Chapter 8

Combining Modalities

Up to now, we have looked at reasoning about knowledge, (various types of) reasoning about uncertainty, reasoning about defaults, and reasoning about counterfactuals, all separately. However, it should be clear that in many cases we want to combine such reasoning. One obvious example is combining counterfactuals and probabilities. In Chapter 7, we had an example of counterfactual reasoning involving a lawyer arguing that his client would not have hit Mrs. McGraw's cow if the brakes had been functioning properly (despite having been drunk and driving in the rain). In fact, the lawyer may want to argue only that, with some reasonable probability, his client would not have hit Mrs. McGraw's cow. This requires reasoning about counterfactuals and probability. Similarly, we may want to combine reasoning about defaults and knowledge, so that an agent can say "I know that birds typically fly but I do not know whether penguins typically fly."

It is easy to construct the syntax of a logic for doing such combined reasoning, by simply combining the syntaxes of the logics of the independent notions. Thus, a statement such as

$$K(\mathit{bird} \rightarrow \mathit{fly}) \wedge \neg K(\mathit{penguin} \rightarrow \mathit{fly}) \wedge \neg K \neg(\mathit{penguin} \rightarrow \mathit{fly})$$

captures the statement at the end of the previous paragraph. It seems that it should be equally straightforward to provide semantics for the logic with combined modalities, by combining the structures that we saw earlier. This is indeed almost true. The only subtlety comes in describing the relationship between the modalities and the notions of "possible world" used by each of the modalities. For example, as we saw in Section 7.5, the set $W_{w,i}$ of "possible worlds" used in counterfactual reasoning includes worlds that the agent knows perfectly well are impossible. Thus, to reason about both counterfactuals and probability requires, in general, two different sets of possible worlds for agent i at world w, call them $W^p_{w,i}$ and $W^c_{w,i}$, where $W^p_{w,i}$ is used when doing probabilistic reasoning and $W^c_{w,i}$ is used for doing counterfactual reasoning. Are they related in any way?

It seems reasonable to require that $W^p_{w,i}$ be a subset of $W^c_{w,i}$—the worlds considered possible for probabilistic reasoning should certainly all be considered possible for counterfactual reasoning—but the converse may not hold. It might also seem reasonable to require, if we are using a preference order to model similarity to w, that worlds in $W^p_{w,i}$ be closer to w than worlds not in $W^c_{w,i}$. However, some thought shows that this may not be so. For example, suppose that there are three primitive propositions, p, q, and r, and the agent knows that p is true if and only if exactly one of q or r is true. Originally, the agent considers two worlds possible, w_1 and w_2, and assigns each of them probability 1/2; the formula $p \wedge q \wedge r$ is true in w_1, while $p \wedge \neg q \wedge \neg r$ is true in w_2. Now what is the closest world to w_1 where q is false? Is it necessarily w_2? That depends on the mechanism by which q was made false, something which is not modeled in the counterfactual structure. The agent could well decide that in the closest world to w_1 where q is false, r is still false; since p is true iff exactly one of q or r is true, p would be false in this world too. That is, the closest world would not be w_2, but one where $\neg p \wedge \neg q \wedge \neg r$ is true.

In this chapter, I focus on two examples of combining modalities—reasoning about knowledge and probability and reasoning about knowledge and belief. However, most of the points made here apply equally well to other cases of combined modalities.

8.1 Reasoning About Knowledge and Probability

As suggested above, constructing the syntax for a combined logic of knowledge and probability is straightforward. Let $\mathcal{L}^{KQU}_n$ be the result of combining the syntaxes of $\mathcal{L}^K_n$ and $\mathcal{L}^{QU}_n$ in the obvious way. $\mathcal{L}^{KQU}_n$ allows statements such as $K_1(\ell_2(\varphi) = 1/3)$—agent 1 knows that, according to agent 2, the probability of φ is 1/3. It also has facilities for asserting uncertainty regarding probability. For example,

$$K_1(\ell_1(\varphi) = 1/2 \vee \ell_1(\varphi) = 2/3) \wedge \neg K_1(\ell_1(\varphi) = 1/2) \wedge \neg K_1(\ell_1(\varphi) = 2/3)$$

says that agent 1 knows that the probability of φ is either 1/2 or 2/3, but he does not know which.

The semantics of $\mathcal{L}^{KQU}_n$ can be given using *(Kripke) structures for knowledge and probability.* Not surprisingly, these are tuples of the form

$M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n, \pi)$, where $\mathcal{PR}_i(w) = (W_{w,i}, \mathcal{F}_{w,i}, \mu_{w,i})$. The rules for assigning truth values to formulas in $\mathcal{L}_n^{KQU}$ at worlds in such a structure are just the result of combining the rules for knowledge and probability separately, from Sections 2.2 and 6.3. I omit the details here.

What is the relationship between $\mathcal{K}_i(w)$ and $W_{w,i}$? It might seem obvious that we should take them to be equal; the agent then puts probability on the set of worlds she considers possible. But this may not always be the right thing to do. The following example may help to clarify the issues.

Example 8.1.1 Alice chooses a number, either 0 or 1, and writes it down. She then tosses a fair coin. If the outcome of the coin toss agrees with the number chosen (that is, if the number chosen is 1 and the coin lands heads or the number chosen is 0 and the coin lands tails), then she performs an action **a**; otherwise she does not. Suppose that Bob does not know Alice's choice. What is the probability, according to Bob, that Alice performs action **a**? What is the probability according to Alice? (For definiteness, assume that both of these probabilities are to be assessed in the situation after Alice has chosen the number, but before the coin is tossed.)

It should be clear that, according to Alice, who knows the number chosen, the probability (before she tosses the coin) that she performs action **a** is 1/2. There is also a reasonable argument to show that, even according to Bob (who does not know the number chosen) the probability is 1/2. Clearly from Bob's viewpoint, if Alice chose 0, then the probability that Alice performs action **a** is 1/2 (since the probability of the coin landing heads is 1/2); similarly, if Alice chose 1, then the probability of her performing action **a** is 1/2. Since, no matter what number Alice chose, the probability according to Bob that Alice performs action **a** is 1/2, it seems reasonable to say that Bob *knows* that the probability of Alice's performing action **a** is 1/2.

