $\otimes$ and $\oplus$. With more constraints, we can typically prove more about how the agent's beliefs will change. In future work, we plan to examine the consequences of using richer constraints.

Third, we address the issue of observations in belief change. We argue that it is unreasonable to assume that all observations are equally likely. This assumption has been implicitly made in virtually all works on the subject. We show, by example, that when we consider the plausibility of making an observation as well as the plausibility of changes in the world, we can capture situations that cannot be captured using standard models. Using a variant of the Markov assumption, we can succinctly specify the prior plausibility on runs given an assessment of transition plausibilities and the plausibility of making different observations at each state of the world.

One of the standard themes of the belief change literature is the effort to characterize a class of belief-change operators by postulates, such as the AGM postulates for belief revision [Alchourrón, Gärdenfors, and Makinson 1985] and the Katsuno-Mendelzon postulates for belief update [Katsuno and Mendelzon 1991a]. As the discussion in Section 6.8 shows, characterizing Markovian belief change with such postulates does not provide additional insight. The trouble is that such postulates do *not* let us reason about relative likelihoods of transitions and independence of transitions. But this is precisely the kind of reasoning that motivates the use of Markovian plausibility (and probability) measures. Such reasoning clearly plays a key role in modeling the stolen car problem and its variants. If the language is too weak to allow us to express such reasoning (as is the case for the language used to express the AGM postulates), then we cannot distinguish the class of Markovian plausibility measures from the class of all plausibility measures. This suggests that the right language to reason about belief change should allow us to talk about transitions and their relative plausibility. As our examples show, applications are often best described in these terms.

Chapter 7

Conclusion

We believe that the contribution of this dissertation is twofold. First, we developed a framework that allows us to model belief in dynamic systems in a natural way. In this framework, we view belief change in terms of the agents' knowledge, their assessment of the plausibility of possible situations, and how both change in time. Moreover, the analysis of AGM revision and KM update in this framework allows us to formally ground the intuitions about these approaches (e.g., revision reasons about a static world), to understand their limitations (e.g., update's propensity to believe that changes happen as late as possible), and to understand the assumptions they embody (e.g., that all observations are equally plausible).

Second, our framework provides the necessary foundations for more suitable notions of belief change that do not embody such strong assumptions. Since our framework distinguishes between assumptions about the belief change process and assumptions about the system (i.e., the interactions between the agents and their environment), we can easily extend current proposals to deal with actions, observations, muliple agents, etc.; see Section 5.7. Additionally, we proposed a new approach to belief change that imports ideas from probabilistic reasoning to allow for a combination of revision and update.

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 framework [Friedman and Halpern 1994b], we used preference orderings. We introduced plausibility 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., [Boutilier 1994c]), but also to approaches based on κ-rankings (e.g., [Goldszmidt and Pearl 1992]), probability measures (e.g., [Halpern and Tuttle 1993]), or any other measure of uncertainty. Second, plausibility measures allow us to get to the "core" aspects of our representation of uncertainty. For example, in our analysis of qualitative plausibility measures we identify the minimal conditions that are necessary and sufficient to get the KLM properties. From then on we examined plausibility measures that satisfied this condition, which simplified much of the formal machinery (e.g., the completeness proofs in Chapters 2 and 3). Third, plausibility makes it easier for us to incorporate intuitions from other approaches. For example, we adopted solutions from probability theory to deal with the problem of minimal change (in our definition of conditioning in Chapter 4), and the problem of compact specification of a prior over runs (in our use of the Markov assumption in Chapter 6). Finally, plausibility measures have the advantage of having 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 [Friedman, Halpern, and Koller 1996], once we examine structures with infinitely many worlds, qualitative plausibility measures can capture natural ordering of events that cannot be captured by preferential orderings, possibility measures, or κ-rankings.

As we have tried to argue throughout this dissertation, 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 4.3 or the iterated prisoners' dilemma game of Section 4.8). This means that most of the results in the current belief change literature are not directly applicable to 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. As we have seen, this allows (and in some sense, forces us) to deal with issues that have been ignored in previous work. For example, we have identified an implicit assumption that has been made in virtually all previous work on belief change. This assumption states that all possible observations are equally likely. As we argued in Section 6.7, this assumption is unreasonable and does not allow us to model many situations of interest. In this case, our framework forced us to recognize this assumption, since we had to explicitly describe the agent's local state and what her beliefs are in different local states.

Of course, more work needs to be done to show that this indeed the appropriate foundations for dealing with problems of belief change. One central problem that we intend to address in future work is that of reasoning about action and change in the face of uncertainty. Such reasoning appears, for example, in "classical" planning and execution. Most planners make many defeasible assumptions about the state of the world and the effects of actions. Such assumptions are crucial to make planning computationally feasible. During plan execution, however, the agent must monitor these assumptions, and change them in light of new observations. While this problem has been cited as one of the original motivations for examining belief change in AI, it is currently solved in practice using *ad hoc* domain specific technqiues. We believe that using our framework we can provide a formally justified method for specifying knowledge about unlikely events and principled methods for belief change in light of new observations.

To summarize, for the reasons cited above, we believe that our framework will prove to provide the necessary foundations for dealing with problems of belief change and qualitative treatment of uncertainty in intelligent agents.

Appendix A

Proofs for Chapter 2

A.1 Proofs for Section 2.4

Lemma 2.4.1: *Let W be set of possible worlds and let π be a function that maps each world in W to a truth assignment to $\mathcal{L}$. Let $T \subseteq \mathcal{L}_{\mathrm{def}}$ be a set of defaults that is closed under the rules of system* **P** *that satisfies the following condition:*

$(*)$ *if $\varphi \rightarrow \psi \in T$, $[\![\varphi]\!] = [\![\varphi']\!]$, and $[\![\psi]\!] = [\![\psi']\!]$, then $\varphi' \rightarrow \psi' \in T$, for all formulas $\varphi, \varphi', \psi, \psi' \in \mathcal{L}$.*

There is a plausibility structure $PL_T = (W, \mathrm{Pl}_T, \pi)$ such that $\mathrm{Pl}_T([\![\varphi]\!]) \leq \mathrm{Pl}_T([\![\psi]\!])$ if and only if $\varphi \vee \psi \rightarrow \psi \in T$. Moreover, $PL_T \models \varphi \rightarrow \psi$ if and only if $\varphi \rightarrow \psi \in T$.

Proof: We start by noting that the condition imposed on T ensures that Pl_T is well defined, i.e., sets that are described by different formulas are compared in a consistent manner. We now examine whether there exists a plausibility measure (W, Pl_T) satisfying $(*)$. It suffices to show that the order relation on Pl_T is reflexive, transitive, and satisfies A1.

Reflexivity Applying REF and LLE, we get that $(\varphi \vee \varphi) \rightarrow \varphi \in T$ for all φ.

Transitivity This is a direct consequence of the following lemma of Kraus, Lehmann and Magidor.

> **Lemma A.1.1:** [Kraus, Lehmann, and Magidor 1990, Lemma 5.5] *Let T be a set of defaults closed under applications of the rules of system* **P**. *Then T is closed under application of the rule:*
>
> - *From $\varphi_1 \vee \varphi_2 \rightarrow \varphi_2$ and $\varphi_2 \vee \varphi_3 \rightarrow \varphi_3$ infer $\varphi_1 \vee \varphi_3 \rightarrow \varphi_3$.*

A1 Assume $[\![\varphi]\!] \subseteq [\![\psi]\!]$. It follows that $[\![\psi]\!] = [\![\psi \vee \varphi]\!]$. Since T is closed under REF we get that $(\varphi \vee \psi) \rightarrow (\varphi \vee \psi) \in T$. Using $(*)$, we get that $(\varphi \vee \psi) \rightarrow \psi \in T$. Thus, $\mathrm{Pl}_T([\![\varphi]\!]) \leq \mathrm{Pl}_T([\![\psi]\!])$.

Finally, we need to show that $(W, \mathrm{Pl}_T, \pi) \models \varphi \rightarrow \psi$ if and only if $\varphi \rightarrow \psi \in T$. We start by observing that LLE, REF, and AND imply that $\varphi \rightarrow \psi \in T$ if and only if $(\varphi \wedge \neg\psi) \vee (\varphi \wedge \psi) \rightarrow \varphi \wedge \psi \in T$. We conclude that

$(**)$ $\varphi \rightarrow \psi \in T$ if and only if $\mathrm{Pl}_T([\![\varphi \wedge \neg\psi]\!]) \leq \mathrm{Pl}_T([\![\varphi \wedge \psi]\!])$.

Thus, it suffices to show that $\mathrm{Pl}_T([\![\varphi \wedge \neg\psi]\!]) \leq \mathrm{Pl}_T([\![\varphi \wedge \psi]\!])$ if and only if either $\mathrm{Pl}_T([\![\varphi \wedge \neg\psi]\!]) < \mathrm{Pl}_T([\![\varphi \wedge \psi]\!])$ or $\mathrm{Pl}_T([\![\varphi]\!]) = \perp$. The "if' direction is trivial. For the "only if" direction, suppose by way of contradiction that we have $\mathrm{Pl}_T([\![\varphi]\!]) > \perp$, $\mathrm{Pl}_T([\![\varphi \wedge \neg\psi]\!]) \leq \mathrm{Pl}_T([\![\varphi \wedge \psi]\!])$, and $\mathrm{Pl}_T([\![\varphi \wedge \psi]\!]) \leq \mathrm{Pl}_T([\![\varphi \wedge \neg\psi]\!])$. From $(**)$, we have that $\varphi \rightarrow \psi \in T$ and $\varphi \rightarrow \neg\psi \in T$. By AND and $(*)$, we have that $\varphi \rightarrow$*false* $\in T$. But then $\mathrm{Pl}_T([\![\varphi]\!]) = \perp$, which contradicts our assumptions. ∎

A.2 Proofs for Section 2.5

Theorem 2.5.2: *If* $\Delta \vdash_{\mathbf{P}'} \varphi \rightarrow \psi$, *then* $\Delta \models_{\mathcal{S}^{PL}} \varphi \rightarrow \psi$.

Proof: We need to show that LLE, RW, and REF are sound in $\mathcal{S}^{PL}$. Let $PL = (W, \mathrm{Pl}, \pi)$ be a plausibility structure. We proceed as follows.

LLE Assume that $\vdash_{\mathcal{L}} \varphi \Leftrightarrow \varphi'$. Then, by definition, $[\![\varphi]\!] = [\![\varphi']\!]$. The soundness of LLE immediately follows.

RW Assume that $\vdash_{\mathcal{L}} \psi \Rightarrow \psi'$ and that $PL \models \varphi \rightarrow \psi$. We want to show that $PL \models \varphi \rightarrow \psi'$. If $\mathrm{Pl}([\![\varphi]\!]) = \perp$, this is immediate. On the other hand, if $\mathrm{Pl}([\![\varphi]\!]) > \perp$, then $\mathrm{Pl}([\![\varphi \wedge \neg\psi]\!]) < \mathrm{Pl}([\![\varphi \wedge \psi]\!])$. Since $\vdash_{\mathcal{L}} \psi \Rightarrow \psi'$ we have that $[\![\psi]\!] \subseteq [\![\psi']\!]$. It follows that $[\![\varphi \wedge \neg\psi']\!] \subseteq [\![\varphi \wedge \neg\psi]\!]$ and $[\![\varphi \wedge \psi]\!] \subseteq [\![\varphi \wedge \psi']\!]$. Using A1, we conclude that $\mathrm{Pl}([\![\varphi \wedge \psi']\!] > \mathrm{Pl}([\![\varphi \wedge \neg\psi']\!]$, so $PL \models \varphi \rightarrow \psi'$.

REF By definition, $[\![\varphi \wedge \neg\varphi]\!] = \emptyset$ and $\mathrm{Pl}(\emptyset) = \perp$. Thus, if $\mathrm{Pl}([\![\varphi]\!]) > \perp$, then $PL \models \varphi \rightarrow \varphi$. On the other hand, if $\mathrm{Pl}([\![\varphi]\!]) = \perp$, then $PL \models \varphi \rightarrow \varphi$ vacuously.

∎

Proposition 2.5.3: *A plausibility measure satisfies A2 if and only if it satisfies A2′.*

Proof: Let (W, Pl) be a plausibility space. Assume that Pl satisfies A2. Let A, B_1, and B_2 be sets such that $\mathrm{Pl}(A \cap B_1) > \mathrm{Pl}(A \cap \overline{B_1})$ and $\mathrm{Pl}(A \cap B_2) > \mathrm{Pl}(A \cap \overline{B_2})$. Set $C = A \cap B_1 \cap B_2$, $D = A \cap B_1 \cap \overline{B_2}$, and $E = A \cap \overline{B_1}$. It is easy to verify that C, D, and E are pairwise disjoint. Since $C \cup D = A \cap B_1$, we have that $\mathrm{Pl}(C \cup D) > \mathrm{Pl}(E)$. Moreover, since $C \cup E \supseteq A \cap B_2$, $D \subseteq A \cap \overline{B_2}$, and $\mathrm{Pl}(A \cap B_2) > \mathrm{Pl}(A \cap \overline{B_2})$, we can apply A1 and conclude that $\mathrm{Pl}(C \cup E) > \mathrm{Pl}(D)$. Applying A2, we conclude that $\mathrm{Pl}(C) > \mathrm{Pl}(D \cup E)$, i.e., $\mathrm{Pl}(A \cap B_1 \cap B_2) > \mathrm{Pl}(A \cap \overline{(B_1 \cap B_2)})$. This gives us A2′.

Now assume that Pl satisfies A2$'$. Let C, D, and E be pairwise disjoint sets such that $\mathrm{Pl}(C \cup D) > \mathrm{Pl}(E)$ and $\mathrm{Pl}(C \cup E) > \mathrm{Pl}(D)$. Let $A = C \cup D \cup E$, $B_1 = C \cup D$, and $B_2 = C \cup E$. Then we have $\mathrm{Pl}(A \cap B_1) = \mathrm{Pl}(C \cup D) > \mathrm{Pl}(E) = \mathrm{Pl}(A \cap \overline{B_1})$ and $\mathrm{Pl}(A \cap B_2) = \mathrm{Pl}(C \cup E) > \mathrm{Pl}(D) = \mathrm{Pl}(A \cap \overline{B_2})$. From A2$'$, we have that $\mathrm{Pl}(A \cap B_1 \cap B_2) > \mathrm{Pl}(A \cap \overline{(B_1 \cap B_2)})$, i.e., $\mathrm{Pl}(C) > \mathrm{Pl}(D \cup E)$. This gives us A2. ∎

Theorem 2.5.4: $\mathcal{P} \subseteq \mathcal{S}^{QPL}$ *if and only if for all* Δ, φ, *and* ψ, *if* $\Delta \vdash_{\mathbf{P}} \varphi \rightarrow \psi$, *then* $\Delta \models_{\mathcal{S}} \varphi \rightarrow \psi$.

Proof: To prove the "if" direction it suffices to show that each rule in system **P** is sound in qualitative structures. Let $PL = (W, \mathrm{Pl}, \pi)$ be a qualitative plausibility structure. The soundness of LLE, RW, and REF is proved in Theorem 2.5.2. To deal with the remaining cases, we proceed as follows.

AND Assume that $PL \models \varphi \rightarrow \psi_1$ and $PL \models \varphi \rightarrow \psi_2$. If $\mathrm{Pl}([\![\varphi]\!]) = \bot$, then $PL \models \varphi \rightarrow \psi_1 \wedge \psi_2$ vacuously. Assume that $\mathrm{Pl}([\![\varphi]\!]) > \bot$. Let $A = [\![\varphi]\!]$, $B_1 = [\![\psi_1]\!]$, and $B_2 = [\![\psi_2]\!]$. Since $PL \models \varphi \rightarrow \psi_1$ and $PL \models \varphi \rightarrow \psi_2$, we have that $\mathrm{Pl}(A \cap B_1) > \mathrm{Pl}(A \cap \overline{B_1})$ and $\mathrm{Pl}(A \cap B_2) > \mathrm{Pl}(A \cap \overline{B_2})$. Proposition 2.5.3 states that Pl satisfies A2$'$, and thus we get that $\mathrm{Pl}(A \cap B_1 \cap B_2) > \mathrm{Pl}(A \cap \overline{(B_1 \cap B_2)})$, so $PL \models \varphi \rightarrow \psi_1 \wedge \psi_2$.

CM Again assume that $PL \models \varphi \rightarrow \psi_1$ and $PL \models \varphi \rightarrow \psi_2$. If $\mathrm{Pl}([\![\varphi \wedge \psi_1]\!]) = \bot$, then $PL \models \varphi \wedge \psi_1 \rightarrow \psi_2$ vacuously. Assume that $\mathrm{Pl}([\![\varphi \wedge \psi_1]\!]) > \bot$. Let A, B_1, and B_2 be defined as in the treatment of AND above. Again, we have $\mathrm{Pl}((A \cap B_1 \cap B_2) > \mathrm{Pl}(A \cap \overline{(B_1 \cap B_2)})$. Since $A \cap B_1 \cap \overline{B_2} \subseteq A \cap \overline{(B_1 \cap B_2)}$, we conclude that $\mathrm{Pl}(A \cap B_1 \cap B_2) > \mathrm{Pl}(A \cap B_1 \cap \overline{B_2})$. Thus, $PL \models \varphi \wedge \psi_1 \rightarrow \psi_2$.

OR Assume that $PL \models \varphi_1 \rightarrow \psi$ and $PL \models \varphi_2 \rightarrow \psi$. If $\mathrm{Pl}([\![\varphi_1]\!]) = \mathrm{Pl}([\![\varphi_2]\!]) = \bot$, then applying A3 we get that $\mathrm{Pl}([\![\varphi_1 \vee \varphi_2]\!]) = \bot$ and thus $PL \models (\varphi_1 \vee \varphi_2) \rightarrow \psi$ vacuously. So assume that $\mathrm{Pl}([\![\varphi_1]\!]) > \bot$. (Identical argument works if $\mathrm{Pl}([\![\varphi_2]\!]) > \bot$.) Set $A = [\![(\varphi_1 \vee \varphi_2) \wedge \psi]\!]$, $B = [\![\varphi_1 \wedge \neg\psi]\!]$, and $C = [\![\varphi_2 \wedge \neg\varphi_1 \wedge \neg\psi]\!]$. To prove that $PL \models (\varphi_1 \vee \varphi_2) \rightarrow \psi$, we must show that $\mathrm{Pl}(A) > \mathrm{Pl}(B \cup C)$. Since $PL \models \varphi_1 \rightarrow \psi$, we have that $\mathrm{Pl}(A) \geq \mathrm{Pl}([\![\varphi_1 \wedge \psi]\!]) > \mathrm{Pl}(B)$. If $\mathrm{Pl}([\![\varphi_2]\!]) = \bot$, then $\mathrm{Pl}(C) = \bot$ and we conclude that $\mathrm{Pl}(A) > \mathrm{Pl}(C)$. On the other hand, if $\mathrm{Pl}([\![\varphi_2]\!]) > \bot$, then since $PL \models \varphi_2 \rightarrow \psi$, we have $\mathrm{Pl}(A) \geq \mathrm{Pl}([\![\varphi_2 \wedge \psi]\!]) > \mathrm{Pl}([\![\varphi_2 \wedge \neg\psi]\!]) \geq \mathrm{Pl}(C)$. From A2, $\mathrm{Pl}(A) > \mathrm{Pl}(B)$, and $\mathrm{Pl}(A) > \mathrm{Pl}(C)$, we get that $\mathrm{Pl}(A) > \mathrm{Pl}(B \cup C)$. Thus, $PL \models (\varphi_1 \vee \varphi_2) \rightarrow \psi$.

To prove the "only if" direction we have to show that if there is some $PL = (W, \mathrm{Pl}, \pi)$ in $\mathcal{S}$ that is not qualitative, then the KLM properties are not sound with respect to P. Assume that Pl does not satisfy A2. Since we have assumed that $\mathcal{F} = \{[\![\varphi]\!] : \varphi \in \mathcal{L}\}$, there are formulas φ, ψ_1, and ψ_2, such that $[\![\varphi]\!]$, $[\![\psi_1]\!]$, and $[\![\psi_2]\!]$ are pairwise disjoint, $\mathrm{Pl}([\![\varphi \vee \psi_2]\!]) > \mathrm{Pl}([\![\psi_1]\!])$, $\mathrm{Pl}([\![\varphi \vee \psi_1]\!]) > \mathrm{Pl}([\![\psi_2]\!])$, and yet $\mathrm{Pl}([\![\varphi]\!]) \not> \mathrm{Pl}([\![\psi_1 \wedge \psi_2]\!])$. Thus, $PL \models (\varphi \vee \psi_1 \vee \psi_2) \rightarrow \neg\psi_1$ and $PL \models (\varphi \vee \psi_1 \vee \psi_2) \rightarrow \neg\psi_2$. However, $PL \not\models (\varphi \vee \psi_1 \vee \psi_2) \rightarrow (\neg\psi_1 \wedge \neg\psi_2)$. This shows that the AND rule is not sound in $\mathcal{S}$. Now assume that there is some $PL = (W, \mathrm{Pl}, \pi)$ that does not satisfy A3. Thus, there are formulas φ_1 and φ_2 such that $\mathrm{Pl}([\![\varphi_1]\!]) = \mathrm{Pl}([\![\varphi_2]\!]) = \bot$ and $\mathrm{Pl}([\![\varphi_1 \vee \varphi_2]\!]) > \bot$. We conclude that $PL \models \varphi_1 \rightarrow$ *false* and $PL \models \varphi_2 \rightarrow$ *false*, but $PL \not\models (\varphi_1 \vee \varphi_2) \rightarrow$ *false*. This shows that the OR rule is not sound in $\mathcal{S}$. ∎

Theorem 2.5.8: *A set $\mathcal{S}$ of qualitative plausibility structures is rich if and only if for all finite Δ and defaults $\varphi{\rightarrow}\psi$, we have that $\Delta \models_{\mathcal{S}} \varphi{\rightarrow}\psi$ implies $\Delta \vdash_{\mathbf{P}} \varphi{\rightarrow}\psi$.*

Proof: For the "if" direction, assume that $\mathcal{S}$ is not rich. We need to show that system **P** is not complete for $\models_{\mathcal{S}}$. It is sufficient to construct Δ, φ, and ψ such that $\Delta \models_{\mathcal{S}} \varphi{\rightarrow}\psi$ but $\Delta \not\vdash_{\mathbf{P}} \varphi{\rightarrow}\psi$.

We start with a lemma, whose straightforward proof is left to the reader.

Lemma A.2.1: *Let $\varphi_1, \ldots, \varphi_n$ be a collection of mutually exclusive formulas. Let Δ consist of the default $\varphi_n{\rightarrow}$false and the defaults $\varphi_i \vee \varphi_j{\rightarrow}\varphi_i$ for all $1 \leq i < j \leq n$. Then $(W, \mathrm{Pl}, \pi) \models \Delta$ if and only if there is some j with $1 \leq j \leq n$ such that*

$$\mathrm{Pl}([\![\varphi_1]\!]) > \mathrm{Pl}([\![\varphi_2]\!]) > \cdots > \mathrm{Pl}([\![\varphi_j]\!]) = \cdots = \mathrm{Pl}([\![\varphi_n]\!]) = \perp.$$

Since $\mathcal{S}$ is not rich, there is a collection $\varphi_1, \ldots, \varphi_n$ that is a counterexample to the definition of richness. Let Δ be the set of defaults defined in Lemma A.2.1. We claim that if $(W, \mathrm{Pl}, \pi) \in \mathcal{S}$ satisfies all the defaults in Δ, then $\mathrm{Pl}([\![\varphi_{n-1}]\!]) = \perp$. To see this, assume that $\mathrm{Pl}([\![\varphi_{n-1}]\!]) > \perp$. Then according to Lemma A.2.1, $\mathrm{Pl}([\![\varphi_1]\!]) > \cdots > \mathrm{Pl}([\![\varphi_{n-1}]\!]) > \mathrm{Pl}([\![\varphi_n]\!]) = \perp$, but this contradicts the assumption that the sequence $\varphi_1, \ldots, \varphi_n$ is a counterexample to richness. Since $\mathrm{Pl}([\![\varphi_{n-1}]\!]) = \perp$ in every structure that satisfies Δ, we conclude that $\Delta \models_{\mathcal{S}} \varphi_{n-1}{\rightarrow}$*false*.

We now show that $\Delta \not\vdash_{\mathbf{P}} \varphi_{n-1}{\rightarrow}$*false*. The easiest way of proving this is by using Theorem 2.3.1. All we need to show is that there is a preferential structure that satisfies Δ but does not satisfies $\varphi_{n-1}{\rightarrow}$*false*. Let $W = \{w_1, \ldots, w_{n-1}\}$, let $\prec$ be such that $w_i \prec w_j$ for all $i < j$, and let π be such that $\pi(w_i)(\varphi_j) =$ **true** if and only if $i = j$. It is straightforward to verify that $(W, \prec, \pi)$ satisfies Δ. However, it easy to see that $(W, \prec, \pi) \not\models \varphi_{n-1}{\rightarrow}$*false*. This concludes the proof of "if" direction.

For the "only if" direction, assume that there is some Δ and $\varphi{\rightarrow}\psi$ such that $\Delta \models_{\mathcal{S}} \varphi{\rightarrow}\psi$ but $\Delta \not\vdash_{\mathbf{P}} \varphi{\rightarrow}\psi$. Using Theorem 2.3.1 we get that $\Delta \not\models_{\mathrm{p}} \varphi{\rightarrow}\psi$. Thus, there is some preferential structure $P = (W, \prec, \pi)$ that satisfies the defaults in Δ but not $\varphi{\rightarrow}\psi$. In fact, as the following lemma shows, we can assume that P is a *linear* structure.

Lemma A.2.2: [Friedman and Halpern 1994c] *Let Δ be a finite set of defaults. If there is a preferential structure that satisfies Δ and does not satisfy $\varphi{\rightarrow}\psi$, then there is a preferential structure $P = (W, \prec, \pi)$ such that $W = \{w_1, \ldots, w_n\}$, $w_i \prec w_j$ for all $i < j$, $P \models \Delta$ and $P \not\models \varphi{\rightarrow}\psi$.*

We now use P to construct a sequence of formulas that will be a counterexample to the richness of $\mathcal{S}$. Let $p_1, \ldots, p_m$ be the propositions that appear in Δ and $\varphi{\rightarrow}\psi$. Since Δ is finite, there is a finite number of such propositions. We note that whether a default $\varphi{\rightarrow}\psi$ is satisfied in P depends only on the minimal world satisfying φ. If $\pi(w_i)$ and $\pi(w_j)$ for some $i < j$ agree on the truth of $p_1, \ldots, p_m$, then w_j cannot be a minimal world for any formula defined using only $p_1 \ldots p_m$. Thus, we can assume, without loss of generality, that for all $w_i \neq w_j$, there is some p_k that is assigned a different truth value by each of the two worlds. We now construct formulas that characterize the truth assignment to $p_1, \ldots, p_m$ in each world in W. Let

$$\varphi_i = \bigwedge_{\{j:\pi(w_i)(p_j)=\mathbf{true}\}} p_j \wedge \bigwedge_{\{j:\pi(w_i)(p_j)=\mathbf{false}\}} \neg p_j$$

for $i = 1, \ldots, n$, and let $\varphi_{n+1} = \neg(\varphi_1 \vee \ldots \vee \varphi_n)$. It is easy to verify that these formulas are mutually exclusive.

We now claim that if *PL* is a plausibility structure where $\mathrm{Pl}([\![\varphi_1]\!]) > \cdots > \mathrm{Pl}([\![\varphi_{n+1}]\!]) = \bot$, then *PL* satisfies the defaults in Δ but not $\varphi \rightarrow \psi$. This will suffice to prove that $\mathcal{S}$ is not rich, since if $\mathcal{S}$ contains such a structure we get a contradiction to the assumption that $\Delta \models_{\mathcal{S}} \varphi \rightarrow \psi$.

Let *PL* be a plausibility structure where $\mathrm{Pl}([\![\varphi_1]\!]) > \cdots > \mathrm{Pl}([\![\varphi_{n+1}]\!]) = \bot$. We want to show that $PL \models \xi \rightarrow \xi'$ if and only if $P \models \xi \rightarrow \xi'$, for all formulas ξ and ξ' defined using only $p_1, \ldots, p_m$.

Let ξ, ξ' be formulas defined over $p_1, \ldots, p_m$. Assume that $P \models \xi \rightarrow \xi'$. There are two cases: either (a) ξ is not satisfied in W, or (b) the minimal world satisfying ξ also satisfies ξ'. In case (a), it is easy to see that $\vdash_{\mathcal{L}} \xi \Rightarrow \varphi_{n+1}$. From A1, we have that $\mathrm{Pl}(\xi) = \bot$, and thus $PL \models \xi \rightarrow \xi'$ vacuously. In case (b) assume that w_i is the minimal world satisfying ξ. Since $P \models \xi \rightarrow \xi'$ we have that $\pi(w_i) \models \xi \wedge \xi'$. A simple argument shows that $\vdash_{\mathcal{L}} \varphi_i \Rightarrow \xi \wedge \xi'$. Thus, using A1, we get that $\mathrm{Pl}([\![\xi \wedge \xi']\!]) \geq \mathrm{Pl}([\![\varphi_i]\!])$. Since w_i is the minimal world satisfying ξ and it also satisfies ξ', we have that $\xi \wedge \neg\xi'$ is not satisfied by $w_1, \ldots, w_i$. Since $\varphi_1, \ldots, \varphi_{n+1}$ are exhaustive, we have that $\vdash_{\mathcal{L}} (\xi \wedge \neg\xi') \Rightarrow (\varphi_{i+1} \vee \ldots \vee \varphi_{n+1})$. Thus, $\mathrm{Pl}([\![\xi \wedge \neg\xi']\!]) \leq \mathrm{Pl}([\![\varphi_{i+1} \vee \ldots \vee \varphi_{n+1}]\!])$. By repeated applications of A2 and the fact that $\mathrm{Pl}([\![\varphi_i]\!]) > \mathrm{Pl}([\![\varphi_j]\!])$ for all $j > i$, we get that $\mathrm{Pl}([\![\varphi_i]\!]) > \mathrm{Pl}[\![\xi \wedge \neg\xi']\!])$. We conclude that $\mathrm{Pl}([\![\xi \wedge \xi']\!]) > \mathrm{Pl}([\![\xi \wedge \neg\xi']\!])$ and thus $PL \models \xi \rightarrow \xi'$.

Now assume that $P \not\models \xi \rightarrow \xi'$. Thus, there is a minimal world w_i that satisfies ξ; moreover, w_i does not satisfy ξ'. This implies that $P \models \xi \rightarrow \neg\xi'$ and $\mathrm{Pl}([\![\xi]\!]) > \bot$. Applying the proof in the previous paragraph, we have that $\mathrm{Pl}([\![\xi \wedge \neg\xi']\!]) > \mathrm{Pl}([\![\xi \wedge \xi']\!])$, and thus $PL \not\models \xi \rightarrow \xi'$. ∎

A.3 Proofs for Section 2.6

We now prove Theorem 2.6.3. We start with two preliminary lemmas.

Lemma A.3.1: *Let* $(W, \mathcal{F}, \mathrm{Pl})$ *be a normal qualitative plausibility space such that* $\mathcal{F}$ *is finite, let* $A^*, B^* \in \mathcal{F}$ *be disjoint sets such that* $\mathrm{Pl}(A^*) \not< \mathrm{Pl}(B^*)$*, and let* $x \geq 2$*. Then there is a probability measure* Pr *over* W *such that* $\Pr(B^*|A^* \cup B^*) \leq \frac{1}{2}$*; moreover if* $\mathrm{Pl}(A) < \mathrm{Pl}(B)$ *then* $\Pr(B|A \cup B) \geq 1 - \frac{1}{x}$*, for all disjoint sets* $A, B \in \mathcal{F}$*.*

Proof: Let $A_1, \ldots, A_n$ be the *atoms* of $\mathcal{F}$, i.e., each $A_i \neq \emptyset$ is in $\mathcal{F}$ and there is no nonempty $B \in \mathcal{F}$ such that $B \subset A_i$. Since $\mathcal{F}$ is finite, every set in $\mathcal{F}$ is a disjoint union of atoms.

We can describe Pl using a set of defaults. Let $p_1, \ldots, p_n$ be a collection of distinct propositions. For each set $A \in \mathcal{F}$ we define $\varphi_A = \bigvee_{A_i \subseteq A} p_i$ if A is nonempty, and define $\varphi_A = \mathit{false}$ if A is empty. Let

$$\begin{aligned}\Delta \;=\; & \{(\varphi_A \vee \varphi_B) \rightarrow \varphi_B : A, B \in \mathcal{F}, A \cap B = \emptyset, \mathrm{Pl}(A) < \mathrm{Pl}(B)\} \cup \\ & \{(p_i \wedge p_j) \rightarrow \mathit{false} : i \neq j\} \cup \\ & \{\neg(p_1 \vee \ldots \vee p_n) \rightarrow \mathit{false}\}.\end{aligned}$$

Let π be a truth assignment to W such that $\pi(w)(p_i) =$ **true** if and only if $w \in A_i$. Then it is easy to check that $PL = (W, \mathcal{F}, \text{Pl}, \pi)$ satisfies Δ and $PL \not\models (\varphi_{A^*} \vee \varphi_{B^*}) \rightarrow \varphi_{B^*}$. Since Pl is a qualitative plausibility measure by assumption, we have that $PL \in \mathcal{S}^{QPL}$. Using Theorem 2.5.8, we get that $\Delta \not\vdash_{\mathbf{P}} (\varphi_{A^*} \vee \varphi_{B^*}) \rightarrow \varphi_{B^*}$. From Theorem 2.3.1 we conclude that there is a preferential structure $P = (W_P, \prec, \pi_P)$ such that P satisfies Δ but $P \not\models (\varphi_{A^*} \vee \varphi_{B^*}) \rightarrow \varphi_{B^*}$. Applying Lemma A.2.2 we can assume that P is linear, i.e., $W_P = \{v_1, \ldots, v_m\}$ for some $m \le n$, and $\prec$ is such that $v_i \prec v_j$ if $i < j$. Moreover, we have that W_P not empty: Since PL is normal, we have that $PL \not\models$ *true*$\rightarrow$*false*. Hence, $P \not\models$ *true*$\rightarrow$*false*, which implies that there are some minimal worlds that satisfy *true*.

We now construct the required probability measure over W. We start by noting that the defaults in Δ imply that for each world v_i, $\pi_P(v_i)$ assigns **true** to exactly one proposition. Thus, without loss of generality, we can assume that $\pi_P(v_i)(p_j) =$**true** if and only if $i = j$. We define Pr by assigning a probability to each atom. The probability of all other sets is induced from this assignment: $\Pr(A) = \sum_{A_i \subseteq A} \Pr(A_i)$, for all $A \in \mathcal{F}$. Let A_i be an atom. We define $\Pr(A_i) = \alpha \cdot (\frac{1}{x})^i$, if $i \le m$; otherwise we define $\Pr(A_i) = 0$. The constant α is a normalization constant that ensures that the probability of atoms sum to 1. It is easy to verify that $\alpha = (x-1)\frac{x^m}{x^m - 1}$.

We now show that Pr satisfies the requirements of the lemma. Assume that $A, B \in \mathcal{F}$ are disjoint and $\text{Pl}(A) < \text{Pl}(B)$. We want to show that $\Pr(B|A \cup B) \ge 1 - \frac{1}{x}$. By definition, $(\varphi_A \vee \varphi_B) \rightarrow \varphi_B \in \Delta$, and thus $P \models (\varphi_A \vee \varphi_B) \rightarrow \varphi_B$. There are two cases: either (a) there is no world in W_P that satisfies $(\varphi_A \vee \varphi_B)$ or (b) the minimal world that satisfies $(\varphi_A \vee \varphi_B)$ also satisfies φ_B. In case (a), it immediately follows that if $A_i \subseteq A \cup B$, then $i > m$. We conclude that $\Pr(A) = \Pr(B) = 0$. Thus, $\Pr(B|A \cup B) = 1$. (Recall that if $\Pr(A \cup B) = 0$ then, by convention, $\Pr(B|A \cup B) = 1$.) In case (b), assume that v_i is the minimal world satisfying $\varphi_A \vee \varphi_B$. Since $P \models (\varphi_A \vee \varphi_B) \rightarrow \varphi_B$, it must be the case that v_i satisfies φ_B. This implies that $\Pr(B) \ge \Pr(A_i)$. Since v_i is the minimal world that satisfies $\varphi_A \vee \varphi_B$, and since v_i does not satisfy φ_B (since A and B are disjoint) we conclude that $\Pr(A) \le \sum_{j > i} \Pr(A_j)$. Simple calculation show that $\sum_{j > i} \Pr(A_j) \le \Pr(A_i) \cdot \frac{1}{x-1}$. Thus, $(x-1)\Pr(A) \le \Pr(A_i) \le \Pr(B)$. This implies that $(x-1)(\Pr(A) + \Pr(B)) \le x \Pr(B)$. Since A and B are disjoint, we have that $\Pr(A \cup B) = \Pr(A) + \Pr(B)$. We get that $(x-1)\Pr(A \cup B) \le x \Pr(B)$, so $\Pr(B|A \cup B) \ge \frac{x-1}{x}$.

Finally we have to show that $\Pr(B^*|A^* \cup B^*) \le \frac{1}{2}$. Since $P \not\models (\varphi_{A^*} \vee \varphi_{B^*}) \rightarrow \varphi_{B^*}$, the minimal world, v_i, satisfying $\varphi_{A^*} \vee \varphi_{B^*}$ does not satisfy φ_{B^*}. Since A^* and B^* are disjoint, this implies that v_i satisfies φ_{A^*}. The argument above shows that $\Pr(A^*|A^* \cup B^*) \ge 1 - \frac{1}{x}$ and $\Pr(A^* \cup B^*) > 0$. Therefore, $\Pr(B^*|A^* \cup B^*) \le \frac{1}{x} \le \frac{1}{2}$. ∎

Lemma A.3.2: *Let* $(W, \mathcal{F}, \text{Pl})$ *be a normal qualitative plausibility space such that* $\mathcal{F}$ *is finite. Then there is a PPD* $PP = \{\Pr_n : n \ge 1\}$ *such that, for all disjoint* $A, B \in \mathcal{F}$,

- *if* $\text{Pl}(A) < \text{Pl}(B)$, *then* $\Pr_n(B|A \cup B) \ge 1 - \frac{1}{n+1}$ *for all* n,
- *if* $\text{Pl}(A) \not< \text{Pl}(B)$, *then for all* n, *there is an* m *such that* $n \le m < n + |\mathcal{F}|^2$ *and* $\Pr_m(B|A \cup B) \le \frac{1}{2}$.

Proof: Let $A_0, \ldots, A_{|\mathcal{F}|-1}$ be some enumeration of the members of $\mathcal{F}$. We define the PPD $PP = \{\Pr_n : n \ge 1\}$ as follows. For each $n \ge 1$ we define i_n, j_n to be the unique integers such that

$n = k \cdot |\mathcal{F}|^2 + i_n \cdot |\mathcal{F}| + j_n$ for some positive integer k. According to Lemma A.3.1 there exists a probability distribution $\Pr_n$ on W such that if $\mathrm{Pl}(A) < \mathrm{Pl}(B)$, then $\Pr_n(B|A \cup B) \geq 1 - \frac{1}{n+1}$, for all disjoint sets $A, B \in F$. Moreover, if $\mathrm{Pl}(A_{i_n}) \not< \mathrm{Pl}(A_{j_n})$, we can ensure that $\Pr_n(A_{j_n}|A_{i_n} \cup A_{j_n}) \leq \frac{1}{2}$. It is easy to verify that this PPD satisfies the requirements of lemma. ∎

Theorem 2.6.3: *Let PL $\in \mathcal{S}^{QPL}$ be a normal plausibility structure with a countable algebra $\mathcal{F}$. Then there is a structure PL$'$ $\in \mathcal{S}^{\epsilon}$ default-isomorphic to PL.*

Proof: Let $PL = (W, \mathcal{F}, \mathrm{Pl}, \pi)$ be a normal qualitative plausibility such that $\mathcal{F}$ is countable. Let $A_1, A_2, \ldots$ be an enumeration of the sets in $\mathcal{F}$. Without loss of generality, we can assume that $A_1 = W$. Since Pl is normal, we must have $\mathrm{Pl}(A_1) > \perp$. For each k, let $\mathcal{F}_k$ be the minimal algebra that contains $A_1, \ldots, A_k$. Clearly, $\mathcal{F}_k$ is a finite algebra, and Pl restricted to $\mathcal{F}_k$ is normal. Thus, there is a PPD $PP^k = \{\Pr_1^k, \Pr_2^k, \ldots\}$ that satisfies the conditions of Lemma A.3.2.

We now use elements of these sequences to construct the desired PPD. We define PP to be the sequence that consists of a segment of PP^1, followed by a segment of PP^2, and so on, such that the length of the segment from PP^k is $|F_k|^2$:

$$\{\Pr_1^1, \ldots, \Pr_{1+|\mathcal{F}_1|^2-1}^1, \Pr_2^2, \ldots, \Pr_{2+|\mathcal{F}_2|^2-1}^2, \ldots, \Pr_k^k, \ldots, \Pr_{k+|\mathcal{F}_k|^2-1}^k, \ldots\}.$$

To show that the plausibility measure that corresponds to (W, PP, π) is default-isomorphic to *PL* we need to show that for all disjoint $A_i, A_j \in \mathcal{F}$, $\mathrm{Pl}(A_i) < \mathrm{Pl}(A_j)$ if and only if $\lim_{n\to\infty} \Pr_n(A_j|A_i \cup A_j) = 1$.

Assume that $\mathrm{Pl}(A_i) < \mathrm{Pl}(A_j)$. Let $\epsilon > 0$. Set $m_\epsilon = \max(i, j, \lceil\frac{1}{\epsilon}\rceil)$, and $N_\epsilon = \sum_{k=1}^{m_\epsilon} |\mathcal{F}_k|^2$. Let $n > N_\epsilon$. Then our construction is such that $\Pr_n = \Pr_l^k$ for some $l \geq k \geq m_\epsilon$. Since $k \geq max(i, j)$, we have that $A_i, A_j \in \mathcal{F}_k$. Moreover, according to Lemma A.3.2, $\Pr_l^k(A_j|A_i \cup A_j) \geq 1 - \frac{1}{l+1} \geq 1 - \frac{1}{m_\epsilon+1} \geq 1 - \epsilon$. Thus, we conclude that $\lim_{n\to\infty} \Pr_n(A_j|A_i \cup A_j) = 1$.

Assume that $\mathrm{Pl}(A_i) \not< \mathrm{Pl}(A_j)$. Let $k > \max(i, j)$. According to Lemma A.3.2, there is an m such that $k \leq m < k + |\mathcal{F}_k|^2$ such that $\Pr_m^k(A_j|A_i \cup A_j) \leq \frac{1}{2}$. Moreover, by the our construction, there is an n such that $\Pr_n = \Pr_m^k$. Thus, for infinitely many n, $\Pr_n(A_j|A_i \cup A_j) \leq \frac{1}{2}$. We conclude that $\lim_{n\to\infty} \Pr_n(A_j|A_i \cup A_j) \neq 1$. ∎

A.4 Proofs for Section 2.7

Proposition 2.7.2: *If E is an expectation structure, then PL$_E$ is a plausibility structure. Furthermore, $E \models \varphi{\rightarrow}\psi$ if and only if PL$_E \models \varphi{\rightarrow}\psi$.*

Proof: Suppose $E = (\mathcal{L}, \trianglelefteq)$ is an expectation structure, and $PL_E = (W_E, \mathrm{Pl}_E, \pi_E)$. To show that PL_E is a plausibility structure, we need to check that the ordering defined by Pl_E is reflexive, transitive, and satisfies A1. Note that it is easy to show, using standard arguments, that $[\![\varphi]\!] \subseteq [\![\psi]\!]$ if and only if $\vdash_{\mathcal{L}} \varphi \Rightarrow \psi$. The proof that PL_E is plausibility measure follows in a straightforward manner and is left as an exercise to the reader.

We now show that $E \models \varphi \rightarrow \psi$ if and only if $PL_E \models \varphi \rightarrow \psi$. Assume that $E \models \varphi \rightarrow \psi$. According to Theorem 2.7.1 there are two possible cases: either (a) $\vdash_{\mathcal{L}} \varphi \Rightarrow \psi$ or (b) $(\varphi \Rightarrow \neg\psi) \lhd (\varphi \Rightarrow \psi)$. In case (a), we apply REF and RW (which are valid in all plausibility structures) to get that $PL_E \models \varphi \rightarrow \psi$. Now consider case (b). It is clear that *true* $\not\trianglelefteq (\varphi \Rightarrow \neg\psi)$, for if *true* $\trianglelefteq (\varphi \Rightarrow \neg\psi)$, then *true* $\lhd (\varphi \Rightarrow \psi)$ by E1. We also have $(\varphi \Rightarrow \psi) \trianglelefteq$ *true* by using E2, so by E1 again, we get that *true* $\lhd$ *true*, which is a contradiction. Since *true* $\not\trianglelefteq (\varphi \Rightarrow \neg\psi)$ and $(\varphi \Rightarrow \neg\psi) \lhd (\varphi \Rightarrow \psi)$, by definition, we have that $\mathrm{Pl}_E([\![\varphi \wedge \neg\psi]\!]) \leq \mathrm{Pl}_E([\![\varphi \wedge \psi]\!])$. Moreover, since *true* $\not\trianglelefteq (\varphi \Rightarrow \neg\psi)$, we have that $\varphi \wedge \psi$ is consistent. Thus, $\not\vdash_{\mathcal{L}} \varphi \wedge \psi \Rightarrow \varphi \wedge \neg\psi$. Furthermore, since $(\varphi \Rightarrow \neg\psi) \lhd (\varphi \Rightarrow \psi)$, we also have that $(\varphi \Rightarrow \psi) \not\lhd (\varphi \Rightarrow \neg\psi)$. This implies, by definition, that $\mathrm{Pl}_E([\![\varphi \wedge \neg\psi]\!]) \not\geq \mathrm{Pl}_E([\![\varphi \wedge \psi]\!])$. Hence, $\mathrm{Pl}_E([\![\varphi \wedge \neg\psi]\!]) < \mathrm{Pl}_E([\![\varphi \wedge \psi]\!])$, and so $PL_E \models \varphi \rightarrow \psi$.

Now assume that $PL_E \models \varphi \rightarrow \psi$. If $\vdash_{\mathcal{L}} \varphi \Rightarrow \psi$, then $E \models \varphi \rightarrow \psi$. Assume that $\not\vdash_{\mathcal{L}} \varphi \Rightarrow \psi$. This implies that $\varphi \wedge \neg\psi$ is consistent in $\mathcal{L}$. We claim that $\mathrm{Pl}_E([\![\varphi]\!]) > \bot$. To see this, assume that $\mathrm{Pl}_E([\![\varphi]\!] \leq \mathrm{Pl}([\![\mathit{false}]\!])$. Since *true* $\trianglelefteq \neg$*false*, the definition of Pl_E implies that $\vdash_{\mathcal{L}} \varphi \Rightarrow \mathit{false}$. But this contradict the assumption that φ is consistent in $\mathcal{L}$. We conclude that $\mathrm{Pl}_E([\![\varphi]\!]) > \bot$, and thus since $PL_e \models \varphi \rightarrow \psi$, $\mathrm{Pl}_E([\![\varphi \wedge \psi]\!]) > \mathrm{Pl}_E([\![\varphi \wedge \neg\psi]\!])$. It is straightforward to show that since $\varphi \wedge \neg\psi$ is consistent in $\mathcal{L}$, we get that $\neg(\varphi \wedge \psi) \lhd \neg(\varphi \wedge \neg\psi)$. Rewriting this equation, we get that $(\varphi \Rightarrow \neg\psi) \lhd (\varphi \Rightarrow \psi)$, and thus $E \models \varphi \rightarrow \psi$. ∎

A.5 Proofs for Section 2.8

We now prove Lemma 2.8.3. We first need one preliminary result, which collects a number of properties of the N operator.

Lemma A.5.1: *(a)* $\vdash_{\mathbf{C}} N(\varphi \wedge \psi) \Rightarrow N\varphi$.

(b) $\vdash_{\mathbf{C}} (N\varphi \wedge N\psi) \Rightarrow N(\varphi \wedge \psi)$.

(c) $\vdash_{\mathbf{C}} (N\varphi \wedge N(\varphi \Rightarrow \psi)) \Rightarrow N\psi$.

(d) If $\vdash_{\mathbf{C}} \varphi$ *then* $\vdash_{\mathbf{C}} N\varphi$.

(e) $\vdash_{\mathbf{C}} N\varphi \Rightarrow (\psi \rightarrow \varphi)$.

(f) $\vdash_{\mathbf{C}} (N(\varphi \Leftrightarrow \varphi') \wedge N(\psi \Leftrightarrow \psi')) \Rightarrow ((\varphi \rightarrow \psi) \Leftrightarrow (\varphi' \rightarrow \psi'))$.

Proof: Recall that $N\varphi$ is defined as $\neg\varphi \rightarrow$*false*.

We start with part (a).

1. $\vdash_{\mathbf{C}} (\neg(\varphi \wedge \psi) \rightarrow \mathit{false}) \Rightarrow (\neg(\varphi \wedge \psi) \rightarrow \neg\varphi)$	RC2
2. $\vdash_{\mathbf{C}} (\neg(\varphi \wedge \psi) \rightarrow \mathit{false}) \Rightarrow ((\neg(\varphi \wedge \psi) \wedge \neg\varphi) \rightarrow \mathit{false})$	1, C4
3. $\vdash_{\mathbf{C}} (\neg(\varphi \wedge \psi) \rightarrow \mathit{false}) \Rightarrow (\neg\varphi \rightarrow \mathit{false})$	2, RC1
4. $\vdash_{\mathbf{C}} N(\varphi \wedge \psi) \Rightarrow N\varphi$	3 rewritten

To prove part (b), we proceed as follows:

1. $\vdash_{\mathbf{C}} ((\neg\varphi \rightarrow \mathit{false}) \wedge (\neg\psi \rightarrow \mathit{false})) \Rightarrow ((\neg\varphi \vee \neg\psi) \rightarrow \mathit{false})$ C3
2. $\vdash_{\mathbf{C}} ((\neg\varphi \rightarrow \mathit{false}) \wedge (\neg\psi \rightarrow \mathit{false})) \Rightarrow (\neg(\varphi \wedge \psi) \rightarrow \mathit{false})$ 1, RC1
3. $\vdash_{\mathbf{C}} (N\varphi \wedge N\psi) \Rightarrow N(\varphi \wedge \psi)$ 2 rewritten

For part (c), we proceed as follows:

1. $\vdash_{\mathbf{C}} (N\varphi \wedge N(\varphi \Rightarrow \psi)) \Rightarrow N(\varphi \wedge (\varphi \Rightarrow \psi))$ (b)
2. $\vdash_{\mathbf{C}} (N\varphi \wedge N(\varphi \Rightarrow \psi)) \Rightarrow N(\varphi \wedge \psi)$ 1, RC1
3. $\vdash_{\mathbf{C}} (N\varphi \wedge N(\varphi \Rightarrow \psi)) \Rightarrow N\psi$ (a), 2

To prove part (d), we assume that $\vdash_{\mathbf{C}} \varphi$.

1. $\vdash_{\mathbf{C}} \varphi$ assumption
2. $\vdash_{\mathbf{C}} \mathit{true} \Leftrightarrow \varphi$ 1, C0
3. $\vdash_{\mathbf{C}} \mathit{false} \rightarrow \mathit{false}$ C1
4. $\vdash_{\mathbf{C}} \neg\varphi \rightarrow \mathit{false}$ 3, RC1

To prove part (e), we proceed as follows:

1. $\vdash_{\mathbf{C}} (\psi \wedge \varphi) \rightarrow \varphi$ C1, RC2
2. $\vdash_{\mathbf{C}} (\neg\varphi \rightarrow \mathit{false}) \Rightarrow (\neg\varphi \rightarrow \varphi) \wedge (\neg\varphi \rightarrow \psi)$ RC2
3. $((\neg\varphi \rightarrow \varphi) \wedge (\neg\varphi \rightarrow \psi)) \Rightarrow ((\psi \wedge \neg\varphi) \rightarrow \varphi$ C4
4. $\vdash_{\mathbf{C}} (((\psi \wedge \varphi) \rightarrow \varphi) \wedge ((\psi \wedge \neg\varphi) \rightarrow \varphi))) \Rightarrow (\psi \rightarrow \varphi)$ C3, RC1
5. $\vdash_{\mathbf{C}} N\varphi \Rightarrow (\psi \rightarrow \varphi)$ 1, 2, 3, 4

Finally, we prove part (f).

1. $\vdash_{\mathbf{C}} N(\varphi \Leftrightarrow \varphi') \Rightarrow (N(\varphi' \Rightarrow \varphi) \wedge N(\varphi \Rightarrow \varphi'))$ (a)
2. $\vdash_{\mathbf{C}} N(\varphi' \Rightarrow \varphi) \Rightarrow ((\varphi' \wedge \neg\varphi) \rightarrow \psi)$ definition of N, RC2
3. $\vdash_{\mathbf{C}} N(\varphi \Rightarrow \varphi') \Rightarrow (\varphi \rightarrow (\varphi \Rightarrow \varphi'))$ (e)
4. $\vdash_{\mathbf{C}} N(\varphi \Rightarrow \varphi') \Rightarrow (\varphi \rightarrow \varphi')$ 3, C1, C2, RC2
5. $\vdash_{\mathbf{C}} (\varphi \rightarrow \varphi') \wedge (\varphi \rightarrow \psi) \Rightarrow ((\varphi' \wedge \varphi) \rightarrow \psi)$ C4
6. $\vdash_{\mathbf{C}} N(\varphi \Leftrightarrow \varphi') \wedge (\varphi \rightarrow \psi) \Rightarrow (\varphi' \rightarrow \psi)$ 1, 2, 3, 4, 5, C3, RC1
7. $\vdash_{\mathbf{C}} N(\psi \Leftrightarrow \psi') \Rightarrow (\varphi' \rightarrow (\psi \Rightarrow \psi'))$ (e), RC2
8. $\vdash_{\mathbf{C}} N(\varphi \Leftrightarrow \varphi') \wedge N(\psi \Leftrightarrow \psi') \wedge (\varphi \rightarrow \psi) \Rightarrow (\varphi' \rightarrow \psi')$ 6, 7, C2, RC2

∎

Lemma 2.8.3: *PL is a qualitative plausibility structure. Moreover,* $(PL, w_V) \models \varphi$ *if and only if* $\varphi \in V$ *for all formulas* φ *and worlds* w_V.

Proof: We first need to establish that Pl_{w_V} is well-defined, for all $w_V \in W$. Let V be a maximal $\vdash_{\mathbf{C}}$-consistent set. We need to show that if $[\varphi]_{w_V} = [\varphi']_{w_V}$ and $[\psi]_{w_V} = [\psi']_{w_V}$, then $(\varphi \vee \psi) \rightarrow \psi \in V$ if and only $(\varphi' \vee \psi') \rightarrow \psi' \in V$.

We claim that $[\varphi]_{w_V} = [\varphi']_{w_V}$ if and only if $V/N \cup \{\neg(\varphi \Leftrightarrow \varphi')\}$ is inconsistent. The "if" direction is obvious. For the "only if" direction, assume that $V/N \cup \{\neg(\varphi \Leftrightarrow \varphi')\}$ is consistent. This implies that there is a maximal consistent set U such that $\Delta \cup \{\neg(\varphi \Leftrightarrow \varphi')\} \subseteq U$. Clearly, w_U is in the symmetric difference between $[\varphi]_{w_V}$ and $[\varphi']_{w_V}$. Thus, $[\varphi]_{w_V} \neq [\varphi']_{w_V}$.

Now assume that $[\varphi]_{w_V} = [\varphi']_{w_V}$ and $[\psi]_{w_V} = [\psi']_{w_V}$. Therefore, as we have just shown, $V/N \cup \{\neg(\varphi \Leftrightarrow \varphi')\}$ and $V/N \cup \{\neg(\psi \Leftrightarrow \psi)\}$ are both inconsistent. Thus, there exists a formula δ which is the conjunction of a finite number of formulas in V/N such that $\vdash_{\mathbf{C}} \delta \Rightarrow (\varphi \Leftrightarrow \varphi') \wedge (\psi \Leftrightarrow \psi)$. Using part (b) of Lemma A.5.1, and the fact that if $\varphi \in V/N$, then $N\varphi \in V$, we get that $N\delta \in V$. From parts (c) and (d) of Lemma A.5.1, we get that $\vdash_{\mathbf{C}} N\delta \Rightarrow N(\varphi \Leftrightarrow \varphi') \wedge N(\psi \Leftrightarrow \psi')$. Finally, using part (f) of Lemma A.5.1 we get that

$$\vdash_{\mathbf{C}} N\delta \Rightarrow ((\varphi \rightarrow \psi) \Leftrightarrow (\varphi' \rightarrow \psi')).$$

Thus, $\varphi \rightarrow \psi \in V$ if and only if $\varphi' \rightarrow \psi' \in V$. This suffices to prove that Pl_{w_V} is well-defined.

We now need to establish that *PL* is a plausibility structure. Note that the definition of Pl_{w_V} mirrors the construction in Lemma 2.4.1. It is easy to prove that the set of defaults $\{\varphi \rightarrow \psi \in V\}$ is closed under the KLM rules. Thus, we can immediately use the proof of Lemma 2.4.1 to show that $(W_{w_V}, \mathrm{Pl}_{w_V})$ is a plausibility space.

We next show that $(PL, w_V) \models \varphi$ if and only if $\varphi \in V$. This is done by induction on the structure of φ. The only case of interest is if φ is of the form $\varphi' \rightarrow \psi$. Here again the proof is identical to the proof of Lemma 2.4.1. The only difference is the use of axioms in system **C** instead of the corresponding rules in system **P**.