Note that this argument does not assume a probability for the event that Alice chose 0. This is a good thing: No probability is provided by the problem statement, so none should be assumed. ∎

Clearly this example involves reasoning about knowledge and probability, so we should be able to model it in a Kripke structure for knowledge and probability. I in fact consider three potential structures, that differ only in Bob's probability assignment $\mathcal{PR}_B$. Let $M^i = (W, \mathcal{K}_A, \mathcal{K}_B, \mathcal{PR}_A, \mathcal{PR}_B^i, \pi)$, $i = 1, 2, 3$. Most of the components of M^i are defined in (what I hope by now is) the obvious way. $W = \{(0, H), (0, T), (1, H), (1, T)\}$: Alice chose 0 and the coin lands heads, Alice chose 0 and the coin lands tails, and so on. (Actually, it would be cleaner to represent this as a two-step process: first the number is chosen, and then the coin is tossed. This is done in Chapter 8, when I add time to the framework.) Bob cannot distinguish any of these worlds; in any one of them, he considers all four possible. Thus, $\mathcal{K}_B(w) = W$ for all $w \in W$. On the other hand, since Alice knows the number she chose, in a world of the form $(0,x)$, Alice considers only worlds of the form $(0,y)$ possible, while in a world of the form $(1,x)$, Alice considers only worlds of the form $(1,y)$ possible. Alice's probability measures $\mathcal{PR}_A$ is also what we would expect. $W_{w,A} = \mathcal{K}_A(w) = \{(0,H),(0,T)\}$ and $\mu_{w,A}(0,H) = \mu_{w,A}(0,T) = 1/2$ for $w \in \{(0,H),(0,T)\}$, $(0,x)$, and similarly for $w \in \{(1,H),(1,T)\}$.

Suppose that we take as primitive propositions c_0, c_1, h, t, and a, representing that Alice chooses 0, Alice chooses 1, the coin will land heads, the coin will land tails, and Alice performs action a, respectively, and define π in the obvious way. Note that a holds in the worlds $(0,T)$ and $(1,H)$. Then, for example, $(M^i,(1,H)) \models K_A(c_1 \wedge K_A(\ell_A(h) = 1/2 \wedge \ell_A(a) = 1/2)$: in the world where Alice chooses 1 and the coin will land heads, Alice knows that she chooses 1 and knows that the probability that the coin will land heads (and hence that she will perform action **a**) is 1/2.

It remains to define $\mathcal{PR}^i_B$, $i = 1,2,3$. If Bob could assign some probability α to Alice choosing 0, there would be no problem: in all worlds w, we could take $W_{w,B} = W$ and define $\mu_{w,B}$ so that both $(0,H)$ and $(0,T)$ get probability $\frac{\alpha}{2}$, while both $(1,H)$ and $(1,T)$ gets probability $\frac{1-\alpha}{2}$. This gives us a set of probability measures on W, parameterized by α.

It is not hard to show that for any choice of α, the event that action **a** is performed (i.e., the event $[\![a]\!]_M = \{(0,T),(1,H)\}$) has probability 1/2. A Bayesian might feel that each agent should choose *some* α and work with that. There are two arguments against this viewpoint. The first argument is pragmatic. Since the problem statement does not give α, any particular choice may lead to conclusions beyond those justified by the problem. Thus, we should rightly be suspicious of conclusions drawn on the basis of a particular α. The second argument is more philosophical: adding a probability seems unnecessary here. As we shall see in Chapter 8, this is more than just a philosophical issue, since in many problems that arise in computer science, we are faced with a situation that are best thought of as having both probabilistic uncertainty (the outcome of the coin toss, in this case) and nonprobabilistic nondeterminism (Alice's choice).

One solution to this problem might be to use a structure for knowledge and lower probability, so that there are sets of probabilities rather than a single probability. In this case, the set would consist of all possible choices for α. While something like this would work, it seems unnecessary. After all, our informal argument above did not seem to need the assumption that there was some probability of Alice choosing 0. And since the argument did not seem to need it, it seems reasonable to hope to model the argument without using it.

The next thought might be to use nonmeasurable sets. Define $\mathcal{PR}_B^1$ so that $\mathcal{PR}_B^1(w) = (W, \mathcal{F}^1, \mu^1)$ for all $w \in W$, where $\mathcal{F}^1$ is the algebra with basis $\{(0,H),(1,H)\}$ and $\{(0,T),(1,T)\}$ and $\mu^1(\{(0,H),(1,H)\}) = \mu^2(\{(0,T),(1,T)\}) = 1/2$. That is, in M^1, the only events to which Bob assigns a probability are those for which the problem statement gives a probability: the event of the coin landing heads and the event of the coin landing tails. Of course, both these events are assigned probability 1/2.

The problem with this approach is that $\{(0,T),(1,H)\}$ is not in $\mathcal{F}^1$. Thus, it is not true that Bob believes that Alice performs action **a** with probability 1/2. The event that Alice performs action **a** is not assigned a probability at all according to this approach!

The second approach does not make the probability space independent of the world. Rather, it uses a different probability space for Bob, depending on whether Alice chooses 0 or 1. Bob's probability space when Alice chooses 0 consists of the two worlds where 0 is chosen, and similarly when Alice chooses 1. In this probability space, all worlds are measurable and have the obvious probability. In fact, $\mathcal{PR}_B^2(w) = \mathcal{PR}_A(w)$ for all $w \in W$; Bob's probability assignment is the same as Alice's.

Structure M^2 supports the reasoning in the example. In fact, $(M^2, w) \models \ell_B(a) = 1/2$ for every world w, . To see this, consider for example the world $(0,T)$. Since $[\![a]\!]_{M^2} = \{(0,T),(1,H)\}$ and $W_{(0,T),B} = \{(0,T),(0,H)\}$, it follows that $[\![a]\!]_{M^2} \cap W_{(0,T),B} = \{(0,T)\}$. By definition, $\mu_{(0,T)}((0,T)) = 1/2$. Thus, $(M^2,(0,T)) \models \ell_B(a) = 1/2$. Similar argument show that $\ell_B(a) = 1/2$ is true at every other world. It follows that $M^2 \models K_B(\ell_B(a) = 1/2)$: Bob *knows* that the probability that Alice performs action **a** is 1/2. Similarly, in this structure Bob know that the probability that the coin lands heads is 1/2 and that the probability that the coin lands tails is 1/2.

What is the probability that 0 was chosen, according to Bob? In the worlds where 0 is actually chosen—that is, $(0,H)$ and $(0,T)$—it is 1; in the other two worlds, it is 0. So

$$M^2 \models K_B(\ell_B(c_0) = 0 \vee \ell_B(c_0) = 1);$$

similar reasoning shows that

$$M^2 \models K_B((c_0 \Rightarrow \ell_B(c_0) = 1) \wedge (c_1 \Rightarrow \ell_B(c_0) = 0)$$

(Exercise 8.1).

While M^2 seems to capture the reasoning in the example in an elegant way, its very elegance opens up a whole new can of worms. If $W_{w,i}$ can be a strict subset of $\mathcal{K}_i(w)$ (as is the case in M^2), then how should it be chosen? Should there be any constraints on the choice? Consider the structure M^3, where $\mathcal{PR}_B^3(w) = (\{w\}, \mathcal{F}_w, \mu_w)$ for each world $w \in W$. $\mathcal{F}_w$

and μ_w are completely determined in this case, because the set of possible worlds is a singleton: $\mathcal{F}_w$ consists of $\{w\}$ and $\emptyset$, and μ_w assigns probability 1 and 0, respectively, to these sets. M^3 does not support the reasoning of the example; it is easy to check that $M^3 \models K_B(\ell_B(a) = 0 \vee \ell_B(a) = 1)$ (Exercise 8.2). But is there anything intrinsically wrong with the structure M^3? I would argue there isn't. The one-coin example from Chapter 1 shows why.