Finally, we have to prove that Pl_{w_V} satisfies A2 and A3. The proof is identical to the proof of Theorem 2.5.4. ∎

Proposition 2.8.4: *Let* $\mathcal{P} \subseteq \mathcal{S}_c^{QPL}$. *C5 is valid in* $\mathcal{S}$ *if and only if all structures in* $\mathcal{S}$ *are rational.*

Proof: To prove the "if" direction it suffices to show that C5 is sound in rational qualitative structures. Let $PL = (W, \mathcal{P}, \pi)$ be a rational qualitative plausibility structure. Assume that $(PL, w) \models (\varphi \rightarrow \psi)$ and $(PL, w) \models \neg(\varphi \wedge \xi \rightarrow \psi)$. We need to prove that $(PL, w) \models \varphi \rightarrow \neg\xi$. If $\mathrm{Pl}_w([\![\varphi]\!]_{(PL,w)}) = \bot$, then $(PL, w) \models \varphi \rightarrow \neg\xi$ vacuously, and we are done. Now assume that $\mathrm{Pl}_w([\![\varphi]\!]_{(PL,w)}) > \bot$. Let $A = [\![\varphi]\!]_{(PL,w)}$, $B = [\![\psi]\!]_{(PL,w)}$, and $C = [\![\xi]\!]_{(PL,w)}$. Since $(PL, w) \models \varphi \rightarrow \psi$, we have that $\mathrm{Pl}_w(A \cap B) > \mathrm{Pl}_w(A \cap \overline{B})$. Since Pl_w satisfies A5, we have that either $\mathrm{Pl}_w(A \cap B \cap C) > \mathrm{Pl}_w(A \cap \overline{B})$ or $\mathrm{Pl}_w(A \cap B \cap \overline{C}) > \mathrm{Pl}_w(A \cap \overline{B})$. However, since $(PL, w) \models \neg(\varphi \wedge \xi \rightarrow \psi)$, we have that $\mathrm{Pl}_w(A \cap B \cap C) \not> \mathrm{Pl}_w(A \cap \overline{B} \cap C)$. This implies that

$$\mathrm{Pl}_w(A \cap B \cap C) \not> \mathrm{Pl}_w(A \cap \overline{B}). \qquad \text{(A.1)}$$

Thus, we conclude that

$$\mathrm{Pl}_w(A \cap B \cap \overline{C}) > \mathrm{Pl}_w(A \cap \overline{B}). \tag{A.2}$$

Applying A4 with (A.2) as the antecedent, we get that either $\mathrm{Pl}_w(A \cap B \cap C) > \mathrm{Pl}_w(A \cap \overline{B})$ or $\mathrm{Pl}_w(A \cap B \cap \overline{C}) > \mathrm{Pl}_w(A \cap B \cap C)$. Since the former contradicts (A.1), we conclude that $\mathrm{Pl}_w(A \cap B \cap \overline{C}) > \mathrm{Pl}_w(A \cap B \cap C)$. Using A1, A2, and (A.2), we get that $\mathrm{Pl}_w(A \cap \overline{C}) \geq \mathrm{Pl}_w(A \cap B \cap \overline{C}) > \mathrm{Pl}_w((A \cap B \cap C) \cup (A \cap \overline{B})) \geq \mathrm{Pl}_w(A \cap C)$. Thus, $(\mathit{PL}, w) \models \varphi \rightarrow \neg\xi$.

To prove the "only if" direction, we have to show that if there is some $\mathit{PL} = (W, \mathcal{P}, \pi)$ in $\mathcal{S}$ that is not rational, then C5 is violated.

Assume that PL does not satisfy A4. Since we have assumed that $\mathcal{F}_w = \{[\![\varphi]\!]_{(\mathit{PL},w)} : \varphi \in \mathcal{L}\}$, there is a world w, and formulas φ, ψ, and ξ such that $[\![\varphi]\!]_{(\mathit{PL},w)}$, $[\![\psi]\!]_{(\mathit{PL},w)}$, and $[\![\xi]\!]_{(\mathit{PL},w)}$ are pairwise disjoint, $\mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)}) > \mathrm{Pl}_w([\![\psi]\!]_{(\mathit{PL},w)})$, and yet $\mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)}) \not> \mathrm{Pl}_w([\![\psi]\!]_{(\mathit{PL},w)})$ and $\mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)}) \not< \mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)})$. Since $\mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)}) > \mathrm{Pl}_w([\![\psi]\!]_{(\mathit{PL},w)})$, we have that $\mathrm{Pl}_w([\![\varphi \vee \xi]\!]_{(\mathit{PL},w)}) > \mathrm{Pl}_w([\![\psi]\!]_{(\mathit{PL},w)})$. Thus, $(\mathit{PL}, w) \models (\varphi \vee \psi \vee \xi) \rightarrow \neg\psi$. Moreover, since $\mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)}) \not< \mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)})$, we have that $\mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)}) > \perp$. Since we also have that $\mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)}) \not> \mathrm{Pl}_w([\![\psi]\!]_{(\mathit{PL},w)})$, we conclude that $(\mathit{PL}, w) \models \neg((\xi \vee \psi) \rightarrow \neg\psi)$. Finally, since $\mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)}) \not> \mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)})$, we have that $\mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)}) \not> \mathrm{Pl}_w([\![\xi \vee \psi]\!]_{(\mathit{PL},w)})$. Thus, $(\mathit{PL}, w) \models \neg((\varphi \vee \psi \vee \xi) \rightarrow \neg(\psi \vee \xi))$. Define α as $\varphi \vee \psi \vee \xi$, β as $\psi \vee \xi$ and γ as $\neg\psi$. We just proved that $(\mathit{PL}, w) \models (\alpha \rightarrow \gamma) \wedge \neg(\alpha \rightarrow \neg\beta) \wedge \neg(\alpha \wedge \beta \rightarrow \gamma)$, which contradicts C5.

Now assume that PL does not satisfy A5. Then there is a world w, and formulas φ, ψ, and ξ, such that $[\![\varphi]\!]_{(\mathit{PL},w)}$, $[\![\psi]\!]_{(\mathit{PL},w)}$, and $[\![\xi]\!]_{(\mathit{PL},w)}$ are pairwise disjoint, $\mathrm{Pl}_w([\![\varphi \vee \psi]\!]_{(\mathit{PL},w)}) > \mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)})$, and yet $\mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)}) \not> \mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)})$ and $\mathrm{Pl}_w([\![\psi]\!]_{(\mathit{PL},w)}) \not> \mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)})$. Since $\mathrm{Pl}_w([\![\varphi \vee \psi]\!]_{(\mathit{PL},w)}) > \mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)}) \geq \perp$, either $\mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)}) > \perp$ or $\mathrm{Pl}_w([\![\psi]\!]_{(\mathit{PL},w)}) > \perp$, for otherwise, using A3 we would get that $\mathrm{Pl}_w([\![\varphi \vee \psi]\!]_{(\mathit{PL},w)}) = \perp$. Without loss of generality, we assume that $\mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)}) > \perp$. Since $\mathrm{Pl}_w([\![\varphi \vee \psi]\!]_{(\mathit{PL},w)}) > \mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)})$, we have that $(\mathit{PL}, w) \models (\varphi \vee \psi \vee \xi) \rightarrow \neg\xi$. Since $\mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)}) > \perp$ and $\mathrm{Pl}_w([\![\varphi]\!]_{(\mathit{PL},w)}) \not> \mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)})$, we have that $(\mathit{PL}, w) \models \neg((\varphi \vee \xi) \rightarrow \neg\xi)$. Finally, since $\mathrm{Pl}_w([\![\psi]\!]_{(\mathit{PL},w)}) \not> \mathrm{Pl}_w([\![\xi]\!]_{(\mathit{PL},w)})$, we have that $\mathrm{Pl}_w([\![\psi]\!]_{(\mathit{PL},w)}) \not> \mathrm{Pl}_w([\![\varphi \vee \xi]\!]_{(\mathit{PL},w)})$, and thus $(\mathit{PL}, w) \models \neg((\varphi \vee \psi \vee \xi) \rightarrow \psi))$. Define α as $\varphi \vee \psi \vee \xi$, β as $\neg\psi$ and γ as $\neg\xi$. We have proved that $(\mathit{PL}, w) \models (\alpha \rightarrow \gamma) \wedge \neg(\alpha \rightarrow \neg\beta) \neg(\alpha \wedge \beta \rightarrow \gamma)$, which contradicts C5. ∎

We now prove Theorem 2.8.5. We start with a lemma:

Lemma A.5.2: *Let* (W, Pl) *be a qualitative plausibility space, and let* $<^*$ *be a binary relation on subsets of* W *such that* $A <^* B$ *if there is some set* $C \subseteq W$ *such that* $A \cap C = \emptyset$ *and* $\mathrm{Pl}(A) < \mathrm{Pl}(C) \leq \mathrm{Pl}(B)$. *Then*

(a) $<^*$ *is a strict order; that is irreflexive, transitive and anti-symmetric,*

(b) if $A \cap B = \emptyset$*, then* $A <^* B$ *if and only if* $\mathrm{Pl}(A) < \mathrm{Pl}(B)$*,*

(c) if Pl *is rational, then* $<^*$ *is modular, and*

(d) if Pl *is rational and* $A <^* (A \cup B)$*, then* $B \not<^* (A \cup B)$*.*

Proof: We start with part (a). The definition of $<^*$ implies that $A <^* B$ only if Pl$(A) <$ Pl(B). Thus, we get that $<^*$ is irreflexive and anti-symmetric. We now show that $<^*$ is transitive. Assume that $A <^* B$ and $B <^* C$. Since $A <^* B$, there is a set D such that $A \cap D = \emptyset$, and Pl$(A) <$ Pl$(D) \leq$ Pl(B). Moreover, since $B <^* C$, we have that Pl$(B) <$ Pl(C). Thus, we get that Pl$(D) <$ Pl(C), and hence $A <^* C$.

For part (b), let A and B be disjoint sets. If Pl$(A) <$ Pl(B), then $A <^* B$ using the set $C = B$. On the other hand, if $A <^* B$, then Pl$(A) <$ Pl(B).

For part (c), assume that Pl is rational, and that $A <^* B$. Let $C \subseteq W$. We have to show that either $A <^* C$ or $C <^* B$. Since $A <^* B$, there is a set D such that $A \cap D = \emptyset$ and Pl$(A) <$ Pl$(D) \leq$ Pl(B). Since D is the disjoint union of $(D \cap C)$ and $D - C$, we can apply A5 and get that either Pl$(A) <$ Pl$(D \cap C)$ or Pl$(A) <$ Pl$(D - C)$. If Pl$(A) <$ Pl$(D \cap C)$, then since Pl$(D \cap C) \leq$ Pl(C), we get that $A <^* C$ and we are done. If Pl$(A) <$ Pl$(D - C)$, since A, $D - C$ and $C - A$ are pairwise disjoint, we get from A4 that either Pl$(A) <$ Pl$(C - A)$ or Pl$(C - A) <$ Pl$(D - C)$. If Pl$(A) <$ Pl$(C - A)$, we again get that $A <^* C$ and we are done. If Pl$(C - A) <$ Pl$(D - C)$, using A2, we get that Pl$((C - A) \cup A) <$ Pl$(D - C)$. We conclude that Pl$(C) <$ Pl$(D - C)$. Since Pl$(D - C) \leq$ Pl$(D) \leq$ Pl(B), we get that $C <^* B$.

Finally, we prove part (d). Assume that $A <^* (A \cup B)$. Without loss of generality, we can also assume that $B \cap A = \emptyset$ (if not, replace B by $B - A$). We want to show that $B \not<^* (A \cup B)$. By way of contradiction, assume that $B <^* (A \cup B)$. Since $A <^* B$, there is a set C such that $A \cap C = \emptyset$ and Pl$(A) <$ Pl$(C) \leq$ Pl$(A \cup B)$. Since $C \subseteq (C - B) \cup B$, this implies that Pl$(A) <$ Pl$((C - B) \cup B)$. Since A, B, and $C - B$ are pairwise disjoint, we can apply A5 to get that either Pl$(A) <$ Pl(B) or Pl$(A) <$ Pl$(C - B)$. If Pl$(A) <$ Pl(B), since we assumed that $B <^* (A \cup B)$, there is a set D such that $D \cap B = \emptyset$ and Pl$(B) <$ Pl$(D) \leq$ Pl$(A \cup B)$. This implies that Pl$(B) <$ Pl$((D - A) \cup B)$. Moreover, since Pl$(A) <$ Pl(B), we also have that Pl$(A) <$ Pl$((D - A) \cup B)$. Since A, B and $D - A$ are pairwise disjoint, we can apply A2 to get that Pl$(A \cup B) <$ Pl$(D - A) \leq$ Pl(D). But this contradicts the assumption that Pl$(D) \leq$ Pl$(A \cup B)$. If Pl$(A) <$ Pl$(C - B)$, since A, B, and $C - B$ are pairwise disjoint, we have by A4 that either Pl$(A) <$ Pl(B) or Pl$(B) <$ Pl$(C - B)$. We have already seen that if Pl$(A) <$ Pl(B) we get a contradiction. Now assume that Pl$(B) <$ Pl$(C - B)$. Then, since we also have that Pl$(A) <$ Pl$(C - B)$, we can apply A2 to get that Pl$(A \cup B) <$ Pl$(C - B) \leq$ Pl(C). But this contradicts our assumption that Pl$(C) \leq$ Pl$(A \cup B)$. Thus, we must have $B \not<^* (A \cup B)$. ∎

Theorem 2.8.5: *Let* $(W, \mathcal{F}, \text{Pl})$ *be a rational qualitative plausibility space. Then there is a plausibility space* $(W, \mathcal{F}, \text{Pl}')$ *such that* Pl$'$ *is a ranking, and* Pl *and* Pl$'$ *are default-isomorphic; that is,* Pl$(A) >$ Pl(B) *if and only if* Pl$'(A) >$ Pl$'(B)$ *for all disjoint sets* $A, B \in \mathcal{F}$.

Proof: Let $<^*$ be the relation defined in Lemma A.5.2. Define a relation $\approx^*$ on sets such that $A \approx^* B$ if neither $A <^* B$ nor $B <^* A$. Since $<^*$ is modular, standard (and straightforward) arguments show that $\approx^*$ is an equivalence relation. We construct a new plausibility measure based on these two ordering. Let $\mathcal{F}/{\approx^*}$ be the set of equivalence classes of measurable sets. Let $\leq^*$ be the the total order on $\mathcal{F}/\approx^*$ induced by $<^*$. Let Pl$'$ be a plausibility measure on $\mathcal{F}$ whose range is $\mathcal{F}/\approx^*$, defined so that $\mathrm{Pl}'(A) = [A]$ where $[A]$ is the equivalence class containing A. We have to show that Pl′ is a plausibility measure. Assume that $A \subseteq B$. Then $\mathrm{Pl}(A) \leq \mathrm{Pl}(B)$ and clearly $B \not<^* A$. Thus, either $A <^* B$ or $A \approx^* B$. We conclude that $\mathrm{Pl}'(A) \leq \mathrm{Pl}'(B)$, as desired. Since $\leq^*$ is a total order, Pl′ satisfies A4′. Using Lemma A.5.2 part (d), we get that $\mathrm{Pl}'(A \cup B) = \max(\mathrm{Pl}'(A), \mathrm{Pl}'(B))$. Thus, Pl′ also satisfies A5′, and hence is a ranking. Finally, we have to show that Pl′ is default-isomorphic to Pl. Let A and B be disjoint events. By to Lemma A.5.2, $A <^* B$ if and only if $\mathrm{Pl}(A) < \mathrm{Pl}(B)$. Thus, we conclude that $\mathrm{Pl}'(A) < \mathrm{Pl}'(B)$ if and only if $\mathrm{Pl}(A) < \mathrm{Pl}(B)$. ∎

Theorem 2.8.8: *Let $\varphi \in \mathcal{L}^C$, let $\mathcal{A}$ be a subset of $\{R, N, X, U\}$, and let A be the corresponding subset of $\{C5, C6, C7, C8\}$. Then φ is valid in subset of $\mathcal{S}_c^{QPL}$ that satisfies $\mathcal{A}$ if and only if φ is provable from system* **C**$+A$.

Proof: Again, we focus on completeness. We obtain completeness in each case by modifying the proof of Theorem 2.8.2. We construct a canonical model as in that proof, checking consistency with the extended axiom system. The resulting structure has the property that $(PL, w_V) \models \varphi$ if and only if $\varphi \in V$. We just need to show that this structure also satisfies the corresponding semantic restrictions.

Assume that C5 is included as an axiom. Then we get that C5 is valid in *PL*. Using Proposition 2.8.4 we get that *PL* is rational.

Assume that C6 is included as an axiom. If *PL* is not normal, then there is some world w_V where $\mathrm{Pl}_{w_V}([\![true]\!]_{(PL,w_V)}) = \mathrm{Pl}_{w_V}([\![false]\!]_{(PL,w_V)})$. We get that $(PL, w_V) \models true \rightarrow false$, which contradicts C6.

Assume that C7 is included as an axiom. We want to show that *PL* is reflexive. Let V be a maximal consistent set. C7 implies that $V/N \subseteq V$, for if $N\varphi \in V$, then by C7, $\varphi \in V$. Thus, we conclude that $w_V \in W_{w_V}$. Assume, by way of contradiction, that there is a formula φ such that $w_V \in [\![\varphi]\!]_{(PL,w_V)}$ and $\mathrm{Pl}_{w_V}([\![\varphi]\!]_{(PL,w_V)}) = \perp$. Since $w_V \in [\![\varphi]\!]_{(PL,w_V)}$, we have that $\varphi \in V$. Since $\mathrm{Pl}_{w_V}([\![\varphi]\!]_{(PL,w_V)}) = \perp$, we have that $\varphi \rightarrow false \in V$. Thus, $N\neg\varphi \in V$. Using C7, we conclude that $\neg\varphi \in V$, which contradicts the assumption that V is a consistent set. Thus, we conclude that $\mathrm{Pl}_{w_V}(\varphi) > \perp$ for all $\varphi \in V$, as desired.

Finally, assume that C8 is included as an axiom. We want to show that *PL* is normal, that is, we want to show that if $w_U \in W_{w_V}$, then $\mathcal{P}(w_U) = \mathcal{P}(w_V)$. Since the construction of $\mathcal{P}(w_V)$ is determined only by the conditional statements in V, it suffices to show that $\varphi \rightarrow \psi \in U$ if and only if $\varphi \rightarrow \psi \in V$. Assume that $\varphi \rightarrow \psi \in V$. Then, using C8, we get that $N(\varphi \rightarrow \psi) \in V$. Since $w_U \in W_{w_V}$, we have that $V/N \subseteq U$. We conclude that $\varphi \rightarrow \psi \in U$. Now assume that $\varphi \rightarrow \psi \notin V$. Then since V is a maximal consistent set, we have that $\neg(\varphi \rightarrow \psi) \in V$. Applying C8, we get that $\neg(\varphi \rightarrow \psi) \in U$, which implies that $\varphi \rightarrow \psi \neq U$. ∎

Finally, we prove Theorem 2.8.10. We start with a lemma:

Lemma A.5.3: *Let $\varphi \in \mathcal{L}^C$ and let $\mathcal{A}$ be a subset of $\{R, N\}$. If φ is satisfiable in a structure satisfying $\mathcal{A}$, it is satisfiable in a finite structure satisfying $\mathcal{A}$.*

Proof: Let $\varphi \in \mathcal{L}^C$ and let $PL = (W, \mathcal{P}, \pi)$ be a structure satisfying $\mathcal{A}$ such that $(PL, w) \models \varphi$ for some $w \in W$. We now construct a finite structure PL' that satisfies φ. The key idea is that when evaluating φ we only examine subformulas of φ. Thus, we are interested in distinguishing only between worlds that differ in the evaluation of some subformula of φ.

We now make this argument precise. Let $Sub(\varphi)$ be the set of subformulas of φ. We partition W into equivalence classes: For $w \in W$, define $[w] = \{w' \in W : \forall \psi \in Sub(\varphi), (PL, w) \models \psi \text{ if and only if } (PL, w') \models \psi\}$. Thus, $[w]$ contains all the worlds that are indistinguishable, for the purpose of evaluating φ, from w. We now choose, arbitrarily, a *representative* world $\hat{w} \in [w]$ for each equivalence class. These definitions extend to sets of worlds: For $A \subseteq W$, define $[A] = \cup_{w \in A}[w]$, and $\hat{A} = \{\hat{w} : w \in A\}$.

We construct PL' as follows. We set $W' = \hat{W}$. Clearly W' is finite, since there are only a finite number of subformulas of φ. Let π' be the restriction of π to W', and let $\mathcal{P}'(\hat{w}) = (W'_{\hat{w}}, \mathrm{Pl}'_{\hat{w}})$, where $W'_{\hat{w}} = \hat{W}_{\hat{w}}$, and $\mathrm{Pl}'_{\hat{w}}$ is defined so that $\mathrm{Pl}'_{\hat{w}}(A') \leq \mathrm{Pl}_{\hat{w}}(B')$ if $\mathrm{Pl}_{\hat{w}}([A'] \cap W_{\hat{w}}) \leq \mathrm{Pl}_{\hat{w}}([B'] \cap W_{\hat{w}})$.

We need show that PL' is a qualitative plausibility structure. This is easy to prove since $[\cdot]$ preserves subsets, unions, and disjointness of sets: if $A' \subseteq B'$, then $[A'] \subseteq [B']$; $[A' \cup B'] = [A'] \cup [B']$; and $[A' \cap B'] = [A'] \cap [B']$.

Next, we need to show that if $(PL, w) \models \varphi$, then $(PL', \hat{w}) \models \varphi$. In fact, we prove this property for every subformula ψ of φ. The proof is by induction on the structure of of ψ. The only case of interest is for conditional formulas $\psi \rightarrow \xi \in Sub(\varphi)$. This case follows easily once we notice that if the induction hypothesis holds for ψ, then $[[\![\psi]\!]_{(PL', \hat{w})}] \cap W_{\hat{w}} = [\![\psi]\!]_{(PL, \hat{w})} \cap W_{\hat{w}}$.

Finally, we have to show that if PL is rational or normal, then so is PL'. Again, this follows easily from the properties of $[\cdot]$. ∎

Lemma A.5.4: *Let $(W, \mathcal{F}, \mathrm{Pl})$ be a rational and normal qualitative plausibility space such that $\mathcal{F}$ is finite. Then there is a κ-ranking κ on W and a possibility measure Poss on W such that for all disjoint $A, B \in \mathcal{F}$, $\mathrm{Pl}(A) > \mathrm{Pl}(B)$ if and only if $\kappa(A) < \kappa(B)$ and $\mathrm{Pl}(A) > \mathrm{Pl}(B)$ if and only if $Poss(A) > Poss(B)$.*

Proof: By Theorem 2.8.5, there is a ranked plausibility space $(W, \mathcal{F}, \mathrm{Pl}')$ such that Pl' is default-isomorphic to Pl. Since $\mathcal{F}$ is finite, the set $\{d : \exists A \in \mathcal{F}, \mathrm{Pl}'(A) = d\}$ is finite. Moreover, since Pl' is a ranking, this set is totally ordered. Let $d_0 > d_1 > \ldots > d_n$ be an ordered enumeration of this set of values.

Let κ be a κ-ranking such that $\kappa(A) = k$ if $\mathrm{Pl}'(A) = d_k > \perp$ and $\kappa(A) = \infty$ if $\mathrm{Pl}'(A) = \perp$. Similarly, let Poss be a possibility measure such that $\mathrm{Poss}(A) = 1 - k/n$ if $\mathrm{Pl}'(A) = d_k > \perp$ and $\mathrm{Poss}(A) = 0$ if $\mathrm{Pl}'(A) = \perp$. It easy to see that since Pl' is a ranking, we get that $\kappa(A \cup B) = \min(\kappa(A), \kappa(B))$ and that $\mathrm{Poss}(A \cup B) = \max(\mathrm{Poss}(A), \mathrm{Poss}(B))$. It is also easy to see that both measures are isomorphic to Pl' and thus default-isomorphic to Pl. ∎

Theorem 2.8.10: *(a)* **C**$+\{C6\}$ *is a sound and complete axiomatization of $\mathcal{S}_c^{\epsilon}$.*

(b) **C** $+\{C5, C6\}$ *is a sound and complete axiomatization of $\mathcal{S}_c^{\kappa}$ and $\mathcal{S}_c^{Poss}$.*

Proof: Part (a) is an immediate corollary of Theorems 2.6.3 and 2.8.8.

For Part (b), as usual, it is easy to verify soundness. For completeness, it suffices to show that if φ is consistent with system **C**+{C5,C6}, then it is satisfiable in $\mathcal{S}_c^{\kappa}$ and $\mathcal{S}_c^{Poss}$. Suppose that φ is consistent with system **C**+{C5,C6}. By Theorem 2.8.8, φ is satisfiable in a rational and normal structure in $\mathcal{S}_C^{QPL}$. By Lemma A.5.3, we can assume that this structure is finite. The result now follows from Lemma A.5.4. ∎

Appendix B

Proofs for Chapter 3

B.1 Proofs for Section 3.6

Theorem 3.6.2: *K (resp., K45, KD45) is a sound and complete axiomatization for $\mathcal{L}^B$ with respect to $\mathcal{M}$ (resp., $\mathcal{M}^{\text{CONS}}$, $\mathcal{M}^{\text{CONS,NORM}}$).*

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 $\varphi \in \mathcal{L}^B$, we have $(M, w) \models \varphi$ if and only if $(M^+, w) \models \varphi$. Completeness then follows from Theorem 3.2.1.

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

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 $\mathrm{Pl}_{(w,i)}(W_{(w,i)}) > \perp$. ∎

Theorem 3.6.3: *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: Again soundness is straightforward, so we focus on completeness. We sketch a completeness proof following the usual Makinson [Makinson 1966] style of proof. We describe only the parts that are different from the standard proofs. See, for example, Halpern and Moses [Halpern and Moses 1992] for details.

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

Let V be a set of formulas and AX an axiom system. We say that V is *AX-consistent* if for all $\varphi_1, \ldots \varphi_n \in V$, it is not the case that $\mathrm{AX} \vdash \neg(\varphi_1 \wedge \ldots \wedge \varphi_n)$. The set V is a *maximal* consistent set if it is consistent, and for each formula φ, either $\varphi \in V$ or $\neg\varphi \in V$.

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

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

- $W = \{w_V :$ V is a maximal $\mathrm{AX}^{\mathrm{KB}}$-consistent set of formulas$\}$
- $\pi(w_V)(p) =$ **true** if and only if $p \in V$
- $\mathcal{K}_i = \{(w_V, w_U) : V/K_i \subseteq U\}$
- $\mathcal{P}_i(w_V) = (W_{(w_V,i)}, \mathrm{Pl}_{(w_V,i)})$, 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 [Halpern and Moses 1992]). Using a standard induction argument, we can verify that $(M^{\mathrm{KB}}, w_V) \models \varphi$ if and only if $\varphi \in V$.

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

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

Let $M^{\mathrm{KB,CONS,NORM}}$ be the canonical model constructed for $\mathrm{AX}^{\mathrm{KB,CONS,NORM}}$. The argument above shows that $M^{\mathrm{KB,CONS,NORM}}$ 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 $\varphi_1, \ldots, \varphi_m \in V/B_i$ such that $\vdash \neg(\varphi_1 \wedge \ldots \wedge \varphi_m)$. Since $\varphi_1, \ldots, \varphi_n \in V/B_i$, we conclude that $B_i\varphi_1, \ldots, B_i\varphi_m \in V$. Using the K45 axioms for B_i, standard arguments show that $B_i(\varphi_1, \ldots, \varphi_n) \in V$, and hence that $B_i(\mathit{false}) \in V$, which contradicts the consistency of V. ∎

Lemma 3.6.5: *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 $\varphi \in \mathcal{L}^{KB}$ that includes only the modalities K_i and B_i, $(M, w) \models \varphi$ if and only if $(M, w) \models \varphi^*$, 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 $w' \in \mathcal{K}_i(w)$, $(M, w') \models \varphi$ if and only if $(M, w') \models \varphi^*$. The only interesting case is if φ has the from $B_i\varphi'$. Suppose that $(M, w') \models B_i\varphi'$. This implies that $(M, w') \models \mathit{true} \rightarrow_i \varphi'$, i.e., for all $w'' \in$ MP$(\mathcal{P}_i(w'))$ we have $(M, w'') \models \varphi'$. Now let $w'' \in \mathcal{K}_i(w')$. If $(M, w'') \models \neg\alpha$, then $(M, w'') \models \alpha \Rightarrow (\varphi')^*$. If $(M, w'') \models \alpha$ then, by definition, $w'' \in$ MP$(\mathcal{P}_i(w))$, and since we assumed SDP, MP$(\mathcal{P}_i(w')) =$ MP$(\mathcal{P}_i(w))$. Thus, we conclude that $(M, w'') \models \varphi'$, and using the induction hypothesis we get that $(M, w'') \models (\varphi')^*$. We conclude that all worlds in $\mathcal{K}_i(w')$ satisfy $\alpha \Rightarrow (\varphi')^*$, and thus $(M, w') \models K_i(\alpha \Rightarrow (\varphi')^*)$. Now assume that $(M, w') \models K_i(\alpha \Rightarrow (\varphi')^*)$. Let w'' be any world in $\mathcal{K}_i(w')$. Since we assumed SDP, we have that MP$(\mathcal{P}_i(w'')) =$ 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 MP$(\mathcal{P}_i(w''))$ satisfy φ'. Hence, $(M, w'') \models \mathit{true} \rightarrow_i \varphi'$. Since this is true for all $w'' \in \mathcal{K}_i(w')$ we conclude that $(M, w') \models B_i\varphi'$. ∎

B.2 Proofs for Section 3.7

Theorem 3.7.1: *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 Section 2.8.

We proceed as follows. Let V be a set of formulas, define $V/K_i = \{\varphi : K_i\varphi \in V\}$ and $V/N_i = \{\varphi : N_i\varphi \in V\}$. We construct a canonical model as follows. Let $\mathcal{M} = (W, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{P}_i, \ldots, \mathcal{P}_n)$, where

- $W = \{w_V :$ 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
  - $W_{(w_V,i)} = \{w_U : V/N_i \subseteq U\}$,
  - $\mathcal{F}_{(w_V,i)} = \{[\varphi]_{(w_V,i)} : \varphi \in \mathcal{L}^{KC}\}$ where $[\varphi]_{(w_V,i)} = \{w_U \in W_{(w_V,i)} : \varphi \in U\}$, and
  - $\mathrm{Pl}_{(w_V,i)}$ is such that $\mathrm{Pl}_{(w_V,i)}([\varphi]_{(w_V,i)}) \leq \mathrm{Pl}_{(w_V,i)}([\psi]_{(w_V,i)})$ if and only if $(\varphi \vee \psi) \rightarrow_i \psi \in V$.

We need to verify that M is indeed a model in $\mathcal{M}$. Using standard arguments it is easy to show that the $\mathcal{K}_i$ relations are equivalence relations. In the proof of Theorem 2.8.2 we prove that $\mathcal{P}_i(w_V)$ is a well-defined qualitative plausibility space.

Finally, we have to show that $(M, w_V) \models \varphi$ if and only if $\varphi \in V$. As usual, this is done by induction on the structure of φ. We use the standard argument for formulas of the form $K_i\varphi$ and arguments from the proof of Theorem 2.8.2 for formulas of the from $\varphi \rightarrow_i \psi$. We omit the details here. ∎

Theorem 3.7.2: *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 3.7.1. We construct a canonical model as in that proof, checking consistency with the extended axiom system. The resulting model is in $\mathcal{M}$ and has the property that $(M, w_V) \models \varphi$ if and only if $\varphi \in V$. We just need to show that this model 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 $\varphi \rightarrow_i \psi \in V$ if and only if $\varphi \rightarrow_i \psi \in U$, since these statements determine $\mathcal{P}_i$ in our construction. Assume $\varphi \rightarrow_i \psi \in V$. Then, according to C10, $K_i(\varphi \rightarrow_i \psi) \in V$, and thus $\varphi \rightarrow_i \psi \in V/K_i$. Recall that $w_U \in \mathcal{K}_i(w_V)$ only if $V/K_i \subseteq U$. We conclude that $\varphi \rightarrow_i \psi \in U$. 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 Theorem 2.8.8 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 Theorem 2.8.8 goes through unchanged; we do not repeat it here. ∎

Theorem 3.7.3: *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 $2^{|Sub(\varphi)|}$ worlds.*

Proof: The proof of this theorem relies on techniques from [Friedman and Halpern 1996b]. We sketch only the main steps here. The proof is based on a standard filtration argument.

Suppose there is a model M and a world w in M such that $(M, w) \models \varphi$. We now construct a small model M'. Let $Sub^+(\varphi) = Sub(\varphi) \cup \{\neg\varphi : \varphi \in Sub(\varphi)\}$. We say that $V \subseteq Sub^+(\varphi)$ is an *atom* if for each $\varphi \in Sub(\varphi)$, either $\varphi \in V$ or $\neg\varphi \in V$. We say that a world w in M satisfies an atom V if for all $\varphi \in V$, we have $(M, w) \models \varphi$. 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 [Friedman and Halpern 1996b] show that $(M', [w]) \models \psi$ if and only if $(M, w) \models \psi$ for all $\psi \in Sub(\varphi)$; 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 [Friedman and Halpern 1996b]. 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 $\psi {\rightarrow}_i \chi \in Sub(\varphi)$, 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 $\psi \in Sub(\varphi)$. ∎

Theorem 3.7.5: *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 $|\mathrm{Sub}(\varphi)|^3$ 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 [Friedman and Halpern 1996b] 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 $|Sub(\varphi)|^2$ worlds.

Choose $w_0 \in W$ such that $(M, w_0) \models \varphi$. For each formula $\neg K_1\psi \in Sub(\varphi)$ 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 $\{w_0\} \cup \{w_\psi : \neg K_1\psi \in Sub(\varphi)\}$. Note that the cardinality of T is at most $|Sub(\varphi)|$. 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 $|Sub(\varphi)|^3$. 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 $(M', w_0) \models \varphi$, so φ is satisfiable in a small preferential structure. ∎

Theorem 3.7.6: *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 [Fagin and Halpern 1994; Friedman and Halpern 1996b; Halpern and Moses 1992]. 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 [Halpern and Moses 1992]. For the exponential lower bound we use exactly the lower bound described Fagin and Halpern [Fagin and Halpern 1994] for the combination of knowledge and probability (which is in turn based on the lower bound for PDL [Fischer and Ladner 1979]). This lower bound construction uses only formulas involving K_i and probabilistic statements of the form $w_i(\varphi) = 1$ (i.e., the probability of φ is 1). Since $N_i\varphi$ has exactly the same properties as $w_i(\varphi) = 1$, 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 model 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 [Halpern and Moses 1992] or that PDL is decidable in deterministic exponential time [Pratt 1979]. ∎

Theorem 3.7.7: *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: 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 3.7.5, φ is satisfiable in a structure satisfying $\mathcal{A}$ if and only if φ is satisfiable in a structure M of size polynomial in $|\varphi|$. 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 [Halpern and Moses 1992; Friedman and Halpern 1996b]). ∎

Appendix C

Proofs for Chapter 4

C.1 Proofs for Section 4.4

Theorem 4.4.1: *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 system $\mathcal{I}$ such that if $\varphi \in \mathcal{L}^{KCT}$ is consistent, then φ is satisfied in $\mathcal{I}$. The outline of the proof is similar to that of Theorem 3.7.1.

We proceed as follows. Let V be a maximal AX^T-consistent set of formulas in $\mathcal{L}^{KCT}$. We define $V/\bigcirc = \{\varphi : \bigcirc\varphi \in V\}$. We claim that $V/\bigcirc$ is also a maximal AX^T-consistent set. To show that $V/\bigcirc$ is maximal, assume that $\varphi \notin V/\bigcirc$. Then $\bigcirc\varphi \notin V$. From axiom T2, we have that $\bigcirc\neg\varphi \in V$, and thus, $\neg\varphi \in V/\bigcirc$. This shows that $V/\bigcirc$ is maximal. To show that $V/\bigcirc$ is AX^T-consistent, assume that there are formulas $\varphi_1, \dots \varphi_n \in V/\bigcirc$ such that $\vdash_{\mathrm{AX}^T} \neg(\varphi_1 \wedge \dots \wedge \varphi_n)$. From K1, T1 and RT1 we get that *false* $\in V/\bigcirc$. Thus, $\bigcirc$*false* $\in V$. Using T2 we get that $\neg\bigcirc$*true* $\in V$. Using RT1, however, we get that $\bigcirc$*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 system as follows. Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}_1, \dots, \mathcal{P}_n)$, 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
  - $r_i^V(m) = (V/\bigcirc^m)/K_i$,
- $\pi(r^V, m)(p) =$ **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 $\mathrm{Pl}_{(r^V,m,i)}([\varphi]_{(r^V,m,i)}) \leq \mathrm{Pl}_{(r^V,m,i)}([\psi]_{(r^V,m,i)})$ if and only if $(\varphi \vee \psi) {\rightarrow_i} \psi \in V/\bigcirc^m$, where $[\varphi]_{(r^V,m,i)} = \{(r^U, k) \in W_{(r^V,m,i)} : \varphi \in U/\bigcirc^k\}$.

Using the arguments in the proof of Theorem 2.8.2, we can show that $\mathcal{P}_i(r, m)$ is well-defined for all i. Finally, we have to show that $(\mathcal{I}, r^V, m) \models \varphi$ if and only if $\varphi \in r_e^V(m)$. As usual, this is done by induction on the structure of φ. This is identical to the proof in of Theorem 3.7.1 except for the $\bigcirc$ modality, which is handled by standard arguments. We omit the details here. ∎

Theorem 4.4.2: *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 4.4.1. We construct a canonical system as in that proof, checking consistency with the extended axiom system. The resulting system has the property that $(\mathcal{I}, r^V, m) \models \varphi$ if and only if $\varphi \in V/\bigcirc^m$. 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 the proofs of Theorems 2.8.8 and 3.7.2. ∎

C.2 Proofs for Section 4.5

Proposition 4.6.2: *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 $(\mathcal{I}, r, m) \models \bigcirc(\varphi \wedge \psi) {\rightarrow_i} \bigcirc\xi$.*

Proof: Expanding the definition we get that $\mathcal{R}([\![\bigcirc(\varphi \wedge \psi)]\!]_{(r,m)}) = \{r' \in W_{(r,i)} : (r', m) \sim_i (r, m), (r', m+1) \models \varphi \wedge \psi\}$. 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 $(r, m+1) \models \varphi$. We conclude that $\mathcal{R}([\![\bigcirc(\varphi \wedge \psi)]\!]_{(r,m)}) = \mathcal{R}([\![\psi]\!]_{(r,m+1)})$. The lemma now follows directly from Proposition 4.5.1. ∎

C.3 Proofs for Section 4.6

Lemma 4.6.5: *Let $\mathcal{I}$ be a synchronous static system satisfying PRIOR, RANK, SDP, and perfect recall that has finite branching. Then $(\mathcal{I}, r, m) \models B_i\varphi \equiv B_i\bigcirc B_i\varphi$ 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 $(\mathcal{I}, r, m) \models B_i\varphi$. If $\mathrm{Pl}_{(r,m,i)}(W_{(r,m,i)}) = \bot$, then $(\mathcal{I}, r, m) \models B_i\bigcirc B_i\varphi$ vacuously. If $\mathrm{Pl}_{(r,m,i)}(W_{(r,m,i)}) \ne \bot$, then $\mathrm{Pl}_{(r,m,i)}([\![\varphi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\neg\varphi]\!]_{(r,m,i)})$. 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 that $\mathrm{Pl}_{(r,m,i)}(A_\psi \cap [\![\varphi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}(A_\psi \cap [\![\neg\varphi]\!]_{(r,m,i)})$. 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 $\mathrm{Pl}_{(r',m,i)}(A_\psi \cap [\![\varphi]\!]_{(r',m,i)}) > \mathrm{Pl}_{(r',m,i)}(A_\psi \cap [\![\neg\varphi]\!]_{(r',m,i)})$. 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 $\mathrm{Pl}_{(r',m+1,i)}([\![\varphi]\!]_{(r',m+1,i)}) > \mathrm{Pl}_{(r',m+1,i)}([\![\neg\varphi]\!]_{(r',m+1,i)})$. Thus, $(\mathcal{I}, r', m) \models \bigcirc B_i\varphi$. We conclude that $A_\psi \subseteq [\![\bigcirc B_i\varphi]\!]_{(r,m,i)}$, and thus $\mathrm{Pl}_{(r,m,i)}([\![\bigcirc B_i\varphi]\!]_{(r,m,i)}) = \top$. Moreover, since $A_\psi \subseteq [\![\bigcirc B_i\varphi]\!]_{(r,m,i)}$ for all A_ψ such that $\mathrm{Pl}_{(r,m,i)}(A_\psi) = \top$, we get that $\mathrm{Pl}_{(r,m,i)}([\![\neg\bigcirc B_i\varphi]\!]_{(r,m,i)}) \le \max\{\mathrm{Pl}_{(r,m,i)}(A_\psi) : \mathrm{Pl}_{(r,m,i)}(A_\psi) < \top\} < \top$. We conclude that $\mathrm{Pl}_{(r,m,i)}([\![\bigcirc B_i\varphi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\neg\bigcirc B_i\varphi]\!]_{(r,m,i)})$, and thus, $(\mathcal{I}, r, m) \models B_i\bigcirc B_i\varphi$.

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

C.4 Proofs for Section 4.7

Theorem 4.7.4: *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 $\varphi \in \mathcal{L}^{KCT}$, we have$(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', f(r), m) \models \varphi$.*

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 C.4.1: *(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 $\mathrm{Pl}_k(C \cap W_k) = \perp_k$ for all $k \leq j$. If $i < j$, we conclude that $\mathrm{Pl}_i(A \cap W_i) \geq \mathrm{Pl}_i(C \cap W_i) = \perp_i$, 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 $\mathrm{Pl}_i((A \cup B) \cap W_i) > \perp_i$ and $\mathrm{Pl}_j((A \cup C) \cap W_j) > \perp_j$. We claim that $i = j$. Assume, by way of contradiction, that $i < j$. Then, $\mathrm{Pl}_i((A \cup C) \cap W_i) = \perp_i$ and hence $\mathrm{Pl}_i(A \cap W_i) = \perp_i$. Moreover, since $\mathrm{Pl}_{\oplus S_i}(A \cup C) > \mathrm{Pl}_{\oplus S_i}(B)$, we get that $\mathrm{Pl}_i(B \cap W_i) = \perp_i$. Using A3 in S_i, we conclude that $\mathrm{Pl}_i((A \cup B) \cap W_i) = \perp_i$, which contradicts our assumption that $\mathrm{Pl}_i((A \cup B) \cap W_i) > \perp_i$. 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 $\mathrm{Pl}_j((B \cup C) \cap W_j) = \perp_j$ 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 $\mathrm{Pl}_{\oplus S_i}(A) = \perp$ if and only if $\mathrm{Pl}_1(A \cap W_i) = \perp$ 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. ∎

Returning to the proof of Theorem 4.7.4, 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 $\varphi \in \mathcal{L}^{KC}$ (so that φ does not include any temporal modalities) and $l \geq m$. We show that $(\mathcal{I}', r^l, m) \models \varphi$ if and only if $(\mathcal{I}, r, m) \models \varphi$. 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 $(\mathcal{I}, r, m) \models K_i\varphi$. Then for all points $(s, m) \in \mathcal{K}_i(r, m)$, we have $(\mathcal{I}, s, m) \models \varphi$. 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 $(\mathcal{I}', s^k, m) \models \varphi$. We conclude that $(\mathcal{I}', r^l, m) \models K_i\varphi$. Now suppose that $(\mathcal{I}, r, m) \not\models K_i\varphi$. Then there is a point

$(s, m) \in \mathcal{K}_i(r, m)$ such that $(\mathcal{I}, s, m) \models \neg\varphi$. Using the induction hypothesis we conclude that $(\mathcal{I}, s^m, m) \models \neg\varphi$. Since $(s^m, m) \in \mathcal{K}'_i(r^l, m)$, we conclude that $(\mathcal{I}', r^l, m) \not\models K_i\varphi$.

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 C.4.1 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 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 between 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 $(\mathcal{I}, r, m) \models \varphi \rightarrow_i \psi$. Thus, either $\mathrm{Pl}_{(r,m,i)}([\![\varphi]\!]_{(r,m,i)}) = \perp$ or $\mathrm{Pl}_{(r,m,i)}([\![\varphi \wedge \psi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\varphi \wedge \neg\psi]\!]_{(r,m,i)})$. If $\mathrm{Pl}_{(r,m,i)}([\![\varphi]\!]_{(r,m,i)}) = \perp$, then from the coherence of $\mathcal{I}$ it follows that if $A \subseteq W_{(r,l',i)}$ and $\mathcal{R}(A) \subseteq \mathcal{R}([\![\varphi]\!]_{(r,m,i)})$, then $\mathrm{Pl}_{(r,l',i)}(A) = \perp$. This implies that $\mathrm{Pl}^{l'}_{(r,i)}((\mathcal{R}([\![\varphi]\!]_{(r,m,i)})^{l'} \cap \mathcal{R}(W_{(r,l',i)})^{l'}) = \perp$ 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 $\mathrm{Pl}'_{(r^l,m,i)}([\![\varphi]\!]_{(r^l,m,i)}) = \perp$. Thus, we conclude that $(\mathcal{I}', r^l, m) \models \varphi \rightarrow_i \psi$ in this case. Now suppose that $\mathrm{Pl}_{(r,m,i)}([\![\varphi \wedge \psi]\!]_{(r,m,i)}) > \mathrm{Pl}_{(r,m,i)}([\![\varphi \wedge \neg\psi]\!]_{(r,m,i)})$. If we could show that $(\mathcal{R}([\![\varphi]\!]_{(r,m,i)}))^m = \mathcal{R}'([\![\varphi]\!]_{(r^l,m,i)}) \cap \mathcal{R}^m$, and similarly for ψ, then we could apply the argument of the previous paragraph to show that $\mathrm{Pl}'_{(r^l,m,i)}([\![\varphi \wedge \psi]\!]_{(r^l,m,i)}) > \mathrm{Pl}'_{(r^l,m,i)}([\![\varphi \wedge \neg\psi]\!]_{(r^l,m,i)})$. This, in turn, would allow us to conclude that $(\mathcal{I}', r^l, m) \models \varphi \rightarrow_i \psi$. The fact that $(\mathcal{R}([\![\varphi]\!]_{(r,m,i)}))^m = \mathcal{R}'([\![\varphi]\!]_{(r^l,m,i)}) \cap \mathcal{R}^m$ follows from the following chain of equivalences:

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

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

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

Finally, for $r \in \mathcal{R}$, define $f(r) = r^\infty$. We have proved that if $\varphi \in \mathcal{L}^{KC}$, then $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', f(r), m) \models \varphi$. 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 C.4.1, we conclude that the prior $\mathcal{P}'_{(r,i)}$ is qualitative for all agents i and runs $r \in R$. This implies, using Proposition 4.7.1, that $\mathcal{I}'$ satisfies QUAL. Similarly, if $\mathcal{I}$ satisfies RANK, using part (c) of Lemma C.4.1 and Proposition 4.7.1, we get that $\mathcal{I}'$ satisfies RANK.

Suppose that $\mathcal{I}$ satisfies NORM. Then $\mathrm{Pl}_{(r,m,i)}([\![\mathit{true}]\!] > \perp$ for all agents i, runs $r \in \mathcal{R}$, and times m. This implies that $\neg(\mathit{true} \rightarrow_i \mathit{false})$ is valid in $\mathcal{I}$. Suppose that $l \geq m$. Then since $\neg(\mathit{true} \rightarrow_i \mathit{false}) \in \mathcal{L}^{KC}$, we conclude from the proof above that $(\mathcal{I}', r^l, m) \models \neg(\mathit{true} \rightarrow_i \mathit{false})$. Thus, $\mathrm{Pl}'_{(r^l,m,i)}([\![\mathit{true}]\!]_{(r^l,m,i)}) > \perp$. 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 C.4.1, 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 4.7.1 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 $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', r^l, m) \models \varphi$ for all $l \geq m$. Again, we proceed by induction on the structure of formulas. The argument for formulas of the form $K_i\varphi$ goes through unchanged, since the changes to Pl' did not affect the $\mathcal{K}_i$ relations. The argument for formulas of the form $\varphi \rightarrow_i \psi$ goes through with almost no change. The only case that requires attention is if $(\mathcal{I}, r, m) \models \varphi \rightarrow_i \psi$ and $[\![\varphi]\!]_{(r,m,i)} = \perp$. Our earlier arguments showed that $\mathrm{Pl}^{l'}_{(r,i)}((\mathcal{R}([\![\varphi]\!]_{(r,m,i)})^{l'} \cap \mathcal{R}(W_{(r,l',i)}))^{l'}) = \perp$ 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 $\mathrm{Pl}^{\infty}_{(r,i)}((\mathcal{R}([\![\varphi]\!]_{(r,m,i)})^{\infty} \cap \mathcal{R}^{\infty}) = \perp$ unless $r^{\infty} \in \mathcal{R}([\![\varphi]\!]_{(r,m,i)})^{\infty}$. This implies that $(r,m) \in [\![\varphi]\!]_{(r,m,i)}$. But this cannot happen, since $\mathrm{Pl}_{(r,m,i)}([\![\varphi]\!]_{(r,m,i)}) = \perp$ and $\mathcal{I}$ satisfies REF. ∎

Theorem 4.7.6: *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 $\varphi \in \mathcal{L}^{KCT}$, $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', f(r), m) \models \varphi$.*

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 4.7.4 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 4.7.4.

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 C.4.2: *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 $S = S_i$.[1] Define $\otimes\mathcal{S} = \oplus_i S_i$. Part (a) of Lemma C.4.1 guarantees that $\otimes\mathcal{S}$ is a plausibility space. Parts (a), (b), and (c) follow immediately from parts (e), (b), and (c) of Lemma C.4.1, respectively. ∎

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

[1] 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.

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 C.4.3: *(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. ∎

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 4.7.4 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 C.4.3 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 C.4.3, 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 $\varphi \in \mathcal{L}^{KC}$ and $l \geq m$. As in the proof of Theorem 4.7.4, we now proceed by induction on the structure of formulas to show that $(\mathcal{I}', r^l, m) \models \varphi$ if and only if $(\mathcal{I}, r, m) \models \varphi$. The only difference arises in dealing with the $\rightarrow_i$ modality.

As before, parts (d) and (f) of Lemma C.4.1 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 4.7.4. Thus, even though we are using a different prior from that of the proof of Theorem 4.7.4, after conditioning, they are essentially the same. By Lemma C.4.2, 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 C.4.3, $(s,m) \in [r',m]_i$. By part (c) of Lemma C.4.3, 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 C.4.3, 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 4.7.4.

Finally, we need to ensure that $\mathcal{I}'$ satisfies $\mathcal{A}$. The proof of Theorem 4.7.4 shows that if $\mathcal{I}$ satisfies NORM, then so does $\mathcal{I}'$. Using parts (b) and (c) of Lemma C.4.2, 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 4.7.4. Define $\mathcal{P}^\infty_{[r]_i} = ([r]^\infty_i, \mathrm{Pl}^\infty_{[r]_i})$, where $\mathrm{Pl}^\infty_{[r]_i}(\emptyset) = \perp$ 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 4.7.4.

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. ∎

Appendix D

Proofs for Chapter 5

D.1 Proofs for Section 5.4

We start with the proof of Theorems 5.4.2 and 5.4.3. To do this, we need some preliminary definitions and lemmas. Figure D.1 shows the general outline of the intermediate representations we use in these proofs. Roughly speaking, we show how to map from a revision operator $\circ$ and a consistent belief set K to a ranking, and similarly how to map from a ranking to an AGM revision operator. These rankings correspond, in a direct way, to priors in systems in $\mathcal{C}^R$, and thus have close connection to the beliefs of the agent in various states.

These mapping between AGM revision operators and rankings are related to the representation theorems of Boutilier [1994c], Grove [1988], and Katsuno and Mendelzon [1991a]. However, the exact details of our representations are different than those of Boutilier, Grove, and Katsuno and Mendelzon. Thus, for completeness we provide the full proofs here.

We start with the mapping from revision operator applied to a specific belief set to a ranking. As an intermediate step we construct a set of defaults as follows. We then will use the results from Chapter 2 to construct a ranked plausibility structure that satisfies these defaults.

Lemma D.1.1: *Let $\circ$ be an AGM revision operator, let $K \subseteq \mathcal{L}_e$ be a consistent belief set, and let*

$$\Delta_{(\circ,K)} = \{\varphi \rightarrow \psi : \varphi, \psi \in \mathcal{L}_e,\ \psi \in K \circ \varphi\}.$$

Then the following is true:

(a) $\Delta_{(\circ,K)}$ *is closed under the rules of $system$* **P**,

(b) $\varphi \rightarrow$*false* $\notin \Delta_{(\circ,K)}$ *for all consistent $\varphi \in \mathcal{L}_e$, and*

(c) $\Delta_{(\circ,K)}$ *satisfies rational monotonicity; that is, if $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$ and $\varphi \rightarrow \neg\xi \notin \Delta_{(\circ,K)}$, then $\varphi \wedge \xi \rightarrow \psi \in \Delta_{(\circ,K)}$.*

Proof: We start with part (a):

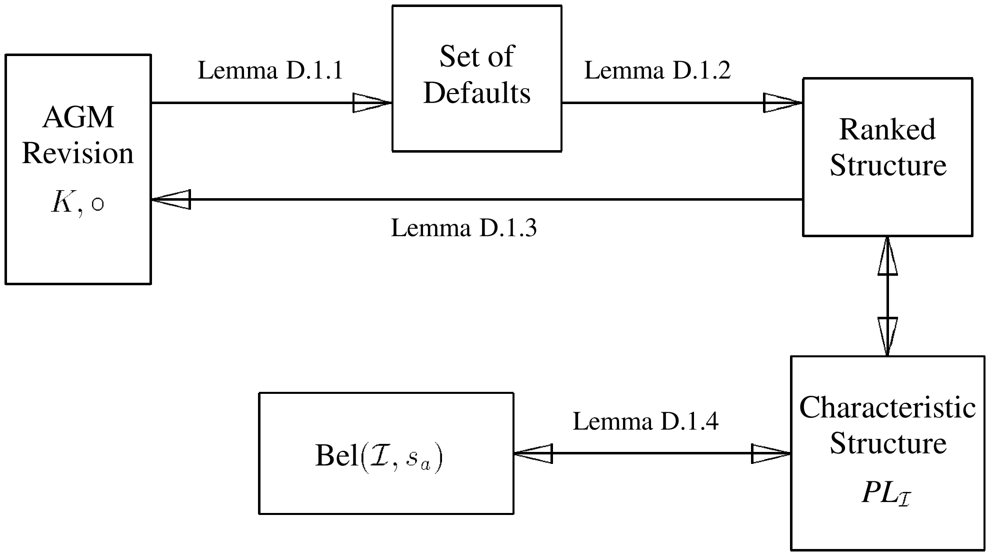

Figure D.1: Schematic description of the entities and lemmas involved in the proof of Theorems 5.4.2 and 5.4.3.

LLE Assume that $\vdash_{\mathcal{L}_e} \varphi \equiv \varphi'$ and that $\varphi{\rightarrow}\psi \in \Delta_{(\circ,K)}$. Thus, $\psi \in K \circ \varphi$. From R5, it follows that $\psi \in K \circ \varphi'$, and thus $\varphi'{\rightarrow}\psi \in \Delta_{(\circ,K)}$.

RW Assume that $\vdash_{\mathcal{L}_e} \psi \Rightarrow \psi'$ and that $\varphi{\rightarrow}\psi \in \Delta_{(\circ,K)}$. Thus, $\psi \in K \circ \varphi$. Since $K \circ \varphi$ is a belief set, it is closed under logical consequence. In particular, $\psi' \in K \circ \varphi$, and hence $\varphi{\rightarrow}\psi' \in \Delta_{(\circ,K)}$.

REF By R2, $\varphi \in K \circ \varphi$, and thus, $\varphi{\rightarrow}\varphi \in \Delta_{(\circ,K)}$.

AND Assume that $\varphi{\rightarrow}\psi_1, \varphi{\rightarrow}\psi_2 \in \Delta_{(\circ,K)}$. Thus, $\psi_1, \psi_2 \in K \circ \varphi$. Since $K \circ \varphi$ is a belief set, $\psi_1 \wedge \psi_2 \in K \circ \varphi$. Thus, $\varphi{\rightarrow}\psi_1 \wedge \psi_2 \in \Delta_{(\circ,K)}$.

OR Assume that $\varphi_1{\rightarrow}\psi, \varphi_2{\rightarrow}\psi \in \Delta_{(\circ,K)}$. There are two cases. If $K \circ (\varphi_1 \vee \varphi_2)$ is inconsistent, then $\psi \in K \circ (\varphi_1 \vee \varphi_2)$ and thus $\varphi_1 \vee \varphi_2{\rightarrow}\psi \in \Delta_{(\circ,K)}$. If $K \circ (\varphi_1 \vee \varphi_2)$ is consistent, then, by R2, $\varphi_1 \vee \varphi_2 \in K \circ (\varphi_1 \vee \varphi_2)$. Thus, we cannot have both $\neg\varphi_1$ and $\neg\varphi_2$ in $K \circ (\varphi_1 \vee \varphi_2)$. Without loss of generality, assume that $\neg\varphi_1 \notin K \circ (\varphi_1 \vee \varphi_2)$. Using R7 and R8, we get that $K \circ ((\varphi_1 \vee \varphi_2) \wedge \varphi_1) = Cl(K \circ (\varphi_1 \vee \varphi_2) \cup \{\varphi_1\})$. Using R6, we get that $K \circ ((\varphi_1 \vee \varphi_2) \wedge \varphi_1) = K \circ \varphi_1$. Thus, we conclude that $K \circ \varphi_1 = Cl(K \circ (\varphi_1 \vee \varphi_2) \cup \{\varphi_1\})$. Since $\varphi_1{\rightarrow}\psi \in \Delta_{(\circ,K)}$, we have that $\psi \in K \circ \varphi_1$. Thus, we get that $\varphi_1 \Rightarrow \psi \in K \circ (\varphi_1 \vee \varphi_2)$. If $\neg\varphi_2 \notin K \circ (\varphi_1 \vee \varphi_2)$, by similar arguments we get that $\varphi_2 \Rightarrow \psi \in K \circ (\varphi_1 \vee \varphi_2)$. This implies that $(\varphi_1 \vee \varphi_2) \Rightarrow \psi \in K \circ (\varphi_1 \vee \varphi_2)$, and thus $\psi \in K \circ (\varphi_1 \vee \varphi_2)$. On the other hand, if $\neg\varphi_2 \in K \circ (\varphi_1 \vee \varphi_2)$, then, since $\varphi_1 \vee \varphi_2 \in K \circ (\varphi_1 \vee \varphi_2)$, we get that $\varphi_1 \in K \circ (\varphi_1 \vee \varphi_2)$, and thus $\psi \in K \circ (\varphi_1 \vee \varphi_2)$.

CM Assume that $\varphi{\rightarrow}\psi_1, \varphi{\rightarrow}\psi_2 \in \Delta_{(\circ,K)}$. If $K \circ \varphi$ is inconsistent, then using R5 we get that φ is inconsistent. Thus, $\varphi \wedge \psi_1$ is inconsistent, so $\psi_2 \in K \circ (\varphi \wedge \psi_1)$. Now assume that $K \circ \varphi$

is consistent. Since $\varphi \rightarrow \psi_1$, we have that $\psi_1 \in K \circ \varphi$. Since $K \circ \varphi$ is consistent, we get that $\neg\psi_1 \notin K \circ \varphi$. Applying R8, we get that $K \circ \varphi \subseteq K \circ (\varphi \wedge \psi_1)$. Since $\varphi \rightarrow \psi_2 \in \Delta_{(\circ,K)}$, we have that $\psi_2 \in K \circ \varphi$. Thus, $\psi_2 \in K \circ (\varphi \wedge \psi_1)$. This implies that $(\varphi \wedge \psi_1) \rightarrow \psi_2 \in \Delta_{(\circ,K)}$.