Example 8.1.2 This time Alice just tosses a fair coin, and looks at the outcome. What is the probability of heads according to Bob? (I am now interested in the probability *after* the coin toss.) Clearly before the coin was tossed, the probability of heads according to Bob was 1/2. There seem to be two competing intuitions regarding the probability of heads after the coin is tossed. One says the probability is still 1/2. After all, Bob has not learned anything about the outcome of the coin toss, so why should he change his valuation of the probability? On the other hand, runs the counterargument, once the coin has been tossed, can we really talk about the probability of heads? It has either landed heads or tails, so at best, Bob can say that the probability is either 0 or 1, but he doesn't know which. ∎

How can we model this example? There are two reasonable candidate structures, which again differ only in Bob's probability assignment. Let $M^i = (\{H,T\}, \mathcal{K}_A, \mathcal{K}_B, \mathcal{PR}_A, \mathcal{PR}_B^i, \pi)$, $i = 4, 5$, where

- $\mathcal{K}_A(w) = \{w\}$, for $w \in \{H,T\}$ (Alice knows the outcome of the coin toss)
- $\mathcal{K}_B(w) = \{H,T\}$ (Bob does not know the outcome)
- $\mathcal{PR}_A(w) = (\{w\}, \mathcal{F}_w, \mu_w)$ (Alice puts the obvious probability on her set of possible worlds, which is a singleton, just as in $\mathcal{PR}_B^3$)
- π assigns the obvious truth values to the primitive propositions h and t (for heads and tails).

It remains to define $\mathcal{PR}_B^4$ and $\mathcal{PR}_B^5$. $\mathcal{PR}_B^4$ is the probability assignment corresponding to the answer 1/2; $\mathcal{PR}_B^4(w) = (\{H,T\}, 2^{\{H,T\}}, \mu)$, where $\mu(H) = \mu(T) = 1/2$, for both $w = H$ and $w = T$. That is, according to $\mathcal{PR}_B^4$, Bob uses the same probability space in both of the worlds he considers possible, and in this probability space, assigns both heads and tails probability 1/2. On the other hand, $\mathcal{PR}_B^5(w) = \mathcal{PR}_A(w) = (\{w\}, \mathcal{F}_w, \mu_w)$. It is easy to see that $M^4 \models K_B(\ell_B(h) = 1/2)$ while $M^5 \models K_B(\ell_B(h) = 0 \vee \ell_B(h) = 1)$ (Exercise 8.3).

Thus, the framework can capture both arguments but, unfortunately, does not say which one is "right". I would argue that there is no one right answer here. A useful way to think of this is in terms of betting games. Consider Example 8.1.2 again. Imagine that besides Bob, Charlie is also watching Alice toss the coin. Before the coin is tossed, Bob may be willing to accept an offer from either Alice or Charlie to bet \$1 for a payoff of \$2 if the coin lands heads. Half the time the coin will land heads and Bob will be \$1 ahead, and half the time the coin will land tails and Bob will lose \$1. On average, he will break even. On the other hand, Bob is clearly not willing to accept such an offer from Alice after the coin was tossed (since we assumed that Alice saw the outcome of the coin toss), although he might still be willing to accept such an offer from Charlie. Roughly speaking, when playing against Charlie, it is appropriate for Bob to act as if the probability of heads is $1/2$, whereas while playing against Alice, he should act is if it is either 0 or 1, but he does not know which.

This example suggests that the choice of probability assignment for Bob should in general depend the adversary that Bob is playing against (and, in particular, what that adversary knows). This is indeed the case. The following general framework, which is appropriate for many examples that arise in practice, helps make this intuition somewhat more precise.

Suppose that $W \subseteq V_1 \times V_2$, for some V_1, V_2, where $V_1 = \{v_1, \ldots, v_k\}$. W can be partitioned into k disjoint sets $W^1, \ldots, W^k$, according to the first component; that is, W^j consists of all the pairs (v_j, v') in W. Assume that for each $v_j \in V_1$, there is a probability measure μ^j on the set W^j for which all sets are measurable. In Example 8.1.1, $V_1 = \{0, 1\}$ and $V_2 = (H, T)$, and there are natural measures on each of $\{(0, H), (0, T)\}$ and $\{(1, H), (1, T)\}$ (which essentially put equal probability on each of H and T). These probability measures can be viewed as defining common prior probabilities, before the agents get any information.

Given a world w, suppose that W^h is the unique element of the partition of W that contains w. Define $\mathcal{PR}_i^i(w) = (W^h \cap \mathcal{K}_i(w), \mu^h|\mathcal{K}_i(w))$. In the special case where there is a measure on all of W (so that $W^h = W$), this amounts to agent i conditioning the prior probability on the set of worlds that i considers possible. $\mathcal{PR}_i^i$ is in fact a reasonable probability assignment; as we shall see in Chapter 8, it is appropriate for many applications. However, it does not take the other agents into account. I claim that to take agent j into account, the probability assignment $\mathcal{PR}_i^j$ should be used, where $\mathcal{PR}_i^j(w) = (W^h \cap \mathcal{K}_i(w) \cap \mathcal{K}_j(w), \mu^h|(\mathcal{K}_i(w) \cap \mathcal{K}_j(w)))$ for $w \in W^h$. (Since all sets are measurable, there is no need to specify the algebra here.) While the following example does not provide a formal justification of this claim, it may help explain the intuition.

Example 8.1.3 Suppose that Alice has two dice, of which one is fair and the other biased. The biased die lands on 3 with probability 1/2, and on all the other numbers with probability 1/10. Alice chooses one of the dice and rolls it. Alice knows which die she used. Although she does not know the outcome of the roll, she knows whether it is an even or an odd number. What is the probability of Alice rolling 3 or less, according to Bob?

The set W of possible worlds is $\{(FD,1),\ldots,(FD,6),(BD,1),\ldots,(BD,6)\}$, where (FD,j) means that the fair die is rolled and lands j, while (BD,j) means that the biased die is rolled and lands j. Since Alice knows which die was used and whether the outcome was even or odd, it is easy to construct $\mathcal{K}_A$. For example, $\mathcal{K}_A((BD,1)) = \{(BD,1),(BD,3),(BD,5)\}$. Since Bob has no idea what die was used or how it landed, $\mathcal{K}_B(w) = W$ for all worlds w. Partition W into two sets, W^{FD} and W^{BD}, where W^{FD} consists of all worlds of the form (FD,j) and W^{BD} consists of all worlds of the form (BD,j). There are obvious probability measures μ^{FD} and μ^{BD} on these sets. Constructing $\mathcal{PR}_A$ is straightforward; just condition the prior probability on Alice's information. Thus, for example, at the world $(BD,1)$, Alice's probability measure puts probability 5/7 on $(BD,3)$ and probability 1/7 on each of $(BD,1)$ and $(BD,5)$.