We now prove part (b). Let $\varphi \in \mathcal{L}_e$ be a consistent formula. Then, using R5, we get that $K \circ \varphi$ is consistent. Thus, $\varphi \rightarrow \mathit{false} \notin \Delta_{(\circ,K)}$.

Finally we prove part (c). Assume that $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$, and $\varphi \wedge \xi \rightarrow \psi \notin \Delta_{(\circ,K)}$. Since $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$, we have that $\psi \in K \circ \varphi$. Now if $\neg\xi \notin K \circ \varphi$, then, using R8, we have that $Cl(K \circ \varphi \cup \{\xi\}) \subseteq K \circ (\varphi \wedge \xi)$. This implies that $\psi \in K \circ (\varphi \wedge \xi)$. However, since we assumed that $\varphi \wedge \xi \rightarrow \psi \notin \Delta_{(\circ,K)}$, we have that $\psi \notin K \circ (\varphi \wedge \xi)$; thus, we get a contradiction. We conclude that $\neg\xi \in K \circ \varphi$. Thus, $\varphi \rightarrow \neg\xi \in \Delta_{(\circ,K)}$. ∎

We now use this result to show that there exists a plausibility structure that corresponds to $\circ$ applied to K.

Lemma D.1.2: *Let $\circ$ be an AGM revision operator, and let $K \subseteq \mathcal{L}_e$ be a consistent belief set. Then there is a plausibility structure $PL = (W, \mathrm{Pl}, \pi)$ such that* Pl *is ranked, $PL \models \varphi \rightarrow \psi$ if and only if $\psi \in K \circ \varphi$, and* $\mathrm{Pl}([\![\varphi]\!]) > \bot$ *for all $\vdash_{\mathcal{L}_e}$-consistent formulas $\varphi \in \mathcal{L}_e$.*

Proof: We use the basic techniques described in the proof of Theorem 2.8.2. Let $\Delta_{(\circ,K)}$ be the set of defaults defined by Lemma D.1.1. We now construct a plausibility space $PL' = (W, \mathrm{Pl}', \pi)$ such that $PL' \models \varphi \rightarrow \psi$ if and only if $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$. We define PL' as follows:

- $W = \{w_V : V \subseteq \mathcal{L}_e$ is a maximal $\vdash_{\mathcal{L}_e}$-consistent set$\}$,
- $\pi(w_V)(p) =$ **true** if $p \in V$, and
- $\mathrm{Pl}'([\![\varphi]\!]) \geq \mathrm{Pl}'([\![\psi]\!])$ if and only if $(\varphi \vee \psi) \rightarrow \varphi \in \Delta_{(\circ,K)}$.

Using Lemma 2.4.1, we get that $PL' \models \varphi \rightarrow \psi$ if and only if $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$. From Lemma D.1.1 (c) and Proposition 2.8.4, it follows that that Pl' is rational. From Theorem 2.8.5, it follows that there is a ranked plausibility measure Pl such that (W, Pl) is default-isomorphic to (W, Pl'). Let $PL = (W, \mathrm{Pl}, \pi)$.

Since PL is default-isomorphic to PL', we have that $PL \models \varphi \rightarrow \psi$ if and only if $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$. Moreover, using Lemma D.1.1, we have that $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$ if and only if $\psi \in K \circ \varphi$. Thus, $PL \models \varphi \rightarrow \psi$ if and only if $\psi \in K \circ \varphi$. Finally, let φ be a $\vdash_{\mathcal{L}_e}$-consistent formula. From Lemma D.1.1 (b), we get that $\varphi \rightarrow \mathit{false} \notin \Delta_{(\circ,K)}$. Since $\Delta_{(\circ,K)}$ is closed under the rules of system **P**, we conclude that $(\varphi \vee \mathit{false}) \rightarrow \mathit{false} \notin \Delta_{(\circ,K)}$. Thus, $\mathrm{Pl}'([\![\varphi]\!]) \not\leq \bot = \mathrm{Pl}'([\![\mathit{false}]\!])$, and thus $\mathrm{Pl}'([\![\varphi]\!]) > \bot$. Since Pl is default-isomorphic to Pl', we conclude that $\mathrm{Pl}([\![\varphi]\!]) > \bot$. ∎

We now prove the converse to Lemma D.1.2.

Lemma D.1.3: *Let $PL = (W, \mathrm{Pl}, \pi)$ be a ranked plausibility structure such that $\pi(w)$ is $\vdash_{\mathcal{L}_e}$-consistent for all worlds w, and $PL \not\models \varphi \rightarrow \mathit{false}$ for all $\vdash_{\mathcal{L}_e}$-consistent formulas $\varphi \in \mathcal{L}_e$; let $K = \{\varphi \in \mathcal{L}_e : PL \models \mathit{true} \rightarrow \varphi\}$. Then there is an AGM revision operator $\circ$ such that $\psi \in K \circ \varphi$ if and only if $PL \models \varphi \rightarrow \psi$.*

Proof: Let $\circ$ be some belief change operation such that $K \circ \varphi = \{\psi : PL \models \varphi \rightarrow \psi\}$. Since this requirement constrains only the result of applying $\circ$ to K, we can assume without loss of generality that $\circ$ satisfies the AGM postulates when applied to belief sets other than K. Thus, we need prove only that $\circ$ satisfies the AGM postulates for revision applied to K. (Note that the proofs for R3 and R4 follow from the proofs for R7 and R8, respectively.)

R1 Since *PL* is qualitative, we have that $\{\psi : PL \models \varphi \rightarrow \psi\}$ is a belief set, that is, closed under logical consequences.

R2 Axiom C1 implies that $PL \models \varphi \rightarrow \varphi$. Thus, $\varphi \in K \circ \varphi$.

R5 By our assumptions, if φ is $\vdash_{\mathcal{L}_e}$-consistent, then $\mathrm{Pl}([\![\varphi]\!]) > \bot$, and thus $PL \not\models \varphi \rightarrow \mathit{false}$. On the other hand, if φ is not $\vdash_{\mathcal{L}_e}$-consistent, then $[\![\varphi]\!] = \emptyset$, and thus $\mathrm{Pl}([\![\varphi]\!]) = \bot$. We conclude that $\mathrm{Pl}([\![\varphi]\!]) = \bot$ if and only if $\vdash_{\mathcal{L}_e} \neg\varphi$. This implies that $PL \models \varphi \rightarrow \mathit{false}$ if and only if $\vdash_{\mathcal{L}_e} \neg\varphi$. Thus, $K \circ \varphi = Cl(\mathit{false})$ if and only if $\vdash_{\mathcal{L}_e} \neg\varphi$.

R6 Assume that $\vdash_{\mathcal{L}_e} \varphi \Leftrightarrow \varphi'$. Then, by our assumption, $\pi(w)(\varphi) = \pi(w)(\varphi')$. Thus, $[\![\varphi \wedge \psi]\!] = [\![\varphi' \wedge \psi]\!]$ for all formulas $\psi \in \mathcal{L}_e$. We conclude that $PL \models \varphi \rightarrow \psi$ if and only if $PL \models \varphi' \rightarrow \psi$. This implies that $K \circ \varphi = K \circ \varphi'$.

R7 There are two cases: either $\mathrm{Pl}([\![\varphi\wedge\psi]\!]) = \bot$ or $\mathrm{Pl}([\![\varphi\wedge\psi]\!]) > \bot$. If $\mathrm{Pl}([\![\varphi\wedge\psi]\!]) = \bot$, then $\varphi\wedge\psi$ is inconsistent. According to R2, we have that $\varphi \in K \circ \varphi$. Thus, $\varphi \wedge \psi \in Cl(K \circ \varphi \cup \{\psi\})$. This implies that $Cl(K \circ \varphi \cup \{\psi\})$ contains *false*, and thus $K \circ (\varphi \wedge \psi) \subseteq Cl(K \circ \varphi \cup \{\psi\})$. If $\mathrm{Pl}([\![\varphi\wedge\psi]\!]) > \bot$, let $\xi \in K \circ (\varphi\wedge\psi)$. We now show that $\xi \in Cl(K \circ \varphi \cup \{\psi\})$. This will show that $K \circ (\varphi \wedge \psi) \subseteq Cl(K \circ \varphi \cup \{\psi\})$. Since $\xi \in K \circ (\varphi \wedge \psi)$, we get that $PL \models (\varphi \wedge \psi) \rightarrow \xi$. Since $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) > \bot$, we get that $\mathrm{Pl}([\![\varphi \wedge \psi \wedge \xi]\!]) > \mathrm{Pl}([\![\varphi \wedge \psi \wedge \neg\xi]\!])$. Then we have that $\mathrm{Pl}([\![\varphi \wedge (\psi \Rightarrow \xi)]\!]) > \mathrm{Pl}([\![\varphi \wedge \neg(\psi \Rightarrow \xi))]\!])$, since $(\varphi \wedge \psi \wedge \xi) \Rightarrow (\varphi \wedge (\psi \Rightarrow \xi))$ and $(\varphi \wedge \neg(\psi \Rightarrow \xi)) \Rightarrow (\varphi \wedge \psi \wedge \neg\xi)$. This also implies that $\mathrm{Pl}([\![\varphi]\!]) > \bot$. Thus, $PL \models \varphi \rightarrow (\psi \Rightarrow \xi)$. So, $(\psi \Rightarrow \xi) \in K \circ \varphi$, and thus $\xi \in Cl(K \circ \varphi \cup \{\psi\})$.

R8 Assume that $\neg\psi \notin K \circ \varphi$. Let $\xi \in Cl(K \circ \varphi \cup \{\psi\})$. We now show that $\xi \in K \circ (\varphi \wedge \psi)$. This will show that $Cl(K \circ \varphi \cup \{\psi\}) \subseteq K \circ (\varphi \wedge \psi)$. Let $A = [\![\varphi \wedge \neg\psi]\!]$, $B = [\![\varphi \wedge \psi \wedge \xi]\!]$, and $C = [\![\varphi \wedge \psi \wedge \neg\xi]\!]$. It is easy to verify that these sets are pairwise disjoint. Since $\varphi \wedge (\psi \Rightarrow \xi) \equiv (\varphi \wedge \neg\psi) \vee (\varphi \wedge \psi \wedge \xi)$ and $(\varphi \wedge \neg(\psi \Rightarrow \xi)) \equiv (\varphi \wedge \psi \wedge \neg\xi)$, we conclude that $[\![\varphi \wedge (\psi \Rightarrow \xi)]\!] = A \cup B$, and $[\![\varphi \wedge \neg(\psi \Rightarrow \xi)]\!] = C$. Since $\xi \in Cl(K \circ \varphi \cup \{\psi\})$, we have that $(\psi \Rightarrow \xi) \in K \circ \varphi$. This means that $PL \models \varphi \rightarrow (\psi \Rightarrow \xi)$. Thus, either $\mathrm{Pl}([\![\varphi]\!]) = \bot$ or $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(C)$. If $\mathrm{Pl}([\![\varphi]\!]) = \bot$, then according to A1, we get that $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) = \bot$. Thus, $PL \models (\varphi \wedge \psi) \rightarrow \xi$ vacuously, and $\xi \in K \circ (\varphi \wedge \psi)$ as desired.

Now assume that $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(C)$. Since Pl is ranked, it satisfies A4′ and A5′. According to A5′, we get that either $\mathrm{Pl}(A) > \mathrm{Pl}(C)$ or $\mathrm{Pl}(B) > \mathrm{Pl}(C)$. Assume that $\mathrm{Pl}(A) > \mathrm{Pl}(C)$ and $\mathrm{Pl}(B) \not> \mathrm{Pl}(C)$. Then, using A4′, we get that $\mathrm{Pl}(A) > \mathrm{Pl}(B)$. Applying A2, we get that $\mathrm{Pl}(A) > \mathrm{Pl}(B \cup C)$. However since $A = [\![\varphi \wedge \neg\psi]\!]$ and $B \cup C = [\![\varphi \wedge \psi]\!]$, this implies that $\neg\psi \in K \circ \varphi$, which contradicts our assumption. Thus, we conclude that $\mathrm{Pl}(B) > \mathrm{Pl}(C)$. Since $B = [\![\varphi \wedge \psi \wedge \xi]\!]$ and $C = [\![\varphi \wedge \psi \wedge \neg\xi]\!]$, we get that $PL \models (\varphi \wedge \psi) \rightarrow \xi$, and thus $\xi \in K \circ (\varphi \wedge \psi)$.

R3 and **R4** Our definition of $\circ$ implies that $K \circ \mathit{true} = K$. According to R6, we have that $K \circ (\mathit{true} \wedge \varphi) = K \circ \varphi$. Combining these two facts, we get that R3 and R4 are special cases of R7 and R8, respectively.

∎

These results show how to map between ranked plausibility structures and AGM revision operators. We now relate systems in $\mathcal{C}^R$ and ranked plausibility structures. Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^R$. Recall that REV2 requires that the prior of $\mathcal{I}$ be a ranking. Thus, we can construct a ranked plausibility structure where worlds are runs in $\mathcal{R}$. We define the *characteristic structure* of $\mathcal{I}$ to be $PL_{\mathcal{I}} = (\mathcal{R}, \mathrm{Pl}_a, \pi_{\mathrm{Pl}_{\mathcal{I}}})$, where Pl_a is the agent's prior over runs and $\pi_{\mathrm{Pl}_{\mathcal{I}}}(r)(p) = \pi(r,0)(p)$ for all $p \in \Phi_e$. Note that $[\![\varphi]\!]_{PL_{\mathcal{I}}} = \mathcal{R}[\varphi]$.

We now use $PL_{\mathcal{I}}$ to describe the beliefs of the agent in each local state.

Lemma D.1.4: *Let $\mathcal{I} \in \mathcal{C}^R$ and let $s_a = \langle o_1, \ldots, o_m \rangle$. Then $\varphi \in Bel(\mathcal{I}, s_a)$ if and only if $PL_{\mathcal{I}} \models (\bigwedge_{i=1}^m o_i) \rightarrow \varphi$. (By convention, if $m = 0$, we take $(\bigwedge_{i=1}^m o_i)$ to be* **true***.)*

Proof: Let $\mathcal{I} \in \mathcal{C}^R$ and let $s_a = \langle o_1, \ldots, o_m \rangle$. There are two cases: either s_a is a local state in $\mathcal{I}$, or it is not.

If s_a is a local state in $\mathcal{I}$, suppose that $r_a(m) = s_a$. Note that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$ if and only if $\mathrm{Pl}_{(r,m)}([\![\varphi]\!]_{(r,m)}) > \mathrm{Pl}_{(r,m)}([\![\neg\varphi]\!]_{(r,m)})$. Recall that, according to the definition of conditioning, $\mathrm{Pl}_{(r,m)}(\cdot)$ is isomorphic to $\mathrm{Pl}_a(\cdot|\mathcal{R}[\cdot; o_1, \ldots, o_m])$. Thus, $\mathrm{Pl}_{(r,m)}([\![\varphi]\!]_{(r,m)}) > \mathrm{Pl}_{(r,m)}([\![\neg\varphi]\!]_{(r,m)})$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi] \mid \mathcal{R}[\cdot; o_1, \ldots, o_m]) > \mathrm{Pl}_a(\mathcal{R}[\neg\varphi] \mid \mathcal{R}[\cdot; o_1, \ldots, o_m])$. Using C1, this is true if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) > \mathrm{Pl}_a(\mathcal{R}[\neg\varphi; o_1, \ldots, o_m])$. Using REV4, this is true if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge \bigwedge_{i=1}^m o_i]) > \mathrm{Pl}_a(\mathcal{R}[\neg\varphi \wedge \bigwedge_{i=1}^m o_i])$. We get that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge \bigwedge_{i=1}^m o_i]) > \mathrm{Pl}_a(\mathcal{R}[\neg\varphi \wedge \bigwedge_{i=1}^m o_i])$. This implies that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$ if and only if $PL_{\mathcal{I}} \models (\bigwedge_{i=1}^m o_i) \rightarrow \varphi$.

If s_a is not a local state in $\mathcal{I}$, then $\mathcal{R}[\cdot; o_1, \ldots, o_m] = \emptyset$, and by definition $\mathrm{Pl}_a(\mathcal{R}[\cdot; o_1, \ldots, o_m]) = \perp$. Using C1 and REV4, we get that $PL_a(\mathcal{R}[\bigwedge_{i=1}^m o_i]) = \perp$, and thus $PL_{\mathcal{I}} \models (\bigwedge_{i=1}^m o_i) \rightarrow \varphi$ for all $\varphi \in \mathcal{L}_e$. Since s_a is not a local state in $\mathcal{I}$, by definition $\mathrm{Bel}(\mathcal{I}, s_a) = \mathcal{L}_e$. Hence, we can conclude that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$ if and only if $PL_{\mathcal{I}} \models (\bigwedge_{i=1}^m o_i) \rightarrow \varphi$. ∎

We now show that given a ranked plausibility structure *PL* we can construct a system whose characteristic structure is default-isomorphic to *PL*.

Lemma D.1.5: *Let $PL_K = (W_K, \mathrm{Pl}_K, \pi_K)$ be a plausibility space that satisfies the conditions of Lemma D.1.3. Then there is a system $\mathcal{I} \in \mathcal{C}^R$ such that $PL_{\mathcal{I}} = PL_K$.*

Proof: Let $PL_K = (W_K, \mathrm{Pl}_K, \pi_K)$ be a plausibility space that satisfies the conditions of Lemma D.1.3. For each world $w \in W_K$ and sequence of observations $o_1, o_2, \ldots$, let $r^{w,o_1,o_2,\cdots}$ be the run defined so that $r_e^{w,o_1,o_2,\cdots}(m) = w$ and $r_a^{w,o_1,o_2,\cdots}(m) = \langle o_1, \ldots, o_m \rangle$ for all m. Let $\mathcal{R} = \{r^{w,o_1,o_2,\cdots} : \pi_k(w)(o_i) = \mathbf{true} \text{ for all } i\}$. Define π so that $\pi(r,m)(p) = \pi_K(r_e(m))(p)$ for $p \in \Phi_e$, and so that

$\pi(r,m)(\mathit{learn}(\varphi)) = \mathbf{true}$ if $o_{(r,m)} = \varphi$ for $\varphi \in \mathcal{L}_e$. Finally, define the prior plausibility Pl_a so that $\mathrm{Pl}_a(R) = \mathrm{Pl}_K(\{w : \exists r \in R(w = r_e(0))\}$. It is easy to check that this definition implies that $\mathrm{Pl}_a(\mathcal{R}[\varphi]) = \mathrm{Pl}_K([\![\varphi]\!]_{PL_K})$. Thus, $PL_{\mathcal{I}} = PL_K$. Since Pl_K is a ranking, Pl_a is also a ranking and thus qualitative.

We now verify that the resulting interpreted system is indeed in $\mathcal{C}^R$. It is easy to check that $\mathcal{I}$ is a belief change system; that is, it satisfies BCS1–BCS6. The construction is such that $r_e(m) = r_e(0)$ for all runs r and times m. Thus, $\mathcal{I}$ satisfies REV1. Since the prior Pl_a is a ranking, this system also satisfies REV2. Lemma D.1.2 implies that if φ is a consistent formula, then $\mathrm{Pl}_K([\![\varphi]\!]_{PL_K}) > \perp$. This implies that $\mathrm{Pl}_a(\mathcal{R}[\varphi]) > \perp$, and thus the system satisfies REV3. Finally, it is easy to show that $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) = \mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o_1 \wedge \ldots \wedge o_m]) = \mathrm{Pl}_K([\![\varphi \wedge o_1 \wedge \ldots \wedge o_m]\!]_{PL_K})$. Thus, the system satisfies REV4. ∎

We are finally ready to prove Theorem 5.4.2.

Theorem 5.4.2: *Let $\circ$ be an AGM revision operator and let $K \subseteq \mathcal{L}_e$ be a consistent belief set. Then there is a system $\mathcal{I}_{(\circ,K)} \in \mathcal{C}^R$ such that $\mathit{Bel}(\mathcal{I}_{(\circ,K)}, \langle\rangle) = K$ and*

$$\mathit{Bel}(\mathcal{I}_{(\circ,K)}, \langle\rangle) \circ \varphi = \mathit{Bel}(\mathcal{I}_{(\circ,K)}, \langle\varphi\rangle)$$

for all $\varphi \in \mathcal{L}_e$.

Proof: Let $\circ$ be an AGM revision operator and let $K \subseteq \mathcal{L}_e$ be a consistent belief set. By Lemmas D.1.2 and D.1.5, there is a system $\mathcal{I}_{(\circ,K)} = (\mathcal{R}_{(\circ,K)}, \pi_{(\circ,K)}, \mathcal{P}_{(\circ,K)}) \in \mathcal{C}^R$ such that $PL_{\mathcal{I}_{(\circ,K)}} \models \varphi \rightarrow \psi$ if and only if $\psi \in K \circ \varphi$. Our construction is such that $\psi \in K \circ \varphi$ if and only if $PL_{\mathcal{I}_{(\circ,K)}} \models \varphi \rightarrow \psi$. Using Lemma D.1.4, we get that $PL_{\mathcal{I}_{(\circ,K)}} \models \varphi \rightarrow \psi$ if and only if $\psi \in \mathrm{Bel}(\mathcal{I}_{(\circ,K)}, \langle\varphi\rangle)$. Thus, $K \circ \varphi = \mathrm{Bel}(\mathcal{I}_{(\circ,K)}, \langle\varphi\rangle)$.

Finally, we show $\mathrm{Bel}(\mathcal{I}_{(\circ,K)}, \langle\rangle) = K$. We start by showing that $K \circ \mathit{true} = K$. Using R3, we get that $K \circ \mathit{true} \subseteq Cl(K \cup \{\mathit{true}\}) = K$. Since K is consistent, by R4, $Cl(K \cup \{\mathit{true}\}) \subseteq K \circ \mathit{true}$. Thus, $K \circ \mathit{true} = K$. By Lemma D.1.4, we have that $\mathrm{Bel}(\mathcal{I}, \langle\rangle) = \mathrm{Bel}(\mathcal{I}, \langle \mathit{true}\rangle)$. Since $\mathrm{Bel}(\mathcal{I}, \langle\mathit{true}\rangle) = K \circ \mathit{true}$, we conclude that $\mathrm{Bel}(\mathcal{I}_{(\circ,K)}, \langle\rangle) = K$. ∎

We next prove Theorem 5.4.3.

Theorem 5.4.3: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$. Then there is an AGM revision operator $\circ_{\mathcal{I}}$ such that*

$$\mathit{Bel}(\mathcal{I}, \langle\rangle) \circ_{\mathcal{I}} \varphi = \mathit{Bel}(\mathcal{I}, \langle\varphi\rangle)$$

for all $\varphi \in \mathcal{L}_e$.

Proof: Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ be a system in $\mathcal{C}^R$. It is easy to verify that $PL_{\mathcal{I}}$ satisfies the conditions of Lemma D.1.3 with $K = \mathrm{Bel}(\mathcal{I}, \langle\rangle)$. This lemma implies that there is a revision operator $\circ_{\mathcal{I}}$ such that $\psi \in K \circ_{\mathcal{I}} \varphi$ if and only if $PL_{\mathcal{I}} \models \varphi \rightarrow \psi$. Using Lemma D.1.4, we have that $\psi \in \mathrm{Bel}(\mathcal{I}, \langle\varphi\rangle)$ if and only if $\mathrm{Pl}_{\mathcal{I}} \models \varphi \rightarrow \psi$. Thus, we have that $K \circ_{\mathcal{I}} \varphi = \mathrm{Bel}(\mathcal{I}, \langle\varphi\rangle)$ for all formulas φ. ∎

Theorem 5.4.4: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ and $s_a = \langle o_1, \ldots, o_m \rangle$ be a local state in $\mathcal{I}$. Then there is an AGM revision operator $\circ_{\mathcal{I},s_a}$ such that*

$$\mathit{Bel}(\mathcal{I}, s_a) \circ_{\mathcal{I},s_a} \varphi = \mathit{Bel}(\mathcal{I}, s_a \cdot \varphi)$$

for all formulas $\varphi \in \mathcal{L}_e$ such that $o_1 \wedge \ldots o_m \wedge \varphi$ is consistent.

Proof: The structure of the proof is similar to that of Theorem 5.4.3. As in that proof, we construct a ranked plausibility structure and use Lemma D.1.3 to find an AGM revision operator. The main difference is that after observing $\varphi_1, \ldots, \varphi_k$, some events are considered impossible. Lemma D.1.3, however, requires that all possible formulas are assigned a positive plausibility. We overcome this problem by assigning a "fictional" positive plausibility to all non-empty events that are ruled out by the previous observations.

We proceed as follows. Let d_0 be a new plausibility value that is less plausible than all positive plausibilities in Pl_a; that is, if $\mathrm{Pl}_a(A) > \perp$, then $\mathrm{Pl}_a(A) > d_0$. Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^R$; let $s_a = \langle o_1, \ldots, o_m \rangle$. We define $\mathit{PL} = (\mathcal{R}, \mathrm{Pl}, \pi_{\mathit{PL}_\mathcal{I}})$, where Pl is such that $\mathrm{Pl}([\![\varphi]\!]) = \max(\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge \bigwedge_{i=1}^m o_i]), d_0)$ for all consistent formulas φ. This definition implies that if φ is consistent with $\bigwedge_{i=1}^m o_i$, then $\mathrm{Pl}([\![\varphi]\!]_{\mathit{PL}}) = \mathrm{Pl}_a(\mathcal{R}[\varphi_1 \wedge \bigwedge_{i=1}^m o_i])$.

We now prove that if φ is consistent with $\bigwedge_{i=1}^m o_i$, then $\mathit{PL} \models \varphi \rightarrow \psi$ if and only if $\mathit{PL}_\mathcal{I} \models (\varphi \wedge \bigwedge_{i=1}^m o_i) \rightarrow \psi$.

For the "if" part, assume that $\mathit{PL}_\mathcal{I} \models (\varphi \wedge \bigwedge_{i=1}^m o_i) \rightarrow \psi$. Since φ is consistent with $\bigwedge_{i=1}^m o_i$ it follows, from REV3, that $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge (\bigwedge_{i=1}^m o_i])) > \perp$. Thus, $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \psi]) > \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi]) \geq \perp$. Thus, $\varphi \wedge \psi$ is consistent with $\bigwedge_{i=1}^m o_i$. This implies that $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) = \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \psi] > \max(d_0, \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi]) = \mathrm{Pl}([\![\varphi \wedge \neg\psi]\!])$. We conclude that $\mathit{PL} \models \varphi \rightarrow \psi$.

For the "only if" part, assume that $\mathit{PL}_\mathcal{I} \not\models (\varphi \wedge (\bigwedge_{i=1}^m o_i)) \rightarrow \psi$. This implies that $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \psi]) \not> \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi])$. Since Pl_a is a ranking, it follows that $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i))\psi]) \leq \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi])$. Since $\perp < \mathrm{Pl}_a(\mathcal{R}[\varphi \wedge (\bigwedge_{i=1}^m o_i)]) = \max(\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \psi]), \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi]))$, we have that $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi]) > \perp$. We conclude that $\mathrm{Pl}([\![\varphi \wedge \neg\psi]\!]) \geq \mathrm{Pl}([\![\varphi \wedge \psi]\!])$. Thus, $\mathit{PL} \not\models \varphi \rightarrow \psi$.

It is easy to verify that *PL* is ranked, and satisfies the requirements of Lemma D.1.3. Thus, there exists a revision operator $\circ_{\mathcal{I},s_a}$ such that $\psi \in K \circ_{\mathcal{I},s_a} \varphi$ if and only if $\mathit{PL} \models \varphi \rightarrow \psi$, where $K = \{\varphi : \mathit{PL} \models \mathit{true} \rightarrow \varphi\}$. Moreover, since for all φ consistent with $\bigwedge_{i=1}^m o_i$ we have that$\mathit{PL} \models \varphi \rightarrow \psi$ if and only if $\mathit{PL}_\mathcal{I} \models (\varphi \wedge (\bigwedge_{i=1}^m o_i)) \rightarrow \psi$, then, from Lemma D.1.4, it follows that $K = \mathrm{Bel}(\mathcal{I}, s_a)$ and that if φ is consistent with $\bigwedge_{i=1}^m o_i$, then $\mathit{PL} \models \varphi \rightarrow \psi$ if and only if $\psi \in \mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi)$. ∎

Theorem 5.4.5: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ whose local states are $\mathcal{E}_{\mathcal{L}_e}$. There is a function $\mathit{Bel}_\mathcal{I}$ that maps epistemic states to belief states such that*

- *if s_a is a local state of the agent in $\mathcal{I}$, then $\mathit{Bel}(\mathcal{I}, s_a) = \mathit{Bel}_\mathcal{I}(s_a)$, and*

- $(\circ, \mathit{Bel}_{\mathcal{I}})$ *satisfies R1′–R8′.*

Proof: Roughly speaking, we define $\mathrm{Bel}_{\mathcal{I}}(s_a) = \mathrm{Bel}(\mathcal{I}, s_a)$ when s_a is a local state in $\mathcal{I}$. If s_a is not in $\mathcal{I}$, then we set $\mathrm{Bel}_{\mathcal{I}}(s_a) = \mathrm{Bel}(\mathcal{I}, s')$, where s' is the longest consistent suffix of s_a. We now make this definition precise, and show that the resulting $\mathrm{Bel}_{\mathcal{I}}$ satisfies R1′–R8′.

We proceed as follows. We define a function $f(\cdot)$ that maps sequences of observations to suffixes as follows:

$$f(\langle o_1, \ldots, o_m \rangle) = \begin{cases} \langle\rangle & \text{if } m = 0, \\ \langle \mathit{false} \rangle & \text{if } m > 0 \text{ and } o_m \text{ is inconsistent,} \\ \langle o_k, \ldots, o_m \rangle & \text{otherwise, with } k \leq m \text{ the minimal index s. t. } \not\vdash_{\mathcal{L}_e} \neg(o_k \wedge \ldots \wedge o_m). \end{cases}$$

Aside from the special case where o_m is inconsistent, we simply choose the longest suffix of s_a that is still consistent. We define $\mathrm{Bel}_{\mathcal{I}}(s_a) = \mathrm{Bel}(\mathcal{I}, f(s_a))$. Clearly, if s_a is a local state in $\mathcal{I}$, then $f(s_a) = s_a$, so $\mathrm{Bel}_{\mathcal{I}}(s_a) = \mathrm{Bel}(\mathcal{I}, s_a)$.

We now have to show that $(\circ, \mathrm{Bel}_{\mathcal{I}})$ satisfies R1′–R8′. The proof outline is as follows. Given a particular state s_a, we construct a ranked plausibility structure that corresponds, in the sense of Lemma D.1.2, to belief change from s_a. We then use Lemma D.1.3 to show that belief changes from s_a satisfies the AGM postulates, i.e., R1–R8. Since this is true from any s_a, we get that $\mathrm{Bel}_{\mathcal{I}}$ satisfies R1′–R8′.

Let $s_a = \langle o_1, \ldots, o_m \rangle$. We define a ranked plausibility space that has the following structure. The most plausible events are the ones consistent with $o_1 \wedge \ldots \wedge o_m$. They are ordered according to the prior ranking conditioned on $o_1 \wedge \ldots \wedge o_m$. The next tier of events are those that are inconsistent with $o_1 \wedge \ldots \wedge o_m$ but are consistent $o_2 \wedge \ldots \wedge o_m$. Again, these are ordered according to the prior ranking conditioned on $o_2 \wedge \ldots \wedge o_m$. We continue this way; the last tier consists of all events that are inconsistent with o_m.

Formally, let $\mathit{PL} = (\mathcal{R}, \mathrm{Pl}, \pi_{\mathit{PL}_{\mathcal{I}}})$, where Pl is such that $\mathrm{Pl}([\![\varphi]\!]) \geq \mathrm{Pl}([\![\psi]\!])$ if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge (\bigwedge_{i=k}^m o_i])) \geq \mathrm{Pl}_a(\mathcal{R}[\psi \wedge (\bigwedge_{i=k}^m o_i])) $ where $k \leq m+1$ is the greatest integer such that for all $j < k$, φ and ψ are both inconsistent with $\bigwedge_{i=j}^m o_i$. It is easy to see that *PL* is ranked, and that if φ is consistent, then $\mathrm{Pl}([\![\varphi]\!]) > \bot$.

Let $\varphi \in \mathcal{L}_e$. We now show that $\mathit{PL} \models \varphi \rightarrow \psi$ if and only if $\psi \in \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi)$. If φ is inconsistent, then $\mathit{PL} \models \varphi \rightarrow \psi$ for all ψ. Moreover, since φ is inconsistent, $f(s_a \cdot \varphi) = \langle \mathit{false} \rangle$, and thus $\mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi) = \mathcal{L}_e$. We conclude that $\varphi \rightarrow \psi$ if and only if $\psi \in \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi)$. If φ is consistent, then let $k \leq m+1$ be the greatest integer such that for all $j < k$, φ is inconsistent with $\bigwedge_{i=j}^m o_i$. It is easy to verify that $f(s_a \cdot \varphi) = \langle o_k, \ldots, o_m, \varphi \rangle$. From Lemma D.1.4, it follows that $\psi \in \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi) = \mathrm{Bel}(\mathcal{I}, \langle o_k, \ldots, o_m, \varphi \rangle)$ if and only if $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \psi]) > \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi])$. We now show that this is the case if and only if $\mathit{PL} \models \varphi \rightarrow \psi$. Suppose that $\mathit{PL}_a(\mathcal{R}[(\varphi \wedge \wedge(\bigwedge_{i=k}^m o_i)) \wedge \psi]) > \mathit{PL}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi])$. Then, clearly, $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \psi]) > \bot$, and thus $\varphi \wedge \psi$ is consistent with $o_k, \ldots, o_m$. Since both $\varphi \wedge \psi$ and $\varphi \wedge \neg\psi$ are inconsistent with $o_j, \ldots, o_m$ for all $j < k$, we have that $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) > \mathrm{Pl}([\![\varphi \wedge \neg\psi]\!])$. On other hand, if $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \psi]) \not> \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi])$, then since Pl_a is a ranking $\mathit{PL}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \psi]) \leq \mathit{PL}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi])$. Moreover, since

φ is consistent with $o_k \wedge \ldots \wedge o_m$, we have that $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge (\bigwedge_{i=k}^m o_i)]) > \perp$. This implies that $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi]) > \perp$ and thus $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) \leq \mathrm{Pl}([\![\varphi \wedge \neg\psi]\!])$. We conclude that $PL \models \varphi \rightarrow \psi$ if and only if $\psi \in \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi)$.

By Lemma D.1.3, there is a revision operator $\circ_{s_a}$ that satisfies R1–R8 such that $\psi \in K \circ \varphi$ if and only if $PL \models \varphi \rightarrow \psi$. It is not hard to check that this implies that the change from $\mathrm{Bel}_{\mathcal{I}}(s_a)$ to $\mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi)$ satisfies R1′–R8′. ∎

Proposition 5.4.6: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ whose local states are $\mathcal{E}_{\mathcal{L}_e}$. There is a function $Bel_{\mathcal{I}}$ that maps epistemic states to belief states such that*

- *if s_a is a local state of the agent in $\mathcal{I}$, then $Bel(\mathcal{I}, s_a) = Bel_{\mathcal{I}}(s_a)$, and*
- *$(\circ, Bel_{\mathcal{I}})$ satisfies R1′–R9′.*

Proof: We show that the function $\mathrm{Bel}_{\mathcal{I}}$ defined in the proof of Theorem 5.4.5 satisfies R9′. Let $s_a = \langle o_1, \ldots, o_m \rangle$, and let $\varphi, \psi \in \mathcal{L}_e$ be formulas such that $\not\vdash_{\mathcal{L}_e} \neg(\varphi \wedge \psi)$. Since φ is consistent with ψ, we get that $f(s_a \cdot \varphi \cdot \psi) = \langle o_k, \ldots, o_m, \varphi, \psi \rangle$, where $k \leq m$ is the least integer such that $\varphi \wedge \psi$ is consistent with $o_k, \ldots, o_m$. For the same reason, we get that $f(s_a \cdot \varphi \wedge \psi) = \langle o_k, \ldots, o_m, \varphi \wedge \psi \rangle$. Using Lemma D.1.4 we immediately get that $\mathrm{Bel}(\mathcal{I}, \langle o_k, \ldots, o_m, \varphi, \psi \rangle) = \mathrm{Bel}(\mathcal{I}, \langle o_k, \ldots, o_m, \varphi \wedge \psi \rangle)$. Thus, we conclude that $\mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi \cdot \psi) = \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi \wedge \psi)$. ∎

Theorem 5.4.7: *Given a function $Bel_{\mathcal{L}_e}$ mapping epistemic states in $\mathcal{E}_{\mathcal{L}_e}$ to belief sets over $\mathcal{L}_e$ such that $Bel_{\mathcal{L}_e}(\langle\rangle)$ is consistent and $(Bel_{\mathcal{L}_e}, \circ)$ satisfies R1′–R9′, there is a system $\mathcal{I} \in \mathcal{C}^R$ whose local states are in $\mathcal{E}_{\mathcal{L}_e}$ such that $Bel_{\mathcal{L}_e}(s_a) = Bel(\mathcal{I}, s_a)$ for each local state s_a in $\mathcal{I}$.*

Proof: According to Theorem 5.4.2, there is a system $\mathcal{I}$ such that $\mathrm{Bel}(\mathcal{I}, \langle\rangle) = \mathrm{Bel}_{\mathcal{L}_e}(\langle\rangle)$ and $\mathrm{Bel}(\mathcal{I}, \langle\varphi\rangle) = \mathrm{Bel}_{\mathcal{L}_e}(\langle\varphi\rangle)$ for all $\varphi \in \mathcal{L}_e$.

We now show that $\mathrm{Bel}(\mathcal{I}, s_a) = \mathrm{Bel}_{\mathcal{L}_e}(s_a)$ for local states s_a in $\mathcal{I}$. We prove this by induction on the length m of s_a. For $m \leq 1$, this is true by our choice of $\mathcal{I}$. For the induction case, let $s_a = \langle o_1, \ldots, o_m \rangle$ be a local state in $\mathcal{I}$. Thus,, $o_1 \wedge \ldots \wedge o_m$ is consistent. From R9′, it follows that $\mathrm{Bel}_{\mathcal{L}_e}(\langle o_1, \ldots, o_m \rangle) = \mathrm{Bel}_{\mathcal{L}_e}(\langle o_1, \ldots, o_{m-2}, o_{m-1} \wedge o_m \rangle)$. Using the induction hypothesis, we have that $\mathrm{Bel}_{\mathcal{L}_e}(\langle o_1, \ldots, o_{m-2}, o_{m-1} \wedge o_m \rangle) = \mathrm{Bel}(\mathcal{I}, \langle o_1, \ldots, o_{m-2}, o_{m-1} \wedge o_m \rangle)$. Using Lemma D.1.4, we get that $\mathrm{Bel}(\mathcal{I}, \langle o_1, \ldots, o_{m-2}, o_{m-1} \wedge o_m \rangle) = \mathrm{Bel}(\mathcal{I}, \langle o_1, \ldots, o_m \rangle)$. Thus, we conclude that $\mathrm{Bel}_{\mathcal{L}_e}(\langle o_1, \ldots, o_m \rangle) = \mathrm{Bel}(\mathcal{I}, \langle o_1, \ldots, o_m \rangle)$. ∎

D.2 Proofs for Section 5.5

In this section we prove Theorem 5.5.1. We start by reviewing the semantic representation of Katsuno and Mendelzon for update. Katsuno and Mendelzon show that there is a direct relation between update operators and distance functions. To make this relation precise, we need to introduce some definitions. An *update structure* is a tuple $U = (W, d, \pi)$, where W is a finite set of worlds, d is a distance function on W, and π is a mapping from worlds to truth assignments for $\mathcal{L}_e$ such that

- $\pi(w)$ is $\vdash_{\mathcal{L}_e}$ consistent,
- if $\not\vdash_{\mathcal{L}_e} \neg\varphi$, then there is some $w \in W$ with $\pi(w)(\varphi) = $ **true**, and
- if $w \neq w'$ then $\pi(w) \neq \pi(w')$ for all $w, w' \in W$.

Given an update structure $U = (W, d, \pi)$, we define $[\![\varphi]\!]_U = \{w : \pi(w)(\varphi) = \textbf{true}\}$. Katsuno and Mendelzon use update structures as semantic representations of update operators. Given an update structure $U = (W, d, \pi)$ and sets $A, B \subseteq W$, Katsuno and Mendelzon define $\min_U(A, B)$ to be the set of worlds in B that are closest to worlds in A, according to d. Formally, $\min_U(A, B) = \{w \in B : \exists w_0 \in A \forall w' \in B\, d(w_0, w') \not< d(w_0, w)\}$.

Theorem D.2.1: [Katsuno and Mendelzon 1991b] *A belief change operator $\diamond$ satisfies U1–U8 if and only if there is an update structure $U = (W, \pi, d)$ such that*

$$[\![\varphi \diamond \psi]\!]_U = \min_U([\![\varphi]\!]_U, [\![\psi]\!]_U).$$

Thus the worlds the agent believes possible after updating with ψ are these worlds that are closest to some world considered possible before learning ψ.

We now show that any system in $\mathcal{C}^U$ corresponds to an update structure. Suppose that $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^U$ is such that the set of environment states is $\mathcal{S}_e$ and the prior of BCS6 is consistent with distance function d. Define an update structure $U_{\mathcal{I}} = (S_e, \pi_e, d)$, where for $p \in \Phi_e$, $\pi_e(s_e)(p) = \pi((s_e, s_a))(p)$ for some choice of s_a. By BCS2, the choice of s_a does not matter. It is easy to see that UPD1 ensures that S_e and π_e satisfy the requirements of the definition of update structures. We want to show that belief change in $\mathcal{I}$ corresponds to belief change in $U_{\mathcal{I}}$ in the sense of Theorem D.2.1. Since Theorem D.2.1 states that any belief change operation defined by an update structure satisfies U1–U8, this will suffice to prove the "if" direction of Theorem 5.5.1. To prove the "only if" direction of Theorem 5.5.1, we show that that for any update structure U, there is a system $\mathcal{I} \in \mathcal{C}^U$ such that $U_{\mathcal{I}} = U$.

We start with preliminary definitions and lemmas for the "if" direction of Theorem 5.5.1. Let $s_a = \langle o_1, \ldots, o_m\rangle$. We define *States*$(\mathcal{I}, s_a) = \{s \in \mathcal{S}_e : s \models \xi$ for all $\xi \in \text{Bel}(\mathcal{I}, s_a)\}$. Clearly, if φ is such that $\text{Bel}(\mathcal{I}, s_a) = Cl(\varphi)$, then *States*$(\mathcal{I}, s_a) = [\![\varphi]\!]_{U_{\mathcal{I}}}$. To show that belief change in $\mathcal{I}$ corresponds to belief change in $U_{\mathcal{I}}$ we have to show that

$$\textit{States}(\mathcal{I}, s_a \cdot \psi) = \min_{U_{\mathcal{I}}}(\textit{States}(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}}).$$

This is proved in Lemma D.2.4. To prove this lemma, we need some preliminary lemmas.

Lemma D.2.2: *Let $\mathcal{I} \in \mathcal{C}^U$, and let $s_a = \langle o_1, \ldots, o_m\rangle$. Then $\varphi \in$ Bel$(\mathcal{I}, s_a)$ if and only if $(\mathcal{I}, r, 0) \models (\bigcirc o_1 \wedge \ldots \wedge \bigcirc^m o_m) \rightarrow \bigcirc^m \varphi$ for some run r in $\mathcal{R}$.*

Proof: The proof of this lemma is analogous to the proof of Lemma D.1.4, using UPD3 and UPD4 instead of REV3 and REV4. We do not repeat the argument here. ∎

We now provide an alternative characterization of *States*$(\mathcal{I}, s_a)$ in terms of the agent's prior on run-prefixes.

Lemma D.2.3: *Let $\mathcal{I} \in \mathcal{C}^U$ and let $s_a = \langle o_1, \ldots, o_m \rangle$. Then $s_m \in$ States$(\mathcal{I}, s_a)$ if and only if there is a sequence of states $[s_0, \ldots, s_m] \subseteq \mathcal{R}[$true$, o_1, \ldots, o_m]$ such that* $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - [s_0, \ldots, s_m])$.

Proof: For the "if" direction, assume that there is a sequence $s_0, \ldots, s_m$ such that $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_m]$, and $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - [s_0, \ldots, s_m])$. By way of contradiction, assume that $s_m \notin \mathit{States}(\mathcal{I}, m)$. Thus, there is a formula $\xi \in \mathrm{Bel}(\mathcal{I}, m)$ such that $s_m \models \neg\xi$. From Lemma D.2.2 it follows that since $\xi \in \mathrm{Bel}(\mathcal{I}, s_a)$, $(\mathcal{I}, r, 0) \models (\bigcirc o_1 \wedge \ldots \wedge \bigcirc^m o_m) \rightarrow \bigcirc^m \xi$ for some run r in $\mathcal{R}$. From the definition of conditioning it follows that $\mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \xi]) > \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \neg\xi])$. Since $s_m \models \neg\xi$, we get that $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \neg\xi]$ and that $\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \xi] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - [s_0, \ldots, s_m]$. From A1, it follows that $\mathrm{Pl}_a([s_0, \ldots, s_m]) < \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - [s_0, \ldots, s_m])$, which contradicts our starting assumption. We conclude that $s_m \in \mathit{States}(\mathcal{I}, s_a)$.

For the "only if" direction, assume that $s_m \in \mathit{States}(\mathcal{I}, a)$. Since $\mathcal{S}_e$ is finite and π_e assigns a different truth assignment to each state in $\mathcal{S}_e$, there is a formula $\xi \in \mathcal{L}_e$ that characterizes s_m; that is, $s \models \xi$ if and only if $s = s_m$. Since $s_m \in \mathit{States}(\mathcal{I}, s_a)$, we have that $\neg\xi \notin \mathrm{Bel}(\mathcal{I}, s_a)$. Using Lemma D.2.2, we get that $(\mathcal{I}, r, 0) \not\models (\bigcirc o_1 \wedge \ldots \wedge \bigcirc^m o_m) \rightarrow \bigcirc^m \neg\xi$ for all runs $r \in \mathcal{R}$. By BCS6, this is true if and only if $\mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m]) > \perp$ and $\mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \xi]) \not< \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \neg\xi])$. By UPD2, there is a sequence $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \xi]$ such that $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m])$ for all $[s'_0, \ldots, s'_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \neg\xi]$. Moreover, without loss of generality, we can assume that $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m])$ for all $[s'_0, \ldots, s'_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \xi]$, since there are only finitely many such sequences. Thus, by UPD2, $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - [s_0, \ldots, s_m])$. ∎

We can now prove that belief change in $\mathcal{I}$ corresponds to belief change in $U_{\mathcal{I}}$.

Lemma D.2.4: *Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^U$ Then*

$$\mathit{States}(\mathcal{I}, s_a \cdot \psi) = \min_{U_{\mathcal{I}}}(\mathit{States}(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}})$$

for all local states s_a and formulas $\psi \in \mathcal{L}_e$.

Proof: Let Pl_a be the prior in $\mathcal{I}$; assume that Pl_a consistent with a distance function d. Let $s_a = \langle o_1, \ldots, o_m \rangle$.

To show that $\min_{U_{\mathcal{I}}}(\mathit{States}(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}}) \subseteq \mathit{States}(\mathcal{I}, s_a \cdot \psi)$, suppose that $s \in \min_{U_{\mathcal{I}}}(\mathit{States}(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}})$. Thus, there is a state $s_m \in \mathit{States}(\mathcal{I}, s_a)$ such that $d(s_m, s') \not< d(s_m, s)$ for all states s' that satisfy ψ. We want to show that $s \in \mathit{States}(\mathcal{I}, s_a \cdot \psi)$. From Lemma D.2.3, it follows that, since $s_m \in \mathit{States}(\mathcal{I}, s_a)$, there is a sequence $s_0, \ldots, s_{m-1}$ such that $[s_0, \ldots, s_m] \in \mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m]$ and $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m] - [s_0, \ldots, s_m])$. We now show that $\mathrm{Pl}_a([s_0, \ldots, s_m, s]) \not< \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m, \psi] - [s_0, \ldots, s_m, s])$. By Lemma D.2.3, this suffices to show that

$s \in States(\mathcal{I}, s_a \cdot \psi)$. Suppose that $[s'_0, \dots, s'_{m+1}] \subseteq \mathcal{R}[true, o_1, \dots, o_m, \psi] - [s_0, \dots, s_m, s]$. If $[s_0, \dots, s_m] = [s'_0, \dots, s'_m]$, then we have that $d(s'_m, s'_{m+1}) \not< d(s_m, s)$. Since Pl_a is consistent with d, it follows that $\mathrm{Pl}_a([s_0, \dots, s_m, s]) \not< \mathrm{Pl}_a([s'_0, \dots, s'_m, s'_{m+1}])$. If $[s_0, \dots, s_m] \neq [s'_0, \dots, s'_m]$, then, since $\mathrm{Pl}_a([s_0, \dots, s_m]) \not< \mathrm{Pl}_a([s'_0, \dots, s'_m])$ and Pl_a is consistent with d, we have that $\mathrm{Pl}_a([s_0, \dots, s_m, s]) \not< \mathrm{Pl}_a([s'_0, \dots, s'_m, s'_{m+1}])$.

Since $\mathrm{Pl}_a([s_0, \dots, s_m, s]) \not< \mathrm{Pl}_a([s'_0, \dots, s'_m, s'_{m+1}])$ for all $[s'_0, \dots, s'_{m+1}] \subseteq \mathcal{R}[true, o_1, \dots, o_m, \psi] - [s_0, \dots, s_m, s]$ and Pl_a is prefix-defined, we have that $\mathrm{Pl}_a([s_0, \dots, s_m, s]) \not< \mathrm{Pl}_a(\mathcal{R}[true, o_1, \dots, o_m, \psi] - [s_0, \dots, s_m, s]$. By Lemma D.2.3, $s \in States(\mathcal{I}, s_a \cdot \psi)$, as desired.

To show that $States(\mathcal{I}, s_a \cdot \psi) \subseteq \min_{U_\mathcal{I}}(States(\mathcal{I}, s_a), [\![\psi]\!]_{U_\mathcal{I}})$, suppose that $s \in States(\mathcal{I}, s_a \cdot \psi)$. By Lemma D.2.3, there is a sequence $s_0, \dots, s_m$ such that $[s_0, \dots, s_m, s]) \subseteq \mathcal{R}[true, o_1, \dots, o_m, \psi]$ and $\mathrm{Pl}_a([s_0, \dots, s_m, s]) \not< \mathrm{Pl}_a(\mathcal{R}[true, o_1, \dots, o_m, \psi] - [s_0, \dots, s_m, s])$. We want to show that $s_m \in States(\mathcal{I}, s_a)$ and that $d(s_m, s') \not< d(s_m, s)$ for all s' that satisfy ψ. This suffices to prove that $s \in \min_{U_\mathcal{I}}(States(\mathcal{I}, s_a), [\![\psi]\!]_{U_\mathcal{I}})$.

To show that $s_m \in States(\mathcal{I}, s_a)$, by Lemma D.2.3, it suffices to show that $\mathrm{Pl}_a([s_0, \dots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[true, o_1, \dots, o_m] - [s_0, \dots, s_m])$. Let $s'_0, \dots, s'_m$ be a sequence such that $[s'_0, \dots, s'_m] \subseteq \mathcal{R}[true, o_1, \dots, o_m]$. By definition, $[s'_0, \dots, s'_m, s] \subseteq \mathcal{R}[true, o_1, \dots, o_m, \psi]$. Thus, from our choice of $s_0, \dots, s_m$, it follows that $\mathrm{Pl}_a([s_0, \dots, s_m, s]) \not< \mathrm{Pl}_a([s'_0, \dots, s'_m, s])$. Since Pl_a is consistent with d, it follows that $\mathrm{Pl}_a([s_0, \dots, s_m]) \not< \mathrm{Pl}_a([s'_0, \dots, s'_m])$. Thus, by Lemma D.2.3, $s_m \in States(\mathcal{I}, s_a)$. To see that $d(s_m, s') \not< d(s_m, s)$ for all s' that satisfy ψ, let $s' \neq s$ be such that $s' \models \psi$. Thus, $[s_0, \dots, s_m, s'] \subseteq [true, o_1, \dots, o_m, \psi]$. From our choice of $s_0, \dots, s_m$, it follows that $\mathrm{Pl}_a([s_0, \dots, s_m, s]) \not< \mathrm{Pl}_a([s_0, \dots, s_m, s'])$. Since Pl_a is consistent with d, it follows that $d(s_m, s') \not< d(s_m, s)$. We conclude that $s \in \min_{U_\mathcal{I}}(States(\mathcal{I}, s_a), [\![\psi]\!]_{U_\mathcal{I}})$. ∎

We now have the tools to prove the "if" direction of Theorem 5.5.1.

Lemma D.2.5: *If $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^U$, then there is a belief change operator $\diamond$ that satisfies U1–U8 such that*

$$Bel(\mathcal{I}, s_a) \diamond \psi = Bel(\mathcal{I}, s_a \cdot \psi)$$

for all local states s_a and formulas $\psi \in \mathcal{L}_e$.

Proof: Let $\mathcal{I} \in \mathcal{C}^U$. Using the arguments we presented above, it easy to check that $U_\mathcal{I}$ is an update structure. By Theorem D.2.1, there is a belief change operator $\diamond$ that satisfies U1–U8 such that $[\![\varphi \diamond \psi]\!]_{U_\mathcal{I}} = \min_{U_\mathcal{I}}([\![\varphi]\!]_{U_\mathcal{I}}, [\![\psi]\!]_{U_\mathcal{I}})$ for all $\varphi, \psi \in \mathcal{L}_e$. From Lemma D.2.4, it follows that $\mathrm{Bel}(\mathcal{I}, s_a) \diamond \psi = \mathrm{Bel}(\mathcal{I}, s_a \cdot \psi)$. ∎

We now prove the "only if" direction of Theorem 5.5.1. Suppose that $\diamond$ is a belief change operator that satisfies U1–U8. According to Theorem D.2.1, there is an update structure $U_\diamond$ that corresponds to $\diamond$. Thus, it suffices to show that there is a system $\mathcal{I}$ such that $U_\mathcal{I} = U_\diamond$.

Lemma D.2.6: *Let $U = (W, d, \pi_U)$ be an update structure. Then there is a system $\mathcal{I} \in \mathcal{C}^U$ such that $U_\mathcal{I} = U$.*

Proof: Given the sequences $w_0, w_1, \ldots \in W$ and $o_1, o_2, \ldots \in \mathcal{L}_e$, let $r^{w_0,w_1,\ldots;o_1,o_2,\cdots}$ be the run defined so that $r_e^{w_0,w_1,\ldots;o_1,o_2,\cdots}(m) = w_m$ and $r_a^{w_0,w_1,\ldots;o_1,o_2,\cdots}(m) = \langle o_1, \ldots, o_m \rangle$. Let $\mathcal{R} = \{r^{w_0,w_1,\ldots;o_1,o_2,\cdots} : \pi_U(w_m)(o_m) = \textbf{true} \text{ for all } m\}$. Define π such that $\pi(r,m)(p) = \pi_U(r_e(m))(p)$ for $p \in \Phi_e$ and $\pi(r,m)(\mathit{learn}(\varphi)) = \textbf{true}$ if $o_{(r,m)} = \varphi$ for $\varphi \in \mathcal{L}_e$.

It is clear that $(\mathcal{R}, \pi)$ satisfies BCS1–BCS5 and UPD1. Thus, all that remains to show is that there is a prior plausibility measure Pl_a that satisfies UPD2–UPD4. This will ensure that $(\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^U$.

We proceed as follows. We define a *preferential space* $(\mathcal{R}, \prec)$ where $r \prec r'$ if and only if there is some m such that $r_e(k) = r'_e(k)$ for all $0 \le k \le m$, $r_e(m+1) \ne r'_e(m+1)$, and $d(r_e(m), r_e(m+1)) < d(r'_e(m), r'_e(m+1))$. Recall that $r \prec r'$ denotes that r is preferred over r'. Thus, this ordering is consistent with the comparison of events of the form $[s_0, \ldots, s_n]$ according to UPD2.

Using the construction of Theorem 2.4.2, there is a plausibility space (R, Pl_a) such that $\mathrm{Pl}_a(A) \ge \mathrm{Pl}_a(B)$ if and only if for all $r \in B - A$, there is a run $r' \in A$ such that $r' \prec r$ and there is no $r'' \in B - A$ such that $r'' \prec r'$. By Theorem 2.5.5, Pl_a is a qualitative plausibility measure. We now show that it satisfies UPD2–UPD4.

We start with UPD2. To show that Pl_A is consistent with d, we need to show that $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$ if and only if there is some $m < n$ such that $s_k = s'_k$ for all $0 \le k \le m$, and $d(s_m, s_{m+1}) > d(s'_m, s'_{m+1})$. Suppose that $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$. Let r be some run in $[s_0, \ldots, '_n]$. Without loss of generality we can assume that $r_e(m) = r_e(n)$ for all $m > n$. Since $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$, there is a run $r' \in [s'_0, \ldots, s'_n]$ such that $r' \prec r$. By definition, this implies that there is an m such that $r_e(k) = r'_e(k)$ for all $0 \le k \le m$, and $d(r'_e(m), r'_e(m+1)) < d(r_e(m), r_e(m+1))$. We claim that $m < n$. For if $m \ge n$, then $r_e(m+1) = r_e(m)$ by construction, so $d(r_e(m), r_e(m+1)) = d(r_e(m), r_e(m)) \le d(r'_e(m), r'_e(m+1))$ and $r' \not\prec r$, a contradiction. Thus, $s_k = s'_k$ for all $0 \le k \le m$, $d(s'_m, s'_{m+1}) < d(s_m, s_{m+1})$.

For the converse, suppose that there is an $m < n$ such that $s_k = s'_k$ for all $0 \le k \le m$, and $d(s'_m, s'_{m+1}) < d(s_m, s_{m+1})$. Let r' be the run where $r'_e(k) = s'_k$ for $k \le n$, $r'_e(k) = s'_n$ for $k \ge n$, and $o_{(r',k)} = \mathit{true}$ for all k. It follows $r' \prec r$ for all runs $r' \in [s_0, \ldots, s_n]$. Thus, $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$.