What about Bob's probability assignment? The assignment $\mathcal{PR}_B^B$ is such that $\mathcal{PR}_B^B(w) = (W^X, \mu^X)$ if $w \in W^X$, for $X \in \{BD, FD\}$. That is, the set of worlds in the probability space used at w is determined by whether the fair coin or biased coin is used at w, even though Bob does not know which was used. Notice that every subset of W^X is measurable. Consider the primitive propositions $o_1,\ldots,o_6$, whose intended meaning is "the outcome is 1", ..., "the outcome is 6"; suppose that π interprets these primitive proposiitons in the intended way. Let $o_{\le 3}$ be an abbreviation for $o_1 \vee o_2 \vee o_3$; thus, $o_{\le 3}$ represents the event of getting 3 or less. Finally, let $M^6 = (W, \mathcal{K}_A, \mathcal{K}_B, \mathcal{PR}_A, \mathcal{PR}_B^B, \pi)$. Then

$$\begin{aligned} M^6 \models \quad & K_B(\ell_B(o_{\le 3}) = 1/2 \vee \ell_B(o_{\le 3}) = 7/10) \wedge \\ & K_B(K_A(\ell_A(o_{\le 3}) = 6/7) \vee K_A(\ell_A(o_{\le 3}) = 2/3) \vee K_A(\ell_A(o_{\le 3}) = 1/3)) \end{aligned}$$

(Exercise 8.4).

According to $\mathcal{PR}_B^B$, Bob knows that the probability of the die landing 3 or less is either 1/2 or 7/10 (depending on whether Alice used the fair or biased die). Do these numbers give him some guidance on how to bet? Suppose that Charlie is as ignorant as Bob about which die was used and how it landed. Charlie offers Bob a bet where Bob get \$$\beta$ if the outcome of the die was actually 3 or less and loses \$1 otherwise. Should Bob accept the bet? Clearly the answer depends on β. If $\beta > 1$, then Bob should certainly accept (assuming he accepts whenever he thinks his expected payoff is nonnegative). If the die is fair, he wins \$$\beta$ with probability 1/2 and loses \$1 with probability 1/2, so he comes out ahead. If the die is biased, he does even better. Similar arguments show that if $\beta < 3/7$, then Bob should certainly reject. For β between 3/7 and 1, both accepting and rejecting are reasonable courses of action. Thus, in this case, the numbers do provide Bob with reasonable guidance as far as betting behavior goes.

On the other hand, if Alice is offering Bob the bet, then this analysis is clearly incorrect. Alice has extra information, and she certainly may be taking that into account when she offers the bet. For example, if Alice chooses the biased die and she learns that the outcome was odd, then she knows that the probability of $o_{\leq 3}$ is 6/7. Bob obviously should take Alice's extra information into account when deciding whether or not to accept an offer to bet. I claim that he can do this by using the probability assignment $\mathcal{PR}_B^A$, where $\mathcal{PR}_B^A(w)\mathcal{PR}_A(w)$. The details of the argument can be found in Exercise 8.5. ∎

8.2 Axiomatizing Knowledge and Probability

We saw that the axioms that characterize reasoning about knowledge depend in part on the assumptions we make about the $\mathcal{K}_i$ operator. For example, if $\mathcal{K}_i$ is reflexive then axiom K2, $K_i\varphi \Rightarrow \varphi$, is sound. Suppose, for definiteness, that $\mathcal{K}_i$ is an equivalence relation, so that knowledge is characterized by the axiom system S5$_n$. Reasoning about probability is characterized by the axiom system $\mathcal{L}_n^{QU}$. Are there additional properties for reasoning about knowledge and probability? As the following result shows, there aren't, at least, as long as no additional assumptions are made about the interactions between knowledge and probability.

Let $\mathcal{M}_n^{K,prob}$ consist of all structures for knowledge and probability for n agents, and let $\mathcal{M}_n^{K,meas}$ consist of all the structures for knowledge and probability for n agents where the probability is measurable, that is, $\mathcal{M}_n^{K,meas}$ consists of all consist of the structures for knowledge and probability of the form $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n, \pi)$ such that $(W, \mathcal{PR}_1, \ldots, \mathcal{PR}_n, \pi) \in \mathcal{M}_n^{meas}$. Let AX$_n^{K,prob}$ consist of the axioms and inference rules of S5$_n$ for knowledge together with the axioms and inference rules of AX$_n^{prob}$ for probability. Let AX$_n^{K,bel}$ consist of the axioms and inference rules of S5$_n$ and AX$_n^{bel}$.

Theorem 8.2.1 *$AX_n^{K,prob}$ (resp., $AX_n^{K,bel}$) is a sound and complete axiomatization with respect to $\mathcal{M}_n^{K,meas}$ (resp., $\mathcal{M}_n^{K,prob}$) for the language $\mathcal{L}_n^{KQU}$.*

Proof Soundness is immediate from the soundness of $S5_n$ and AX_n^{prob}; completeness is beyond the scope of the book. ∎

We typically do want to make some assumptions about the interactions between knowledge and probability. In the rest of this section, I briefly discuss four such possible interactions and their axiomatic characterization.

Suppose that $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n, \pi)$. We have seen that we do not necessarily want to take $W_{w,i} = \mathcal{K}_i(w)$. In the structure M^2 used to model Example 8.1.1, $W_{w,B}$ was a strict subset of $\mathcal{K}_B(w)$. But what if $W_{w,i}$ is not even a subset of $\mathcal{K}_i(w)$? This does not seem so reasonable. In particular, it allows the agent to place positive probability on a fact that he knows to be false; for example, it would allow the formula $K_i \neg p \wedge \ell_i(p) > 0$ to be satisfiable. An agent who places positive probability on an event he knows to be false can be viewed as inconsistent. Thus, the following condition, which prevents this, is called CONS (for *cons*istent).

CONS. For all i and w, if $\mathcal{PR}_i(w) = (W_{w,i}, \mathcal{F}_{w,i}, \mu_{w,i})$, then $W_{w,i} \subseteq \mathcal{K}_i(w)$.

Another asssumption that seems reasonable is that $\mathcal{PR}_{w,i}$ depends only on the agent's knowledge at world w, not on other features of w. This would mean that $\mathcal{PR}_{w,i}$ is the same for all worlds w that agent i cannot distinguish. This property is called SDP (*state-determined probability*).

SDP. For all i, v, and w, if $v \in \mathcal{K}_i(w)$, then $\mathcal{PR}_i(v) = \mathcal{PR}_i(w)$.