To show that Pl_a is prefix-defined, we must show that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) \ge \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_n])$ if and only if for all $[s_0, \ldots, s_n] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_n] - \mathcal{R}[\varphi_0, \ldots, \varphi_n]$, there is some $[s'_0, \ldots, s'_n] \subseteq \mathcal{R}[\varphi_0, \ldots, \varphi_n]$ such that $\mathrm{Pl}_a([s'_0, \ldots, s'_n]) > \mathrm{Pl}_a([s_0, \ldots, s_n])$. Suppose that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) \ge \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_n])$. Let $[s_0, \ldots, s_n] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_n] - \mathcal{R}[\varphi_0, \ldots, \varphi_n]$. Let $r \in [s_0, \ldots, s_n]$ be a run such that $r_e(m) = r_e(n)$ for all $m \ge n$. Since $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) \ge \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_n])$ there is a run $r' \in \mathcal{R}[\varphi_0, \ldots, \varphi_n]$ such that $r' \prec r$. This implies that there is an m such that $r_e(k) = r'_e(k)$ for all $0 \le k \le m$, and $d(r'_e(m), r'_e(m+1)) < d(r_e(m), r_e(m+1))$. As before, we have that $m < n$, and thus $\mathrm{Pl}_a([r'_e(0), \ldots, r'_e(n)]) > \mathrm{Pl}_a([s_0, \ldots, s_n])$. Since $r' \in \mathcal{R}[\varphi_0, \ldots, \varphi_n]$, we also have that $[r'_e(0), \ldots, r'_e(n)] \subseteq \mathcal{R}[\varphi_0, \ldots, \varphi_n]$, as desired.

For the converse, assume that for all $[s_0, \ldots, s_n] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_n] - \mathcal{R}[\varphi_0, \ldots, \varphi_n]$ there is some $[s'_0, \ldots, s'_n] \subseteq \mathcal{R}[\varphi_0, \ldots, \varphi_n]$ such that $\mathrm{Pl}_a([s'_0, \ldots, s'_n]) > \mathrm{Pl}_a([s_0, \ldots, s_n])$. This implies that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) > \mathrm{Pl}_a([s_0, \ldots, s_n])$ for all for all $[s_0, \ldots, s_n] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_n] -$

$\mathcal{R}[\varphi_0, \ldots, \varphi_n]$. Since there are only finitely many sequences of states of length m, we can apply A2, and conclude that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) > \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_n] - \mathcal{R}[\varphi_0, \ldots, \varphi_n])$. Thus, $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) \geq \mathrm{Pl}_a((\mathcal{R}[\psi_0, \ldots, \psi_n]))$.

For UPD3, recall that the construction of Theorem 2.4.2 is such that $\mathrm{Pl}_a(R) > \perp$ for all non-empty $R \subseteq \mathcal{R}$. Since, by our construction, the set $\mathcal{R}[\varphi_0, \ldots, \varphi_n]$ is non-empty for all sequences $\varphi_0, \ldots, \varphi_n$ of consistent formulas, UPD3 must hold.

Finally, we consider UPD4. We have to show that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_{n+1}; o_1, \ldots, o_n]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_{n+1}; o_1, \ldots, o_n])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_n \wedge o_n, \varphi_{n+1}]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \psi_1 \wedge o_1, \ldots, \psi_n \wedge o_n, \psi_{n+1}])$. By construction, $\mathcal{R}[\varphi_0, \ldots, \varphi_{n+1}; o_1, \ldots, o_n] \subseteq \mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_n \wedge o_n, \varphi_{n+1}]$. On the other hand, for each run $r \in \mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_n \wedge o_n, \varphi_{n+1}]$ there is a run $r' \in \mathcal{R}[\varphi_0, \ldots, \varphi_{n+1}; o_1, \ldots, o_n]$ such that $r'_e(m) = r_e(m)$ for all m, and $o_{(r,m)} = o_m$ for $1 \leq m \leq n$. Since the preference ordering on runs is a function only of the environment states, it is clear that r and r' are compared in the same manner; that is for all r'', $r'' \prec r$ if and only if $r'' \prec r'$, and $r \prec r''$ if and only if $r' \prec r''$. Thus, we conclude that for the purposes of the preference ordering, both $\mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_n \wedge o_n, \varphi_{n+1}]$ and $\mathcal{R}[\varphi_0, \ldots, \varphi_{n+1}; o_1, \ldots, o_n]$ are compared in the same manner to other sets. It easy to see that this suffices to show that Pl_a satisfies UPD4. ∎

Finally, we can prove Theorem 5.5.1.

Theorem 5.5.1: *A belief change operator $\diamond$ satisfies U1–U8 if and only if there is a system $\mathcal{I} \in \mathcal{C}^U$ such that*

$$\mathit{Bel}(\mathcal{I}, s_a) \diamond \psi = \mathit{Bel}(\mathcal{I}, s_a \cdot \psi)$$

for all epistemic states s_a and formulas $\psi \in \mathcal{L}_e$.

Proof: The "if" direction follows from Lemma D.2.5. For the "only if" direction, assume that $\diamond$ satisfies U1–U8. By Theorem D.2.1, there is an update structure $U_\diamond$ such that $[\![\varphi \diamond \psi]\!]_{U_\mathcal{I}} = \min_{U_\mathcal{I}}([\![\varphi]\!]_{U_\mathcal{I}}, [\![\psi]\!]_{U_\mathcal{I}})$ for all $\varphi, \psi \in \mathcal{L}_e$. By Lemma D.2.6, there is a system $\mathcal{I} \in \mathcal{C}^U$ such that $U_\mathcal{I} = U_\diamond$. From Lemma D.2.4, it follows that $\mathrm{Bel}(\mathcal{I}, s_a) \diamond \psi = \mathrm{Bel}(\mathcal{I}, s_a \cdot \psi)$ for all local states s_a and formulas $\psi \in \mathcal{L}_e$. ∎

Appendix E

Proofs for Chapter 6

E.1 Proofs for Section 6.3

Proposition 6.3.1: *If $\mathcal{D}$ be an algebraic domain, then $\mathcal{D}^k$ is also an algebraic domain.*

Proof: Let $\mathcal{D} = (D, \bigoplus, \bigotimes)$ be an algebraic domain. We want to prove that $\mathcal{D}^k = (D^k, \bigoplus_k, \bigotimes_k)$ is also an algebraic domain.

We need to prove that $\leq_k$ is partial order, that D is pointed, that $\bigoplus$ is commutative, associative, monotonic and additive, that $\bigotimes$ is commutative, associative, strictly monotonic, bottom preserving, and invertible, and that $\bigotimes$ distributes over $\bigoplus$.

We first prove that $\leq_k$ is a partial order, that is reflexive, transitive, and anti-symmetric. Since $\leq$ is reflexive, it follows that $(x_1, \ldots, x_k) \leq_k (x_1, \ldots, x_k)$. Now assume that $(x_1, \ldots, x_k) \leq_k (y_1, \ldots, y_k)$ and $(y_1, \ldots, y_k) \leq_k (z_1, \ldots, z_k)$. Then, by definition, $x_i \leq y_i$ and $y_i \leq z_i$ for all i. Thus, since $\leq$ is transitive, $x_i \leq z_i$ for all i, and we conclude that $(x_1, \ldots, x_k) \leq_k (z_1, \ldots, z_k)$. Finally, assume that $(x_1, \ldots, x_k) <_k (y_1, \ldots, y_k)$. This implies that there is some i such that $x_i < y_i$. Since $\leq$ is anti-symmetric, we get that $y_i \not\leq x_i$ and thus $(y_1, \ldots, y_k) \not\leq_k (x_1, \ldots, x_k)$.

Since we have that $(\bot, \ldots, \bot) \leq_k (x_1, \ldots, x_k) \leq_k (\top, \ldots, \top)$ for all $(x_1, \ldots, x_k)$, D^k is pointed.

Next, we prove that $\bigoplus_k$ is commutative, monotonic, and additive. Since $\bigoplus_k$ is defined by a pointwise application of $\bigoplus$, which is commutative, we get that $\bigoplus_k$ is commutative. Now assume that $(x_1, \ldots, x_k) \leq_k (y_1, \ldots, y_k)$. Then, since $\bigoplus$ is monotonic, $x_i \bigoplus z_i \leq y_i \bigoplus z_i$ for all i. This implies that $(x_1, \ldots, x_k) \bigoplus_k (z_1, \ldots, z_k) \leq_k (y_1, \ldots, y_k) \bigoplus_k (z_1, \ldots, z_k)$. Finally, we show that $\bigoplus_k$ is additive. Let $(x_1, \ldots, x_k) \in D^k$. Since $\bigoplus$ is additive, we have that $x_i \bigoplus \top = \top$ and that $x_i \bigoplus \bot = x_i$ for all i. This implies that $(x_1, \ldots, x_k) \bigoplus_k (\top, \ldots, \top) = (\top, \ldots, \top)$ and $(x_1, \ldots, x_k) \bigoplus_k (\bot, \ldots, \bot) = (x_1, \ldots, x_k)$.

Next, we prove that $\bigotimes_k$ is commutative, associative, strictly monotonic, and multiplicative. Since $\bigotimes_k$ is defined by a pointwise application of $\bigotimes$, which is commutative and associative, we get that $\bigotimes_k$ is also commutative and associative. Now assume that $(x_1, \ldots, x_k) <_k (y_1, \ldots, y_k)$. This implies that $x_i \le y_i$ for all i, and that there is some j such that $x_j < y_j$. Then, we have that if $x_i = y_i$, then $x_i \bigotimes z_i = y_i \bigotimes z_i$ and if $x_i < y_i$, then since $\bigotimes$ is strictly monotonic, $x_i \bigotimes z_i < y_i \bigotimes z_i$. This implies that $(x_1, \ldots, x_k) \bigotimes_k (z_1, \ldots, z_k) <_k (y_1, \ldots, y_k) \bigotimes_k (z_1, \ldots, z_k)$. Finally, we show that $\bigotimes_k$ is multiplicative. Let $(x_1, \ldots, x_k) \in D^k$. Since $\bigotimes$ is multiplicative, we have that $x_i \bigotimes \bot = \bot$ and $x_i \bigotimes \top = x_i$ for all i. Thus, $(x_1, \ldots, x_k) \bigotimes_k (\bot, \ldots, \bot) = (\bot, \ldots, \bot)$. and

Finally, we need to show that $\bigotimes_k$ distributes over $\bigoplus_k$ and that $\bigotimes_k$ has a unique inverse when $x \ge_k y$. Both properties follow in a straightforward manner from the pointwise definition of $\bigotimes_k$, $\bigoplus_k$ and $\le_k$. ∎

We now prove Theorem 6.3.2. We start with a lemma.

Lemma E.1.1: *Let $(W, \mathcal{F}, \mathrm{Pl})$ be a normal qualitative plausibility space such that $\mathcal{F}$ is finite, and let $A^*, B^* \in \mathcal{F}$ be disjoint sets such that $\mathrm{Pl}(A^*) \not< \mathrm{Pl}(B^*)$. Then there is a plausibility measure Pl' over W based on $\mathcal{D}_\kappa$ such that $\mathrm{Pl}'(A^*) > \mathrm{Pl}'(B^*)$; moreover, if $\mathrm{Pl}(A) < \mathrm{Pl}(B)$ then $\mathrm{Pl}'(A) < \mathrm{Pl}'(B)$, for all disjoint sets $A, B \in \mathcal{F}$.*

Proof: Same as the proof of Lemma A.3.1. We leave the details to the reader. ∎

Theorem 6.3.2: *Let $(W, \mathcal{F}, \mathrm{Pl})$ be a qualitative plausibility space such that $\mathcal{F}$ is finite. Then there is an integer k and a structured plausibility space $(W, \mathcal{F}, \mathrm{Pl}')$ based on $\mathcal{D}_\kappa^k$ such that Pl' is default-isomorphic to Pl.*

Proof: Let $\langle A_1, B_1\rangle, \ldots, \langle A_k, B_k\rangle$ be some enumeration of all pairs of disjoint sets $A_i, B_i \in \mathcal{F}$ such that $\mathrm{Pl}(A_i) \not< \mathrm{Pl}(B_i)$. By Lemma E.1.1, for each pair A_i, B_i there exists a plausibility measure Pl_i over W based on $\mathcal{D}_\kappa$ such that $\mathrm{Pl}_i(A_i) > \mathrm{Pl}_i(B_i)$. Moreover, if A and B are disjoint sets such that $\mathrm{Pl}(A) < \mathrm{Pl}(B)$, then $\mathrm{Pl}_i(A) < \mathrm{Pl}_i(B)$ for all i.

For $A \in \mathcal{F}$, we define $\mathrm{Pl}'(A) = \langle \mathrm{Pl}_1(A), \ldots, \mathrm{Pl}_k(A)\rangle$. It is easy to verify that Pl' is based on $\mathcal{D}_\kappa^k$, using the pointwise definition of $\bigoplus_k$. Let A, B be two disjoint sets in $\mathcal{F}$. We claim that $\mathrm{Pl}(A) < \mathrm{Pl}(B)$ if and only if $\mathrm{Pl}'(A) < \mathrm{Pl}'(B)$. Assume that $\mathrm{Pl}(A) < \mathrm{Pl}(B)$. Then according to our construction, $\mathrm{Pl}_i(A) < \mathrm{Pl}_i(B)$ for all i. Thus, $\mathrm{Pl}'(A) < Pl'(B)$. Now assume that $\mathrm{Pl}(A) \not< \mathrm{Pl}(B)$. Then there is an index i such that $A_i = A$ and $B_i = B$. By construction, $\mathrm{Pl}_i(A) > \mathrm{Pl}_i(B)$. We conclude that $\mathrm{Pl}'(A) \not< \mathrm{Pl}'(B)$. ∎

Proposition 6.3.5: *Let $(W, \mathcal{F}, \mathrm{Pl})$ be a structured qualitative plausibility measure based on an algebraic domain $(D, \bigoplus, \bigotimes)$, and let $d, d' \in D$ be such that $d \ge d'$. If there are pairwise disjoint sets $A, B, C \in \mathcal{F}$ such that $\mathrm{Pl}(A) = \mathrm{Pl}(B) = d$ and $\mathrm{Pl}(C) = d'$, then $d \bigoplus d' = d$.*

Proof: Assume that $d \bigoplus d' \ne d$. Then, since $\bigoplus$ is monotonically non-decreasing, we have that $d \bigoplus d' > d$. Moreover, since $d \ge d'$, we also have that $d \bigoplus d > d'$. This implies that $\mathrm{Pl}(A \cup C) = d \bigoplus d' > d = \mathrm{Pl}(B)$ and that $\mathrm{Pl}(A \cup B) = d \bigoplus d > d' = \mathrm{Pl}(C)$. Applying A2, we get that $\mathrm{Pl}(A) > \mathrm{Pl}(B \cup C)$, Since $\mathrm{Pl}(B \cup C) \ge \mathrm{Pl}(B)$, we get that $\mathrm{Pl}(A) > \mathrm{Pl}(B)$, which contradicts our assumptions. ∎

Lemma 6.3.6: *If* $(W, \mathcal{F}, \mathrm{Pl})$ *is a qualitative decomposable plausibility space, and* $A, B, C, D \in \mathcal{F}$ *are pairwise disjoint sets such that* $\mathrm{Pl}(A)$ *is incomparable to* $\mathrm{Pl}(B)$ *(i.e.,* $\mathrm{Pl}(A) \not\leq \mathrm{Pl}(B)$ *and* $\mathrm{Pl}(B) \not\leq \mathrm{Pl}(A)$*) and* $\mathrm{Pl}(A) = \mathrm{Pl}(C)$*, then* $\mathrm{Pl}(B) \neq \mathrm{Pl}(D)$*.*

Proof: Assume that $\mathrm{Pl}(A) = \mathrm{Pl}(C)$ and $\mathrm{Pl}(B) = \mathrm{Pl}(D)$. Then, since $\mathrm{Pl}(A)$ and $\mathrm{Pl}(B)$ are incomparable, we have that $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(A)$ and $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(B)$. Using our assumption that $\mathrm{Pl}(A) = \mathrm{Pl}(C)$ and $\mathrm{Pl}(B) = \mathrm{Pl}(D)$, we get that $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(C)$ and $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(D)$. Applying A2, we get that $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(C \cup D)$. On the other hand, since Pl is decomposable and $\mathrm{Pl}(A) = \mathrm{Pl}(C)$ and $\mathrm{Pl}(B) = \mathrm{Pl}(D)$, we have that $PL(A \cup B) = \mathrm{Pl}(C \cup D)$ ∎

Theorem 6.3.7: *Under the conditions of Theorem 6.3.3, the unique Markovian plausibility measure of Theorem 6.3.3 is* not *qualitative if there is a state* $s \in \mathcal{S}$ *such that*

- *s is reachable from $s_\vdash$, and*
- *there are two sequences $s_1, \ldots, s_k$ and $s'_1, \ldots, s'_{k'}$ such that $s_1 = s'_1 = s_k = s'_{k'} = s$, neither sequence is a subsequence of the other, and $\bigotimes_{i=1}^{k} t_{s_i, s_{i+1}}$ is incomparable to $\bigotimes_{i=1}^{k'} t_{s'_i, s'_{i+1}}$.*

Proof: Let $\mathcal{D} = (D, \bigoplus, \bigotimes)$ be an algebraic domain, let $\mathcal{S}$ be a finite set of system states, let $\{t_{s,s'} : s \in \mathcal{S}^*, s' \in S\}$ be a set of proper transition plausibilities, and let $(W, \mathcal{F}, \mathrm{Pl})$ be the structured Markovian plausibility space of Theorem 6.3.3.

Suppose that there is a state s and three sequences $\sigma_j = \langle s^j_1, \ldots, s^j_{m_j} \rangle$ for $j = 0, 1, 2$ such that

- $s^j_{m_j} = s$ for $j = 0, 1, 2$,
- $d_0 = (t_{s_\vdash, s^0_1} \bigotimes (\bigotimes_{i=1}^{m_0 - 1} t_{s^0_i, s^0_{i+1}})) > \bot$,
- $d_j = (t_{s, s^j_0} \bigotimes (\bigotimes_{i=0}^{m_j - 1} t_{s^j_i, s^j_{i+1}})) > \bot$ for $j = 1, 2$,
- d_1 is incomparable to d_2, and
- neither σ_1 nor σ_2 is a substring of the other.

Let

$$\begin{aligned} A &= [\sigma_0 \cdot \sigma_1 \cdot \sigma_1 \cdot \sigma_2] \\ B &= [\sigma_0 \cdot \sigma_1 \cdot \sigma_2 \cdot \sigma_2] \\ C &= [\sigma_0 \cdot \sigma_1 \cdot \sigma_2 \cdot \sigma_2] \\ D &= [\sigma_0 \cdot \sigma_2 \cdot \sigma_1 \cdot \sigma_2]. \end{aligned}$$

Let $d = d_0 \bigotimes d_1 \bigotimes d_2$. It is easy to see that $\mathrm{Pl}(A) = \mathrm{Pl}(C) = d \bigotimes d_1$, and that $\mathrm{Pl}(B) = \mathrm{Pl}(D) = d \bigotimes d_2$. Since neither σ_1 nor σ_2 is a substring of the other, the sets A, B, C, and D are pairwise disjoint. Moreover, since d_1 is incomparable to d_2 and $d > \bot$, we have that $d \bigotimes d_1$ is incomparable to $d \bigotimes d_2$. Using Lemma 6.3.6, we conclude that Pl is not qualitative. ∎

E.2 Proofs for Section 6.4

Proposition 6.4.1: *Let $\mathcal{S}_e$ be a finite set of environment states, let π_e be a mapping from states in $\mathcal{S}_e$ to truth assignments to Φ_e, let $\mathcal{D} = (D, \bigoplus, \bigotimes)$ be a ranked algebraic domain, and let $\{t_{s,s'} \in D : s \in \mathcal{S}_e^*, s' \in \mathcal{S}_e\}$ be a set of proper transition plausibilities. Then there exists a BCS $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^M$ with a Markovian prior* Pl_a *based on $\mathcal{D}$ such that* $\mathrm{Pl}_a(S_0 = s) = t_{s_\vdash, s}$ *and* $\mathrm{Pl}_a(S_{m+1} = s' \mid S_m = s) = t_{s,s'}$ *for all states s and s' and all times m.*

Proof: Let $(\mathcal{R}, \pi)$ be the system that contains all runs of the form required by BCS1–BCS5 based on $\mathcal{S}_e$ and π_e.

We now define the prior plausibility Pl_a on $\mathcal{R}$. We use a construction similar that of Theorem 6.3.3. For each pair of sequences $s_0, \ldots, s_m \in \mathcal{S}_e$ and $o_1, \ldots, o_m \in \mathcal{L}_e$ such that $s_i \models o_i$ for all i, we define $\mathrm{Pl}([s_0, \ldots, s_m; o_1, \ldots, o_m]) = t_{s_0} \bigotimes t_{s_0,s_1} \bigotimes \cdots \bigotimes t_{s_{m-1},s_m}$. Let R be a time-m event. We define

$$\mathrm{Pl}_a(R) = \max\{\mathrm{Pl}_a([s_0, \ldots, s_m; o_1, \ldots, o_m]) : [s_0, \ldots, s_m; o_1, \ldots, o_m] \subseteq R\}.$$

This uniquely defines unconditional plausibilities in Pl_a. We then define conditional plausibilities using the same technique as in the proof of Theorem 6.3.3.

It is easy to verify that the resulting prior is Markovian, has the required transition plausibilities, and satisfies UPD4. ∎

E.3 Proofs for Section 6.6

We now describe a sound and complete characterization of safe sets of constraints. We start with few preliminary definitions. Let C be a set of constraints. We define C^+ to consist of C combined with the following additional constraints:

- $\bot \leq y$ for all variables y,
- $y \leq \top$ for all variables y, and
- $\top \leq x_{s,s'}$ if for all $s'' \neq s'$ there is a y such that $(x_{s,s''} < y) \in C$.

The first two items describe constraints that follow from the definition of $\bot$ and $\top$. The last item needs some clarification. Recall that we require that for a proper plausibility assignment a, $\bigoplus_{s'} a(x_{s,s'}) = \top$ for all s. Since we examine ranked domains, we have that $\bigoplus$ is max. Thus, this implies that for all s, there is an s' such that $a(x_{s,s'}) = \top$. If s' is the only state such that $x_{s,s'}$ is not strictly dominated by another variable, then $x_{s,s'}$ must be assigned $\top$.

We define relations $\leq_C$ and $<_C$ that capture the transitive closure of dominance and strict dominance relations in C. Formally, we say that

- $x_1 \leq_C x_n$ if there is a sequence $x_1, \ldots, x_n$ such that either $x_i \leq x_{i+1}$ or $x_i < x_{i+1}$ is in C^+ for all $1 \leq i < n$, and
- $x_1 <_C x_n$ if there is a sequence $x_1, \ldots, x_n$ such that $x_i < x_{i+1}$ is in C^+ for some i, and such that either $x_i \leq x_{i+1}$ or $x_i < x_{i+1}$ is in C^+ for all $1 \leq i < n$.

Thus, if $x \leq_C y$ (resp., $x <_C y$), then the constraints in C force y to be as plausible as (resp., strictly more plausible than) x. The discussion above shows that if a is a proper plausibility assignment that is consistent with C, then a is also consistent with $\leq_C$ and $<_C$:

Lemma E.3.1: *Let C be a set of constraints, and let a be a proper plausibility assignment to some ranked domain $(D, \bigoplus, \bigotimes)$ consistent with C. Then if $y \leq_C y'$ (resp., $y <_C y'$), then $a(y) \leq_D a(y')$ (resp., $a(y) <_D a(y')$).*

Using $<_C$, we can characterize when a set of constraints is safe.

Proposition E.3.2: *Let $\mathcal{S}_e$ be a finite set of states and C a set of constraints on the transition plausibilities over S_e. C is safe if and only if the following conditions are met:*

- *$y \not<_C y$ and $y \not<_C \perp$ for all variables y, and*
- *for all $s \in \mathcal{S}_e^*$ there exists some $s' \in \mathcal{S}_e$ such that $x_{s,s'} \not<_C \top$.*

Proof: For the "if" direction, assume that C satisfies the conditions. To show that C is safe, we construct an appropriate plausibility assignment. Let V be the set of transition variables. Let

$$V_i = \{y \in V - \cup_{0 \leq j < i} V_i : y \not\leq_C \perp, \forall y' \in (V - \cup_{0 \leq j < i} V_i), y \not<_C y'\}$$

for $i \geq 0$, and $V_\infty = \{y \in V : y \leq_C \perp\}$. Roughly speaking, V_i is the set of "largest" (that is, undominated) variables in $V - (V_0 \cup \ldots \cup V_{i-1})$. Since $\mathcal{S}_e$ is finite, so is V. Since for all y we have that $y \not<_C y$, we have that $V = \cup_i V_i \cup V_\infty$.

We now define an assignment $a : V \mapsto D_\kappa$. For $i = 0, 1, \ldots, \infty$ and $y \in V_i$, let $a(y) = i$. It is clear that a satisfies the constraints in C: if $(y < y') \in C$, then $y <_C y'$, so y must be assigned a higher κ-value than y' (recall that in κ-ranking, higher values are less plausible); if $(y \leq y') \in C$, then $y' \not<_C y$, since otherwise we would have that $y <_C y$. Thus, y cannot be assigned a lower κ-value than y'.

Finally, we have to verify that $\bigoplus_{s'} a(x_{s,s'}) = \top$ for all s. Let $s \in \mathcal{S}_e^*$. By assumption, there exists some $s' \in \mathcal{S}_e$ such that $x_{s,s'} \not<_C \top$. Thus, there cannot be a variable y such that $x_{s,s'} <_C y$. It follows that $x_{s,s'} \in V_0$, and thus $a(x_{s,s'}) = 0$. But $\top_\kappa = 0$, so $\bigoplus_{s'} a(x_{s,s'}) = \top$.

For "only if" direction, assume that C is safe. Thus, there is a plausibility assignment a that maps variables in V to some domain $\mathcal{D} = (D, \bigoplus, \bigotimes)$ such that a is consistent with C and $\bigoplus_{s'} a(x_{s,s'}) = \top$ for all s. Since $\leq_D$ is a partial order, we have that $a(y) \not<_D a(y)$. From Lemma E.3.1, it follows that $y \not<_C y$. Similarly, since D is pointed, we have that $a(y) \not<_D \perp$. From Lemma E.3.1, again it follows that $y \not<_C \perp$. We conclude that C satisfies the first requirement of the proposition.

For the second requirement, let $s \in \mathcal{S}_e \cup \{s_\vdash\}$. Since $\bigoplus_{s'} a(x_{s,s'}) = \top$ and since $\bigoplus$ is max in ranked domains, we get that there is some s' such that $a(x_{s,s'}) = \top$. This implies that $a(x_{s,s'}) \not<_D \top$ and hence $x_{s,s'} \not<_C \top$. ∎

Theorem 6.6.3: *Let C be a set of constraints and let* Pl *be a Markovian plausibility measure consistent with C. If $[s_0, \ldots, s_m] \preceq_C [s'_0, \ldots, s'_m]$, then* $\mathrm{Pl}([s_0, \ldots, s_m]) \leq \mathrm{Pl}([s'_0, \ldots, s'_m])$*. Moreover, if $[s_0, \ldots, s_m] \prec_C [s'_0, \ldots, s'_m]$, then* $\mathrm{Pl}([s_0, \ldots, s_m]) < \mathrm{Pl}([s'_0, \ldots, s'_m])$*.*

Proof: Since Pl is structured, we have that $\mathrm{Pl}([s_0, \ldots, s_m]) = t_{s_0,s_0} \bigotimes t_{s_0,s_1} \bigotimes \ldots \bigotimes t_{s_{m-1},s_m}$, and $\mathrm{Pl}([s'_0, \ldots, s'_m]) = t_{s_0,s'_0} \bigotimes t_{s'_0,s'_1} \bigotimes \ldots \bigotimes t_{s'_{m-1},s'_m}$. Moreover, since Pl is consistent with C, then from Lemma E.3.1 we have that

- if $x_{s_{i-1},s_i} \leq_C x_{s'_{\sigma(i)-1},s'_{\sigma(i)}}$ then $t_{s_{i-1},s_i} \leq t_{s'_{\sigma(i)-1},s'_{\sigma(i)}}$,
- if $x_{s_{i-1},s_i} <_C x_{s'_{\sigma(i)-1},s'_{\sigma(i)}}$, then $t_{s_{i-1},s_i} < t_{s'_{\sigma(i)-1},s'_{\sigma(i)}}$,
- if $x_{s_{i-1},s_i} \leq_C \perp$, then $t_{s_{i-1},s_i} = \perp$, and
- if $\perp <_C x_{s_{i-1},s_i}$, then $\perp < t_{s_{i-1},s_i}$.