Although SDP is a standard assumption (it is almost always made in the economics literature, for example), as we have seen, for some analyses it is inappropriate. For example, the structure M^2 used to capture Example 8.1.1 does not satisfy it. M^2 does, however, satisfy a weaker property, called *uniformity*. Roughly speaking, uniformity holds if $\mathcal{K}_i(s)$ can be partitioned into subsets such that at every point in a given subset T, the probability space is the same. Uniformity is formalized as follows:

UNIF. For all i, v, and w, if $\mathcal{PR}_i(w) = (W_{w,i}, \mathcal{F}_{w,i}, \mu_{w,i})$ and $v \in W_{w,i}$, then $\mathcal{PR}_i(v) = \mathcal{PR}_i(w)$.

In the presence of CONS, UNIF says that $\mathcal{K}_i(w)$ can be partitioned in such a way that, for each cell W' in the partition, $\mathcal{PR}_i(u) = \mathcal{PR}_i(v)$ and $W_{u,i} \subseteq W'$ for all worlds $u, v \in W'$. That is, the probability spaces are the same for worlds in the same cell of the partition and are contained in the cell (Exercise 8.6). Actually, without loss of generality, $W_{w_j,i} = W'$; that is, the probability is placed on the cell itself. This is what happens, for example, in the structure M^2. In the presence of CONS, SDP implies UNIF, although in general it does not (Exercise 8.7).

One last assumption, which is particularly prevalent in the economics literature, is the *common prior* (CP) assumption. This assumption asserts that the agents have a common prior probability on the set of all worlds and each agent's probability assignment at world w is induced from this common prior by conditioning on his set of possible worlds. Thus, CP implies SDP and CONS (and hence UNIF), since it requires that $\mathcal{K}_i(w) = W_{w,i}$.

CP. There exists a probability space $(W, \mathcal{F}_W, \mu_W)$ such that $\mathcal{PR}_i(w) = (\mathcal{K}_i(w), \mathcal{F}_W|\mathcal{K}_i(w), \mu_{w,i}))$ for for all agents i and worlds $w \in W$, where $\mathcal{F}_W|\mathcal{K}_i(w)$ consists of all sets of the form $U \cap \mathcal{K}_i(w)$ for $U \in \mathcal{F}_W$, and $\mu_{w,i} = \mu_W|\mathcal{K}_i(w)$ if $\mu_W(\mathcal{K}_i(w)) > 0$. (There are no constraints on $\mu_{w,i}$ if $\mu_W(\mathcal{K}_i(w)) = 0$.)

Until quite recently, the common prior assumption was almost an article of faith among economists. It says that differences in beliefs among agents can be completely explained by differences in information. Essentially, the picture is that agents start out with identical prior beliefs (the common prior) and then condition on the information that they later receive. If their later beliefs differ, it must thus be due to the fact that they have received different information.

CP limits the class of structures in interesting ways. For example, suppose that $M = (\{w_1, w_2\}, \mathcal{K}_1, \mathcal{K}_2, \mathcal{PR}_1, \mathcal{PR}_2, \pi)$ and that both agents consider both worlds possible, that is, $\mathcal{K}_1(w_1) = \mathcal{K}_1(w_2) = \mathcal{K}_2(w_1) = \mathcal{K}_2(w_2) = \{w_1, w_2\}$. It is easy to see that the only way that this structure can be consistent with CP is if $\mathcal{PR}_1(w_1) = \mathcal{PR}_2(w_1) = \mathcal{PR}_1(w_1) = \mathcal{PR}_2(w_2)$ (Exercise 8.8). But there are less trivial constraints placed by CP, as the following example shows.

Example 8.2.2 Consider the frame F described in Figure 8.1. (Recall from Section 2.2.4 that a frame is a Kripke structure without the interpretation π.) There are four worlds in this frame, $\{w_1, \ldots, w_4\}$. $\mathcal{K}_1$ partitions the worlds into two equivalence classes, $\{w_1, w_2\}$ and $\{w_3, w_4\}$; $\mathcal{K}_2$ partitions them into two other equivalence classes, $\{w_1, w_3\}$ and $\{w_2, w_4\}$. Whatever two worlds agent 1 considers possible, he ascribes them both probability 1/2. Agent 2, however, thinks that w_3 is twice as likely w_1 and w_2 is twice as likely as w_4. It is easy to see that this frame does not satisfy CP (Exercise 8.11). ∎

There are axioms that characterize each of CONS, SDP, and UNIF. Define an *i-likelihood formula* to be one of the form $a_1\ell_i(\varphi_1) + \cdots + a_k\ell_i(\varphi_k) \geq b$. That is, it is a formula where the only likelihood terms involve agent i. Consider the following three axioms:

KP1. $K_i\varphi \Rightarrow (\ell_i(\varphi) = 1)$.

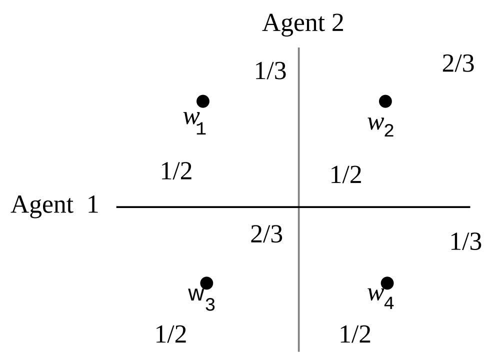

Figure 8.1: A frame that does not satisfy CP.

KP2. $\varphi \Rightarrow K_i\varphi$ if φ is an i-likelihood formula.

KP3. $\varphi \Rightarrow (\ell_i(\varphi) = 1)$ if φ is an i-likelihood formula or the negation of an i-likelihood formula.

In a precise sense, KP1 captures CONS, KP2 captures SDP, and KP3 captures UNIF. KP1 essentially says that the set of worlds that agent i considers possible has probability 1 (according to agent i). It is easy to see that KP1 is sound in structures satisfying CONS. Since SDP says that agent i knows his probability space (in that it is the same for all worlds in $\mathcal{K}_i(w)$), it is easy to see that SDP implies that in a given world, agent i knows all i-likelihood formulas that are true in that world. Thus, KP2 is sound in structures satisfying SDP. Finally, since a given i-likelihood formula has the same truth value at all worlds where agent i's probability assignments is the same, the soundness of KP3 in structures satisfying UNIF is easy to verify.

As stated, KP3 applies to both i-likelihood formulas and their negations, while KP2 as stated applies to only i-likelihood formulas. It is straightforward to show, using the axioms of S5$_n$, that KP2 also applies to negated i-likelihood formulas (Exercise 8.9). With this observation, it is almost immediate that Axioms KP1 and KP2 together imply KP3, which is reasonable since CONS and SDP together imply UNIF (Exercise 8.10).

The next theorem makes the correspondence between various properties and axioms precise.

Theorem 8.2.3 *Let $\mathcal{A}$ be a subset of {CONS,SDP,UNIF} and let A be the corresponding subset of {KP1,KP2,KP3}. Then $AX_n^{K,prob} \cup A$ (resp., $AX_n^{K,bel} \cup A$) is a sound and complete axiomatization for the language $\mathcal{L}_n^{KQU}$ for structures in $\mathcal{M}_n^{K,meas}$ (resp., $\mathcal{M}_n^{K,prob}$) satisfying $\mathcal{A}$.*

Proof As usual, soundness is straightforward (Exercise 8.12) and completeness is beyond the scope of this book. ∎

Despite the fact that CP puts some nontrivial constraints on structures, it turns out that CP adds no new properties in the language $\mathcal{L}_n^{KQU}$ beyond those already implied by CONS and SDP.

Theorem 8.2.4 *$AX_n^{K,prob} \cup \{KP1, KP2\}$ is a sound and complete axiomatization with respect to structures in $\mathcal{M}_n^{K,meas}$ satisfying CP, for the language $\mathcal{L}_n^{KQU}$.*

Although CP does not lead to any new axioms in the language $\mathcal{L}_n^{KQU}$, things change significantly if we add *common knowledge* to the language. Common knowledge of φ holds if everyone knows φ, everyone knows that everyone knows φ, everyone knows that everyone knows that everyone knows, and so on. It is straightforward to extend the logic of knowledge introduced in Section 2.2 to capture common knowledge. We add the modal operator C (for common knowledge) to the language $\mathcal{L}_n^{KQU}$ to get the language $\mathcal{L}_n^{KQUC}$. Let $E^1\varphi$ be an abbreviation for $K_1\varphi \wedge \ldots \wedge K_n\varphi$ and let $E^{m+1}(\varphi)$ be an abbreviation $E^1(E^m\varphi)$. Thus, $E\varphi$ is true if all the agents in $\{1, \ldots, n\}$ know φ, while $E^3\varphi$, for example, is true if everyone knows that everyone knows that everyone knows φ. Given a structure $M \in \mathcal{M}_n^{K,prob}$, define

$$(M, w) \models C\varphi \text{ iff } (M, w) \models E^k\varphi \text{ for all } k \geq 1.$$

In the language $\mathcal{L}_n^{KQUC}$, CP does result in interesting new axioms. In particular, in the presence of CP, agents cannot disagree on the expected value of random variables. For example, if there are two agents, Alice and Bob, it cannot be common knowledge that the expected value of a random variable X is $1/2$ according to Alice and $2/3$ according to Bob. This property can be expressed in the language $\mathcal{L}_2^{KQUC}$. Consider the following axiom:

CP$_2$. If $\varphi_1, \ldots, \varphi_m$ are *mutually exclusive* formulas (that is, if $\neg(\varphi_i \wedge \varphi_j)$ is an instance of a propositional tautology for $i \neq j$), then

$$\neg C(a_1\ell_1(\varphi_1) + \cdots + a_m\ell_1(\varphi_m) > 0 \wedge a_1\ell_2(\varphi_1) + \cdots + a_m\ell_2(\varphi_m) < 0).$$

Notice that $a_1\ell_1(\varphi_1) + \cdots + a_m\ell_1(\varphi_m)$ is the expected value according to agent 1 of a random variable that takes on the value a_i in the worlds where φ_i is true, while $a_1\ell_2(\varphi_1) + \cdots + a_m\ell_2(\varphi_m)$ is the expected value of the same random variable according to agent 2. Thus, CP$_2$ says that it cannot be common knowledge that the expected value of this random variable according to agent 1 is positive while the expected value according to agent 2 is negative.

It can be shown that CP_2 is valid in structures in $\mathcal{M}_2^{K,meas}$ satisfying CP; moreover, there is a natural generalization CP_n that is valid in structures $\mathcal{M}_n^{K,meas}$ satisfying CP (Exercise 8.13). However, these axioms (together with standard axioms for reasoning about common knowledge) are still not quite enough to get completeness. It turns out that we need to strengthen them slightly, although the details are beyond the scope of this book.

8.3 Knowledge and Belief

Philosophers have long discussed the relationship between knowledge and belief. To distinguish them, I use the modal operator K for knowledge and B for belief (or K_i and B_i if there are many agents). Does knowledge entail belief; that is, does $K\varphi \Rightarrow B\varphi$ hold? (This has been called the *entailment* property.) Do agents know their beliefs; that is do $B\varphi \Rightarrow KB\varphi$ and $\neg B\varphi \Rightarrow K\neg B\varphi$ hold? Are agents introspective with regard to their beliefs; that is, do $B\varphi \Rightarrow BB\varphi$ and $\neg B\varphi \Rightarrow B\neg B\varphi$ hold? While it is beyond the scope of this book to go into the philosophical problems, it is interesting to see how notions like CONS, SDP, and UNIF, as defined in the previous section, can help illuminate them.

To begin with, we must decide how to model belief. We have actually seen two approaches. Chapter **??** discussed the use of accessibility relations $\mathcal{K}_i$ for modeling knowledge and belief. To reason about both simultaneously, Kripke structures for knowledge and belief of the form $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{B}_1, \ldots, \mathcal{B}_n, \pi)$ can be used, where $\mathcal{K}_1, \ldots, \mathcal{K}_n$ are used to capture knowledge and $\mathcal{B}_1, \ldots, \mathcal{B}_n$ are used to capture belief. Let $\mathcal{L}_n^{KB}$ be the language with modal operators $K_1, \ldots, K_n$ for knowledge and $B_1, \ldots, B_n$ for belief. As expected, the semantics for $B_i\varphi$ is

$$(M, w) \models B_i\varphi \text{ iff } (M, w') \models \varphi \text{ for all } w' \in \mathcal{B}_i(w).$$

(The semantics of knowledge remains unchanged: $(M, w) \models K_i\varphi$ iff $(M, w') \models \varphi$ for all $w' \in \mathcal{K}_i(w)$.)

Another approach is to use plausibility measures to capture belief, as discussed in Section 7.2: an agent believes φ if φ is more plausible than $\neg\varphi$. Formally, given a structure $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PL}_1, \ldots, \mathcal{PL}_n, \pi)$ for knowledge and plausibility, define

$$(M, w) \models B_i\varphi \text{ iff } \mathrm{Pl}_{w,i}([\![\varphi]\!]_M \cap W_{w,i}) > \mathrm{Pl}_{w,i}([\![\neg\varphi]\!]_M \cap W_{w,i}),$$
$$\text{where } \mathcal{PL}_i(w) = (W_{w,i}, \mathrm{Pl}_{w,i}).$$

According to this definition, $B_i\varphi$ can be viewed as an abbreviation for *true* $\rightarrow_i \varphi$. This abbreviation suggests that $B_i\varphi$ can be interpreted as "φ is typically true". Belief is typically assumed to be closed under conjunction, that is, $(B_i\varphi \wedge B_i\psi) \Leftrightarrow B_i(\varphi \wedge \psi)$ is assumed to hold. By the results of Section 7.2, this equivalence holds for qualitative plausibility measures. Thus, when dealing with belief, I restrict attention to structure for knowledge and *qualitative* plausibility, where are the plausibility measures that arise are qualitative.