Combining these inequalities with the fact that $\bigotimes$ is commutative, associative, monotonically increasing, and multiplicative, we get the desired conclusions. ∎

Theorem 6.6.5: *Let C be a safe set of constraints over $\mathcal{S}_e$, let $\mathcal{I} \in \mathcal{C}^M$ be a BCS system with a structured Markovian prior plausibility that is consistent with C, and let $s_a = \langle o_1, \ldots, o_m \rangle$ be a local state. If $s_m \models \varphi$ for all $[s_0, \ldots, s_m] \subseteq \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$, then $\varphi \in \mathit{Bel}(\mathcal{I}, s_a)$.*

Proof: We start by showing that $\mathrm{Pl}_a(\mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])) > \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m]))$. Let $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$. Since $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_m]$ but $[s_0, \ldots, s_m] \not\subseteq \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$, there must be some $[s'_0, \ldots, s'_m] \subseteq \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$ such that $[s_0, \ldots, s_m] \prec_C [s'_0, \ldots, s'_m]$. Applying Theorem 6.6.3, we get that $\mathrm{Pl}_a([s_0, \ldots, s_m]) < \mathrm{Pl}_a([s'_0, \ldots, s'_m])$. Using A1, we get that $\mathrm{Pl}_a([s_0, \ldots, s_m]) < \mathrm{Pl}_a(\mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m]))$. Since this is true for all $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$, and since there are only finitely many such prefixes, we can use A2 to get that

$$\mathrm{Pl}_a(\mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])) > \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])) \tag{E.1}$$

We now use this fact to show that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$. Since $s_m \models \varphi$ for all $[s_0, \ldots, s_m] \subseteq \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m])$, we have that $\mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m]) \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \varphi]$ and that $\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \neg\varphi] \subseteq \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - \mathit{MAX}^m(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m]))$. Using A1 and Eq. (E.1), we get that $\mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \varphi]) > \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \neg\varphi])$. Using Eq. (6.3) we have that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$. ∎

Proposition 6.6.6: *Let C be a safe set of constraints over $\mathcal{S}_e$, and let $\mathcal{I} \in \mathcal{C}^M$ be a ranked Markovian system with a prior plausibility that is consistent with C, and let $s_a = \langle o_1, \ldots, o_m \rangle$ be a local state. If all the m-prefixes in (i.e., that are subsets of) $MAX^m(\mathcal{R}[true, o_1, \ldots, o_m])$ are equivalent according to $\approx_C$, then $\varphi \in Bel(\mathcal{I}, s_a)$ if and only if $s_m \models \varphi$ for all $[s_0, \ldots, s_m] \subseteq MAX^m(\mathcal{R}[true, o_1, \ldots, o_m])$.*

Proof: The "if" direction follows from Theorem 6.6.5. For the "only if" direction, assume that there is some $[s_0, \ldots, s_m] \subseteq MAX^m(\mathcal{R}[true, o_1, \ldots, o_m])$ such that $s_m \models \neg\varphi$. Since all the prefixes in $MAX^m(\mathcal{R}[true, o_1, \ldots, o_m])$ are equivalent according to $\approx_C$, then by Theorem 6.6.3, all these prefixes have the same plausibility. Moreover, the plausibility of any other m-prefix in $\mathcal{R}[true, o_1, \ldots, o_m]$ must be smaller. Since $\bigoplus$ is max, $\mathrm{Pl}_a(\mathcal{R}[true, o_1, \ldots, o_{m-1}, o_m \wedge \varphi]) \leq \mathrm{Pl}_a(\mathcal{R}[true, o_1, \ldots, o_{m-1}, o_m \wedge \neg\varphi])$. Using Eq. (6.3) we get that $\varphi \notin \mathrm{Bel}(\mathcal{I}, s_a)$. ∎

E.4 Proofs for Section 6.8

We now prove Theorem 6.8.1. Roughly speaking, the proof is as follows. We use repeated applications of Theorem 6.3.2 to construct a sequence of plausibility structures for finite subalgbras of $\mathcal{F}$. We then define a plausibility measure that assigns to each set the sequence of plausibilities it receives in this sequence of plausibility structures. We then must show that this plausibility measure is based on an algebraic domain.

Let $\{(D_i, \bigoplus_i, \bigotimes_i) : i \geq 1\}$ be a sequence of algebraic domains. We define a domain $\mathcal{D}^*$ as follows. We start by defining a relation $\leq^*$ between sequences in $\times_i D_i$: $(x_1, x_2, \ldots) \leq (y_1, y_2, \ldots)$ if there is an n such that for all $i \geq n$, we have $x_i \leq_i y_i$. It straightforward to verify that this relation is a preorder. Define $\vec{x} \approx^* \vec{y}$ if $\vec{x} \leq^* \vec{y}$ and $\vec{y} \leq^* \vec{x}$. Since $\leq^*$ is a preorder, $\approx^*$ is an equivalence relation. We define D^* to be $(\times_i D_i)/\approx^*$, the set of equivalence classes in $\times_i D_i$. We define $\bigoplus^*$ and $\bigotimes^*$ in a pointwise manner on $\times_i D_i$. It is easy to check that they are well defined: if $\vec{x} \approx^* \vec{x}'$ and $\vec{y} \approx^* \vec{y}'$ then $\vec{x} \bigoplus^* \vec{y} \approx^* \vec{x}' \bigoplus^* \vec{y}'$ and $\vec{x} \bigotimes^* \vec{y} \approx^* \vec{x}' \bigotimes^* \vec{y}'$.

Lemma E.4.1: *Let $\{\mathcal{D}_i = (D_i, \bigoplus_i, \bigotimes_i) : i \geq 1\}$ be a sequence of algebraic domains. Then $\mathcal{D}^*$ as defined above is also an algebraic domain.*

Proof: The proof is identical to the proof of Proposition 6.3.1, except that we apply it to infinite sequences. We omit details here. ∎

Theorem 6.8.1: *Let $(W, \mathcal{F}, \mathrm{Pl})$ be a qualitative plausibility space such that $\mathcal{F}$ is countable. Then there is an algebraic domain $\mathcal{D}$ and a structured plausibility space $(W, \mathcal{F}, \mathrm{Pl}')$ based on $\mathcal{D}$ such that Pl' is default-isomorphic to Pl.*

Proof: Let $A_1, A_2, \ldots$ be some enumeration of the sets in $\mathcal{F}$. Without loss of generality, we can assume that $A_1 = \top$. Let $\mathcal{F}_k$ to be the minimal algebra that contains $A_1, \ldots, A_k$. Clearly $\mathcal{F}_k$ is finite for all k. Since $\mathcal{F}_k$ is finite, by Theorem 6.3.2, there is a structured plausibility measure Pl_k that is based on an algebraic domain $\mathcal{D}_k$ such that $\mathrm{Pl}_k(A) > \mathrm{Pl}_k(B)$ if and only if $\mathrm{Pl}(A) > \mathrm{Pl}(B)$ for all disjoint pairs $A, B \in \mathcal{F}_k$.

Let $\mathcal{D}^*$ be the algebraic domain defined by the sequence $\{\mathcal{D}_i : i \geq 1\}$. We define Pl' so that $\mathrm{Pl}'(A)$ is the equivalence class in D^* of the sequence $(\perp, \ldots, \perp, \mathrm{Pl}_k(A), \mathrm{Pl}_{k+1}(A), \ldots)$, where k is the least integer such that $A \in \mathcal{F}_k$.

We need to show that Pl' is indeed default-isomorphic to Pl. Let A and B be disjoint sets in $\mathcal{F}$. By construction, there is an integer k_0 such that $A, B \in \mathcal{F}_{k_0}$. Let $k \geq k_o$, then, by construction, $\mathrm{Pl}_k(A) > \mathrm{Pl}_k(B)$ if and only if $\mathrm{Pl}(A) > \mathrm{Pl}(B)$. Since this is true for all $k > k_0$, we have that $\mathrm{Pl}'(A) > \mathrm{Pl}'(B)$ if and only if $\mathrm{Pl}(A) > \mathrm{Pl}(B)$.

Finally, we need to show that $(W, \mathcal{F}, \mathrm{Pl}')$ is a structured plausibility space. Let A and B be disjoint sets in $\mathcal{F}$. Since Pl_k is structured, we get that $\mathrm{Pl}_k(A \cup B) = \mathrm{Pl}_k(A) \bigoplus_k \mathrm{Pl}_k(B)$ for all k such that $A, B \in \mathcal{F}_k$. Thus, $\mathrm{Pl}'(A \cup B) = \mathrm{Pl}'(A) \bigoplus^* \mathrm{Pl}'(B)$. Now let $A, B \in \mathcal{F}$. Again, since Pl_k is structured, we get that $\mathrm{Pl}_k(A \cap B) = \mathrm{Pl}_k(A|B) \bigotimes_k \mathrm{Pl}_k(B)$ for all k such that $A, B \in \mathcal{F}_k$. Thus, $\mathrm{Pl}'(A \cap B) = \mathrm{Pl}'(A|B) \bigotimes^* \mathrm{Pl}'(B)$. ∎

Theorem 6.8.2: *Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ be a BCS with a countable number of global states. Then there exists a system $\mathcal{I}' = (\mathcal{R}, \pi, \mathcal{P}')$ with a structured prior Pl'_a such that $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', r, m) \models \varphi$ for all runs r, times m, and formulas φ.*

Proof: Since there is a countable number of global states, we have that the minimal algebra $\mathcal{F}_0$ containing events of the form $[s_0, \ldots, s_m; o_1, \ldots, o_k]$ is also countable. Moreover, since only events in this algebra play a role in evaluating formulas in $\mathcal{L}^{KPT}$, we can assume without loss of generality that Pl_a, the prior in $\mathcal{I}$, is defined over $\mathcal{F}_0$. By Theorem 6.8.2, there is a structured plausibility space $(\mathcal{R}, \mathcal{F}_0, \mathrm{Pl}'_a)$ such that Pl'_a is default-isomorphic to Pl_a. Let $\mathcal{I}' = (\mathcal{R}, \pi, \mathcal{P}')$ where $\mathcal{P}'$ is defined by the prior Pl'_a. Clearly, $\mathrm{Pl}'_{(r,m)}$ is default-isomorphic to $\mathrm{Pl}_{(r,m)}$ for all runs r and times m. It follows that $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', r, m) \models \varphi$ for all runs r, times m, and formulas φ. ∎

Theorem 6.8.3: *Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ be a BCS. Then there is a systems $\mathcal{I}' = (\mathcal{R}', \pi', \mathcal{P}')$ and a one-to-one function f from $\mathcal{R}$ to $\mathcal{R}'$ such that $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', f(r), m) \models \varphi$ for all runs r, times m, and formulas φ. Moreover, if the prior of $\mathcal{I}$ is structured, then $\mathcal{I}' \in \mathcal{C}^{POM}$, and if $\mathcal{I}$ also satisfies UPD4, then $\mathcal{I}' \in \mathcal{C}^M$.*

Proof: Roughly speaking, the proof goes as follows. We construct $\mathcal{I}'$ such that the environment state at time m records the sequence of environment states between time 0 and time m and all the observations made until time m. This construction defines an obvious one-to-one mapping between points in $\mathcal{I}$ and points in $\mathcal{I}'$. We define Pl'_a as the projection of Pl_a via this mapping. We show that $\mathcal{I}'$ is isomorphic to $\mathcal{I}$, in the sense that corresponding points in $\mathcal{I}$ and $\mathcal{I}'$ satisfy exactly the same formulas. Additionally, we have to show that the prior in $\mathcal{I}'$ is Markovian. This however, follows trivially, since the environment states at time m record all past events. Thus, conditioning on time-$(m-1)$ events does not provide additional information when we know the state of the system at time m.

We proceed as follows. Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ be a BCS. Given $r \in \mathcal{R}$, let $f(r)$ be the run r' such that $r'_e(m) = (\langle r_e(0), \ldots, r_e(m)\rangle, r_a(m-1))$ and $r'_a(m) = r_e(m)$. Let $\mathcal{I}' = (\mathcal{R}', \pi', \mathcal{P}')$, where $R' = f(\mathcal{R})$, $\pi'(f(r), m) = \pi(r, m)$, and $\mathcal{P}'$ is defined by a prior Pl'_a such that $\mathrm{Pl}'_a(f(A)) = \mathrm{Pl}_a(A)$. It is easy to verify that $\mathcal{I}'$ is isomorphic to $\mathcal{I}$, in the sense that $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}', f(r), m) \models \varphi$ for all runs r, times m, and formulas φ.

Now assume that that the prior Pl_a of $\mathcal{I}$ is structured. Clearly, since Pl'_a is isomorphic to Pl_a, it is also structured. We now show that $\mathcal{I}'$ satisfy the requirements of the theorem.

To show that $\mathcal{I}' \in \mathcal{C}^{POM}$, we have to show that

- $\mathit{Ind}_{\mathrm{Pl}'_a}(S_{m+1} = s_{m+1}; E \mid S_m = s_m)$, where E is any time-m event,
- $\mathit{Ind}_{\mathrm{Pl}'_a}(O_m = o; E \mid S_m = s)$, where E is any time-$(m-1)$ event (i.e., an event described by $S_0, \ldots, S_{m-1}$ and $O_1, \ldots, O_{m-1}$),
- if $\mathrm{Pl}'_a(S_m = s) > \bot$ and $\mathrm{Pl}'_a(S_n = s) > \bot$, then $\mathrm{Pl}'_a(S_{m+1} = s' \mid S_m = s) = \mathrm{Pl}'_a(S_{n+1} = s' \mid S_n = s)$ and $\mathrm{Pl}'_a(O_m = o \mid S_m = s) = \mathrm{Pl}'_a(O_n = o \mid S_n = s)$ for all m and n.

Let $s'_m = (\langle s_0, \ldots, s_m\rangle, \langle o_1, \ldots, o_{m-1}\rangle)$ be some environment state in $\mathcal{I}'$. Let E be any time-$(m-1)$ event. Since s'_m records all the states of the system from time 0 up to time $m-1$, it is clear that either $\mathrm{Pl}'_a(S_m = s'_m, E) = \bot$ or $\mathrm{Pl}'_a(S_{m+1} = s'_{m+1} \mid E, S_m = s'_m) = \mathrm{Pl}'_a(S_{m+1} = s'_{m+1} \mid S_m = s'_m)$, and thus $\mathit{Ind}_{\mathrm{Pl}'_a}(S_{m+1} = s_{m+1}; E \mid S_m = s_m)$. Similar arguments show that $\mathit{Ind}_{\mathrm{Pl}'_a}(O_m = o; E \mid S_m = s_m)$. To show that Pl'_a satisfies the last requirement, notice that in $\mathcal{I}'$, $r_e(n) = r_e(m)$ if and only if $m = n$.

Now assume that $\mathcal{I}$ satisfies UPD4. We want to show that $\mathcal{I}' \in \mathcal{C}^M$. Thus, we need to show that Pl'_a restricted to the algebra of events of the form $S_m = s$ is a Markov chain, and that Pl'_a satisfies UPD4. Since Pl'_a is a partially observable Markov chain, then it is also a Markov chain. To see that Pl'_a satisfies UPD4, notice that $\mathcal{R}'[\varphi_0, \ldots, \varphi_k; o_1, \ldots, o_k] = f'(\mathcal{R}[\varphi_0, \ldots, \varphi_k; o_1, \ldots, o_k])$. Thus, since Pl_a satisfies UPD4 and since $\mathrm{Pl}'_a(f'(A)) = \mathrm{Pl}_a(A)$ for all $A \subseteq \mathcal{R}$, we conclude that Pl'_a also satisfies UPD4. ∎

Theorem 6.8.4: *Let $\varphi \in \mathcal{L}^{KPT}$.*

- *φ is valid in $\mathcal{C}^{BCS}$ if and only if φ is valid in $\mathcal{C}^{POM}$.*
- *φ is valid in $\mathcal{C}^{UPD4}$ if and only if φ is valid in $\mathcal{C}^{M}$.*

Proof: Since $\mathcal{C}^{POM} \subset \mathcal{C}^{BCS}$ and $\mathcal{C}^M \subset \mathcal{C}^{UPD}$, it suffices to show that if φ is satisfiable in $\mathcal{C}^{BCS}$ (resp., $\mathcal{C}^{UPD4}$), then it is satisfiable in $\mathcal{C}^{POM}$ (resp., $\mathcal{C}^M$).

Assume that φ is satisfiable in a system $\mathcal{I}$ such that $\mathcal{I} \in \mathcal{C}^{BCS}$ (resp., $\mathcal{C}^{UPD4}$). Using standard arguments, we can assume, without loss of generality, that $\mathcal{I}$ has a countable number of global states. By Theorem 6.8.2, we can also assume that $\mathcal{I}$ has a structured prior. Finally, by Theorem 6.8.3, there exists a system $\mathcal{I}' \in \mathcal{C}^{POM}$ (resp., in $\mathcal{C}^M$) in which φ is satisfiable. ∎

Bibliography

Adams, E. (1975). *The Logic of Conditionals*. Dordrecht, Netherlands: D. Reidel.

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

Aumann, R. J. (1993). Backwards induction and common knowledge of rationality. Presented at the Summer workshop of the Stanford Institute for Theoretical Economics.

Bacchus, F., A. J. Grove, J. Y. Halpern, and D. Koller (1993). Statistical foundations for default reasoning. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '93)*, pp. 563–569. Available by anonymous ftp from logos.uwaterloo.ca/pub/bacchus or via WWW at http://logos.uwaterloo.ca.

Battigalli, P. and G. Bonanno (1997). The logic of belief persistency. *Economics and philosophy*. Accepted for publication.

Ben-David, S. and R. Ben-Eliyahu (1994). A modal logic for subjective default reasoning. In *Proc. 9th IEEE Symp. on Logic in Computer Science*, pp. 477–486.

Ben-Porath, E. (1992). Rationality, Nash equilibrium and backward induction in perfect information games. Working paper, The Sackler Institute of Economic Studies, Tel-Aviv University.

Bicchieri, C. (1989). Self refuting theories of strategic interaction: A paradox of common knowledge. *Erkenntnis 30*, 69–85.

Boutilier, C. (1992). Normative, subjective and autoepistemic defaults: adopting the Ramsey test. In *Principles of Knowledge Representation and Reasoning: Proc. Third International Conference (KR '92)*, pp. 685–696.

Boutilier, C. (1993). Revision sequences and nested conditionals. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '93)*, pp. 519–525.

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

Boutilier, C. (1994b). An event-based abductive model of update. In *Proc. Tenth Biennial Canadian Conference on Artificial Intelligence*, pp. 241–248.

Boutilier, C. (1994c). Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence 68*, 33–85.

Boutilier, C. (1995). Generalized update: belief change in dynamic settings. In *Proc. Fourteenth International Joint Conference on Artificial Intelligence (IJCAI '95)*, pp. 1550–1556.

Boutilier, C. (1996). A unified model of qualitative belief change: A dynamical systems perspective. Submitted to publication.

Brafman, R., J.-C. Latombe, Y. Moses, and Y. Shoham (1994). Knowledge as a tool in motion planning under uncertainty. In R. Fagin (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. Fifth Conference*, pp. 208–224. San Francisco, Calif.: Morgan Kaufmann.

Burgess, J. (1981). Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic 22*, 76–84.

Crocco, G. and P. Lamarre (1992). On the connection between non-monotonic inference systems and conditional logics. In B. Nebel, C. Rich, and W. Swartout (Eds.), *Proc. Third International Conference on Principles of Knowledge Representation and Reasoning (KR '92)*, pp. 565–571. San Francisco: Morgan Kaufmann.

Darwiche, A. (1992). *A Symbolic Generalization of Probability Theory*. Ph. D. thesis, Stanford University.

Darwiche, A. and J. Pearl (1994). On the logic of iterated belief revision. In R. Fagin (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. Fifth Conference*, pp. 5–23. San Francisco: Morgan Kaufmann.

Davis, R. and W. Hamscher (1988). Model-based reasoning: troubleshooting. In H. Shrobe and T. A. A. for Artificial Intelligence (Eds.), *Exploring AI*, pp. 297–346. SF: Morgan Kaufmann.

De Finetti, B. (1972). *Probability, Induction and Statistics*. New York: John Wiley & Sons, Inc.

de Kleer, J. (1990). Using crude probability estimates to guide diagnosis. *Artificial Intelligence 45*, 381–392.

de Rijke, M. (1992). Meeting some neighbors. Research Report LP-92-10, University of Amsterdam.

Dean, T. and K. Kanazawa (1989). A model for reasoning about persistence and causation. *Computational Intelligence 5*, 142–150.

del Val, A. and Y. Shoham (1992). Deriving properties of belief update from theories of action. In *Proc. National Conference on Artificial Intelligence (AAAI '92)*, pp. 584–589.

del Val, A. and Y. Shoham (1993). Deriving properties of belief update from theories of action (II). In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '93)*, pp. 732–737.

del Val, A. and Y. Shoham (1994). A unified view of belief revision and update. *Journal of Logic and Computation 4*.

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

Dubois, D. and H. Prade (1990). An introduction to possibilistic and fuzzy logics. In G. Shafer and J. Pearl (Eds.), *Readings in Uncertain Reasoning*, pp. 742–761. San Francisco, Calif.: Morgan Kaufmann.

Dubois, D. and H. Prade (1991). Possibilistic logic, preferential models, non-monotonicity and related issues. In *Proc. Twelfth International Joint Conference on Artificial Intelligence (IJCAI '91)*, pp. 419–424.

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

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi (1995). *Reasoning about Knowledge*. Cambridge, Mass.: MIT Press.

Fikes, R. E. and N. J. Nilsson (1972). STRIPS: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence 2*, 189–208.

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

Freund, M. and D. Lehmann (1994). Belief revision and rational inference. Technical Report TR 94-16, Hebrew University.

Friedman, N. and J. Y. Halpern (1994a). Conditional logics of belief change. In *Proc. National Conference on Artificial Intelligence (AAAI '94)*, pp. 915–921.

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

Friedman, N. and J. Y. Halpern (1994c). On the complexity of conditional logics. In J. Doyle, E. Sandewall, and P. Torasso (Eds.), *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*, pp. 202–213. San Francisco: Morgan Kaufmann.

Friedman, N. and J. Y. Halpern (1995). Plausibility measures: a user's manual. In *Proc. Eleventh Conference on Uncertainty in Artificial Intelligence (UAI '95)*, pp. 175–184.

Friedman, N. and J. Y. Halpern (1996a). Belief revision: A critique. In *Principles of Knowledge Representation and Reasoning: Proc. Fifth International Conference (KR '96)*.

Friedman, N. and J. Y. Halpern (1996b). 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)*, 1994.

Friedman, N. and J. Y. Halpern (1996c). Plausibility measures and default reasoning. Submitted for publication. A preliminary version of this work appeared in *Proc. National Conference on Artificial Intelligence (AAAI '96)*, 1996, pages 1297–1304.

Friedman, N. and J. Y. Halpern (1996d). A qualitative Markov assumption and its implications for belief change. In *Proc. Twelfth Conference on Uncertainty in Artificial Intelligence (UAI '96)*, pp. 263–273.

Friedman, N. and J. Y. Halpern (1997a). Modeling belief in dynamic systems. part I: foundations. *Artificial Intelligence*. Accepted for publication. Also published as IBM Technical Report RJ 9965. 1995.

Friedman, N. and J. Y. Halpern (1997b). Modeling belief in dynamic systems. Part II: revision and update. In preparation. 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.".

Friedman, N., J. Y. Halpern, and D. Koller (1996). Conditional first-order logic revisited. In *Proc. National Conference on Artificial Intelligence (AAAI '96)*, pp. 1305–1312.

Fuhrmann, A. (1989). Reflective modalities and theory change. *Synthese 81*, 115–134.

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

Gabbay, D. M., C. J. Hogger, and J. A. Robinson (Eds.) (1993). *Nonmonotonic Reasoning and Uncertain Reasoning*, Volume 3 of *Handbook of Logic in Artificial Intelligence and Logic Programming*. Oxford, U.K.: Oxford University Press.

Gärdenfors, P. (1986). Belief revision and the Ramsey test for conditionals. *Philosophical Review 91*, 81–93.

Gärdenfors, P. (1988). *Knowledge in Flux*. Cambridge, Mass.: MIT Press.

Gärdenfors, P. and D. Makinson (1988). Revisions of knowledge systems using epistemic entrenchment. In M. Vardi (Ed.), *Proc. Second Conference on Theoretical Aspects of Reasoning about Knowledge*, pp. 83–95. San Francisco: Morgan Kaufmann.

Gärdenfors, P. and D. Makinson (1989). Relations between the theory of change and nonmonotonic logic. In A. Fuhrmann and M. Morreau (Eds.), *Logic of Theory Change. Workshop proceedings*, pp. 185–205. Springer-Verlag.

Geffner, H. (1992a). *Default Reasoning*. Cambridge, Mass.: MIT Press.

Geffner, H. (1992b). High probabilities, model preference and default arguments. *Mind and Machines 2*, 51–70.

Ginsberg, M. L. (1986). Counterfactuals. *Artificial Intelligence 30*, 35–79.

Ginsberg, M. L. (Ed.) (1987). *Readings in Nonmonotonic Reasoning*. San Francisco, Calif.: Morgan Kaufmann.

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

Goldszmidt, M. and J. Pearl (1992). Rank-based systems: A simple approach to belief revision, belief update and reasoning about evidence and actions. In *Principles of Knowledge Representation and Reasoning: Proc. Third International Conference (KR '92)*, pp. 661–672.

Grahne, G., A. Mendelzon, and R. Rieter (1992). On the semantics of belief revision systems. In Y. Moses (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. Fourth Conference*, pp. 132–142. San Francisco: Morgan Kaufmann.

Greco, G. H. (1987). Fuzzy integrals and fuzzy measures with their values in complete lattices. *Journal of Mathematical Analysis and Applications 126*, 594–603.

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

Halpern, J. Y. and R. Fagin (1989). Modelling knowledge and action in distributed systems. *Distributed Computing 3*(4), 159–179. 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".

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

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

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

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca, N.Y.: Cornell University Press.

Howard, R. A. (1971). *Dynamic Probabilistc Systems. Volume I: Markov Models*. New York: John Wiley & Sons, Inc.

Jeffreys, H. (1961). *Theory of Probability*. Oxford, U.K.: Oxford University Press.

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

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

Katsuno, H. and K. Satoh (1991). A unified view of consequence relation, belief revision and conditional logic. In *Proc. Twelfth International Joint Conference on Artificial Intelligence (IJCAI '91)*, pp. 406–412.

Kautz, H. A. (1986). Logic of persistence. In *Proc. National Conference on Artificial Intelligence (AAAI '86)*, pp. 401–405.

Keller, A. M. and M. Winslett (1985). On the use of an extended relational model to handle changing incomplete information. *IEEE Transactions on Software Engineering SE-11*(7), 620–633.

Kemeny, J. G. and J. L. Snell (1960). *Finite Markov Chains*. New Jersey: Van Nostrand.

Kraus, S. and D. Lehmann (1988). Knowledge, belief, and time. *Theoretical Computer Science 58*, 155–174.

Kraus, S., D. Lehmann, and M. Magidor (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence 44*, 167–207.

Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in finitely repeated Prisoners' Dilemma. *Journal of Economic Theory 27*(2), 245–252.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics 22*, 76–86.

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

Lehmann, D. (1995). Belief revision, revised. In *Proc. Fourteenth International Joint Conference on Artificial Intelligence (IJCAI '95)*, pp. 1534–1540.

Lehmann, D. and M. Magidor (1992). What does a conditional knowledge base entail? *Artificial Intelligence 55*, 1–60.

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

Levi, I. (1988). Iteration of conditionals and the Ramsey test. *Synthese 76*, 49–81.

Lewis, D. K. (1973). *Counterfactuals*. Cambridge, Mass.: Harvard University Press.

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

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

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

Pearl, J. (1990). System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In R. Parikh (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. Third Conference*, pp. 121–135. San Francisco: Morgan Kaufmann.

Piccione, M. and A. Rubinstein (1996). On the interpretation of decision problems with imperfect recall. This issue.

Pratt, V. R. (1979). Models of program logics. In *Proc. 20th IEEE Symp. on Foundations of Computer Science*, pp. 115–122.

Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The Foundations of Probability and other Logical Essays*. New York: Harcourt Brace.

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

Reny, P. (1992). Rationality in extensive form games. *Journal of Economic Perspectives 6*, 103–118.

Rott, H. (1991). Two methods of constructing contractions and revisions of knowledge systems. *Journal of Philosophical Logic 20*, 149–173.

Schlechta, K. (1995). Defaults as generalized quantifiers. *Journal of Logic and Computation 5*(4), 473–494.

Schlechta, K. (1996). Filters and partial orders. Submitted for publication.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, N.J.: Princeton University Press.

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

Shoham, Y. (1988). Chronological ignorance: experiments in nonmonotonic temporal reasoning. *Artificial Intelligence 36*, 271–331.

Spohn, W. (1988). Ordinal conditional functions: a dynamic theory of epistemic states. In W. Harper and B. Skyrms (Eds.), *Causation in Decision, Belief Change and Statistics*, Volume 2, pp. 105–134. Dordrecht, Netherlands: Reidel.

Stalnaker, R. C. (1992). Knowledge, belief and counterfactual reasoning in games. Forthcoming in *Proceedings of the Second Castiglioncello Conference*, edited by C. Bicchieri and B. Skyrms.

Voorbraak, F. (1992). Generalized Kripke models for epistemic logic. In Y. O. Moses (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. Fourth Conference*, pp. 214–228. San Francisco, Calif.: Morgan Kaufmann.

Wang, Z. and G. J. Klir (1992). *Fuzzy Measure Theory*. New York: Plenum Press.

Weber, S. (1991). Uncertainty measures, decomposability and admissibility. *Fuzzy Sets and Systems 40*, 395–405.

Weydert, E. (1994a). Doxastic normality logic: A qualitative probabilistic modal framework for defaults. In A. Fuhrmann and H. Rott (Eds.), *Logic, Action and Information*. New York: de Gruyter.

Weydert, E. (1994b). General belief measures. In R. López de Mantarás and D. Poole (Eds.), *Proc. Tenth Conference on Uncertainty in Artificial Intelligence (UAI '94)*, pp. 575–582. San Francisco: Morgan Kaufmann.

Williams, M. (1994). Transmutations of knowledge systems. In J. Doyle, E. Sandewall, and P. Torasso (Eds.), *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*, pp. 619–629. San Francisco: Morgan Kaufmann.

Winslett, M. (1988). Reasoning about action using a possible models approach. In *Proc. National Conference on Artificial Intelligence (AAAI '88)*, pp. 89–93.