The definition of belief in terms of plausibility is equivalent to that in terms of the binary relation $\mathcal{B}_i$ as long as the set W of possible worlds is finite. That is, if W is finite, a structure for knowledge and belief can be transformed into a structure for knowledge and qualitative plausibility satisfying the same formulas, and vice versa (Exercise 8.14). However, as shown in Chapter **??**, plausibility meaures are more expressive if W is infinite. Moreover, plausibility can be used to express conditional belief statements $(\psi \rightarrow_i \varphi)$ as well as unconditional belief statements (*true* $\rightarrow_i \varphi$); this added expressive power is useful for dealing with belief revision, as we shall see in Chapter **??**. Thus, I use (qualitative) plausibility to model belief from here on in.

Analogues of CONS, UNIF, SDP, and CP can be defined in structures for knowledge and plausibility: simply replace $\mathcal{PR}_i$ by $\mathcal{PL}_i$ throughout. Interestingly, these properties are closely related to some of the issues regarding the relationship between knowledge and belief, as the following proposition shows.

Proposition 8.3.1 *Let* $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PL}_1, \ldots, \mathcal{PL}_n)$ *be a structure for knowledge and qualitative plausibility.*

(a) If M *satisfies CONS, then* $M \models K_i\varphi \Rightarrow (\psi \rightarrow_i \varphi)$ *for all formulas* φ *and* ψ*; in particular* $M \models K_i\varphi \Rightarrow B_i\varphi$ *for all* φ*.*

(b) If M *satisfies SDP, then* $M \models (\psi \rightarrow_i \varphi) \Rightarrow K_i(\psi \rightarrow_i \varphi)$ *and* $M \models \neg(\psi_i \rightarrow_i \varphi) \Rightarrow K_i\neg(\psi_i \rightarrow_i \varphi)$ *for all formulas* φ *and* ψ*; in particular,* $M \models B_i\varphi \Rightarrow K_iB_i\varphi$ *and* $M \models \neg B_i\varphi \Rightarrow K_i\neg B_i\varphi$ *for all formulas* φ*.*

(c) If M *satisfies UNIF, then* $M \models (\psi \rightarrow_i \varphi) \Rightarrow B_i(\psi \rightarrow_i \varphi)$ *and* $M \models \neg(\psi_i \rightarrow_i \varphi) \Rightarrow B_i\neg(\psi_i \rightarrow_i \varphi)$ *for all formulas* φ *and* ψ*; in particular,* $M \models B_i\varphi \Rightarrow B_iB_i\varphi$ *and* $M \models \neg B_i\varphi \Rightarrow B_i\neg B_i\varphi$ *for all formulas* φ*.*

Proof See Exercise 8.15. ∎

Thus, CONS gives the entailment property, with SDP, agents know their beliefs, and with UNIF, agents are introspective regarding their beliefs.

Exercises

8.1 Show that $M^2 \models K_B(\ell_B(c_0) = 0 \vee \ell_B(c_0) = 1)$ and $M^2 \models K_B((c_0 \Rightarrow \ell_B(c_0) = 1) \wedge (c_1 \Rightarrow \ell_B(c_1) = 1)$.

8.2 Show that $M^3 \models K_B(\ell_B(a) = 0 \vee \ell_B(a) = 1)$.

8.3 Show that $M^4 \models K_B(\ell_B(h) = 1/2)$ while $M^5 \models K_B(\ell_B(h) = 0 \vee \ell_B(h) = 1)$.

8.4 Show that

$$\begin{array}{ll} M^6 \models & K_B(\ell_B(o_{\leq 3}) = 1/2 \vee \ell_B(o_{\leq 3}) = 7/10) \wedge \\ & K_B(K_A(\ell_A(o_{\leq 3}) = 6/7) \vee K_A(\ell_A(o_{\leq 3}) = 2/3) \vee K_A(\ell_A(o_{\leq 3}) = 1/3)). \end{array}$$

* **8.5** Let M^7 be the result of replacing $\mathcal{PR}_B^B$ in M^6 by $\mathcal{PR}_B^A$.

(a) Show that

$$\begin{array}{ll} M^7 \models & K_B(\ell_B(o_{\leq 3}) = 6/7 \vee \ell_B(o_{\leq 3}) = 2/3 \vee \ell(o_{\leq 3}) = 1/3) \wedge \\ & K_B(K_A(\ell_A(o_{\leq 3}) = 6/7) \vee K_A(\ell_A(o_{\leq 3}) = 2/3) \vee K_A(\ell_A(o_{\leq 3}) = 1/3)). \end{array}$$

Suppose that Alice has a certain *strategy* for deciding what bet to offer Bob. Formally, a *strategy* for Alice in this case is a function from worlds to the payoff β that Alice offers for bets on $o_{\leq 3}$. Assume that Alice's strategy depends only on her information, so that if $(w, w') \in \mathcal{K}_A$, then Alice must offer the same payoff at both w and w'. Similarly, a strategy for Bob is a function from worlds to rules for accepting/rejecting offers to bet that is the same at all worlds that Bob cannot distinguish. For example, at a given world w, Alice's strategy may be to offer Bob a bet on $o_{\leq 3}$ with a payoff of \$2 if $o_{\leq 3}$ is true, while Bob's strategy may be to reject all offers with payoff less than \$2.50 (so, in particular, Bob will reject this offer).

(b) Show that if Bob uses the probabilities computed according to M^7 as a guide, then he will not lose money. That is, if Bob accepts bets only if the payoff is at least \$6 (in which case he is guaranteed to at least break even even if the probability of $o_{\leq 3}$ is 6/7), then he does not lose money no matter what strategy Alice uses.

(c) Show that Bob cannot do better than this, in that if if he uses any other threshold for accepting bets, then there is a strategy that Alice could use that would guarantee that Bob would lose money.

8.6 Show that if M satisfies UNIF and CONS, then for each world w, we can partition $\mathcal{K}_i(w)$ into subsets $W_1, \ldots, W_k$ such that if $u, v \in W_j$, then $\mathcal{PR}_i(u) = \mathcal{PR}_i(v)$, and if $\mathcal{PR}_i(u) = (W', \mathcal{F}_{u,i}, \mu_{u,i})$, then $W' \subseteq W_j$.

8.7 Show that CONS and SDP together imply UNIF, but that SDP by itself does not imply UNIF. (For the second half, describe a structure that satisfies SDP and not UNIF.)

8.8 Given the definition of W, $\mathcal{K}_1$, and $\mathcal{K}_2$, show that the only way the structure M described just before Example 8.2.2 can be consisitent with CP is if $\mathcal{PR}_1(w_1) = \mathcal{PR}_2(w_1) = \mathcal{PR}_1(w_1) = \mathcal{PR}_2(w_2)$.

8.9 Suppose that φ is an i-likelihood formula. Show that $\neg\varphi \Rightarrow K_i\neg\varphi$ is provable from KP2 and S5$_n$.

8.10 Show that Axioms KP1 and KP2 (together with Prop and MP) imply KP3.

8.11 Show that the frame F in Example 8.2.2 does not satisfy CP.

8.12 Show that KP1, KP2, and KP3 are valid in structures in $\mathcal{M}_n^{K,meas}$ that satisfy CONS, SDP, and UNIF, respectively.

* **8.13** (a) Show that CP$_2$ is valid in structures in $\mathcal{M}_2^{K,meas}$ satisfying CP.

(b) Consider the following axiom:

CP$_n$. If $\varphi_1, \ldots, \varphi_m$ are mutually exclusive formulas and a_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, are rational numbers such that $\sum_{i=1}^n a_{ij} = 0$, for $j = 1, \ldots, m$, then

$$\neg C(a_{11}\ell_1(\varphi_1) + \cdots + a_{1m}\ell_1(\varphi_m) > 0 \wedge \ldots \wedge a_{n1}\ell_n(\varphi_1) + \cdots + a_{nm}\ell_n(\varphi_m) > 0).$$

(i) Show that CP$_2$ is equivalent to the axiom that results from CP$_n$ above when $n = 2$. (This justifies using the same name for both.)

(ii) Show that CP$_n$ is valid in structures in $\mathcal{M}_n^{K,meas}$ satisfying CP.

* **8.14** (a) Given a structure for knowledge and belief $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{B}_1, \ldots, \mathcal{B}_n, \pi)$, define plausibility assignments $\mathcal{PL}_1, \ldots, \mathcal{PL}_n$ by taking $\mathcal{PL}_i(w) = (W, \mathrm{Pl}_{w,i})$, where

$$\mathrm{Pl}_{w,i}(U) = \begin{cases} 1 & \text{if } U \supseteq \mathcal{B}_i(w) \\ 0 & \text{otherwise.} \end{cases}$$

Let $M' = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PL}_1, \ldots, \mathcal{PL}_n, \pi)$. Show that M' is a structure for knowledge and qualitative plausibility and that M and M' agree on all formulas in $\mathcal{L}^{KB}$, that is, if $\varphi \in \mathcal{L}_n^{KB}$ then, for all $w \in W$,

$$(M, w) \models \varphi \text{ iff } (M', w) \models \varphi.$$

(b) Given a structure $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PL}_1, \ldots, \mathcal{PL}_n, \pi)$ for knowledge and qualitative plausibility, where W is finite, define a binary relation $\mathcal{B}_i$ by setting $\mathcal{B}_i(w) = \cap\{U : \mathrm{Pl}_{w,i}(U) > \mathrm{Pl}_{w,i}(\overline{U})\}$. Let $M' = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{B}_1, \ldots, \mathcal{B}_n, \pi)$. Show that M and M' agree on all formulas in $\mathcal{L}^{KB}$, that is, if $\varphi \in \mathcal{L}_n^{KB}$ then, for all $w \in W$,

$$(M, w) \models \varphi \text{ iff } (M', w) \models \varphi.$$

(c) Show that the construction in part (b) does not work if W is infinite, by constructing a structure M for knowledge and qualitative plausibility for which the set W of possible worlds is infinite and the structure M' for knowledge and belief does not agree with M on all formulas in $\mathcal{L}^{KB}$.

8.15 Prove Proposition 8.3.1.

8.16 State analogues of CONS, SDP, and UNIF in the case where a binary relation $\mathcal{B}_i$ is used to model belief, and prove an analogue of Proposition 8.3.1 for your definition.

* **8.17** Another property of interest relating knowledge and belief is called *certainty*. It is characterized by the following two axioms:

$$\begin{array}{ll} B\varphi \Rightarrow BK\varphi & \text{(positive certainty)} \\ \neg B\varphi \Rightarrow B\neg K\varphi & \text{(negative certainty).} \end{array}$$

(a) Show that if B satisfies the axioms of KD45, K satisfies the axioms of S5, and the entailment property holds, then negative certainty follows from positive certainty.

(b) Show that if B satisfies the axioms of KD45, K satisfies the axioms of S5, the entailment property holds, and positive introspection holds, then B is equivalent to K; that is $B\varphi \Leftrightarrow K\varphi$ is provable. Thus, under these assumptions, an agent cannot hold false beliefs: $\neg\varphi \wedge B\varphi$ is inconsistent. (This result holds even if B does not satisfy the introspective properties K4 and K5.)

Notes

The logic of probability and knowledge considered here was introduced in [Fagin and Halpern 1994]. Although this was the first paper to consider a combined logic of probability and knowledge, combinations of probability with other modal operators had been studied earlier. Propositional probabilistic variants of *temporal logic* (a logic for reasoning about time—see Chapter 8) were considered by Hart and Sharir [1984] and Lehmann and Shelah [1982], while probabilistic variants of *dynamic logic* (a logic for reasoning about actions) were studied by Feldman [1984] and Kozen [1985]. Monderer and Samet [1989] also consider a semantic model that allows the combination of probability and knowledge, although they did not introduce a formal logic for reasoning about them.

The idea of viewing different probability assignments as playing against different adversaries is explored in detail in [Halpern and Tuttle 1993]. The notions of CONS, SDP, UNIF were formalized in [Fagin and Halpern 1994]; Theorems 8.2.1 and 8.2.3 are taken from there.

The common prior assumptions and its implications have been well studied in the economics literature; some significant references include [Aumann 1976; Harsanyi 1968; Morris 1995]. The fact that CP implies no disagreement in expectation (and, in a sense, can be characterized by this property) was observed by Bonanno and Nehring [1996], Feinberg [1995, 1996], Morris [1994], and Samet [pear]. The axiom CP_2 (and CP_n in Exercise 8.13) is taken from [Feinberg 1996; Samet pear]. An axiomatization of $\mathcal{L}_n^{KQUC}$ in the presence of CP can be found in [Halpern 1998a], from where Theorem 8.2.4 is also taken.

There has been a great deal of work on logics of knowledge and belief; see, for example, [**?**; Hoek 1993; Kraus and Lehmann 1988; Lamarre and Shoham 1994; Lenzen 1978; Lenzen 1979; Moses and Shoham 1993; Voorbraak 1991]. The use of plausibility to model belief is discussed in [Friedman and Halpern 1994], from where Proposition 8.3.1 and Exercise 8.14 are taken. The observation in Exercise 8.17 is due to Lenzen [1978, 1979]; see [**?**] for further discussion of this issue.