CHAPTER 5

Rationality and knowledge in game theory

Eddie Dekel and Faruk Gul

1 INTRODUCTION

The concepts of knowledge and rationality have been explicitly applied by economists and game theorists to obtain no-trade results and to characterize solution concepts of games. Implicitly these two concepts underlie much recent work in these fields, ranging from information economics through refinements to attempts to model bounded rationality. Our discussion of the game theoretic and economic literatures on knowledge and rationality will focus on these foundational issues and on the characterizations of solution concepts. Our discussion builds on Harsanyi's (1967) foundation for games of incomplete information and Aumann's (1976) model of common knowledge, on the one hand, and, on the other hand, Bernheim's (1984) and Pearce's (1984) characterization result that, if rationality is common knowledge, then in normal-form games players will choose strategies that survive iterated deletion of strongly dominated strategies.[1]

We begin in section 2 by responding to recent explicit and implicit criticisms of this research agenda. These criticisms are based on the idea that various paradoxes demonstrate that common knowledge of rationality is a problematic notion. We argue instead that these paradoxes are the result of confusing the conclusions that arise from (equilibrium) analysis with assumptions about behavior.[2] A simple example concerns the paradox of cooperation in the Prisoners' Dilemma. Clearly the outcome of the PD will be symmetric. Thus, some have argued, the non-cooperative outcome cannot be the consequence of common knowledge of rationality since, by symmetry, if a player chooses to cooperate so will her opponent. This confuses the conclusion that equilibrium play is symmetric, with the presumption that when a player considers deviating from equilibrium she can assume that all others will symmetrically deviate with her.[3] The section

examines this and other paradoxes in more detail. We do not intend to argue that rationality is an uncontroversial assumption, and we agree that an important agenda is to understand its role better and to consider alternative assumptions. Nevertheless, we do claim that rationality is a fruitful assumption and that, in fact, it sheds much light on these paradoxes.

The second fundamental issue we examine is the underpinnings for the basic model of asymmetric information in economics and game theory. (This is the topic of section 3; this section is rather abstract and, except for subsection 3.1 and 3.3.1, can be skipped with (almost) no loss of continuity.) The very definition of common knowledge, and hence its applications to characterizations of solution concepts, requires that there is a model of the environment, including a state space and information partitions, that is common knowledge among the players. We ask what are the justifications for this assumption and what are the limitations of these justifications. We address these questions using both the Bayesian model familiar to economists, and the syntactic models which have been recently introduced into economics; we also examine the connection between the syntactic view of knowledge as information and the Bayesian notion of knowledge, belief with probability 1, which we call certainty. One limitation we point out is that the justifications for assuming a commonly known model do not imply that the assumption that players have a prior belief on the state space is warranted. Thus, we raise a concern with notions such as *ex ante* efficiency and the common prior assumption (CPA), which rely on the existence of such a prior.

Sections 4 through 6.2 are the heart of the chapter; here we characterize solution concepts, using the model of knowledge and certainty developed in section 3, in terms of assumptions concerning the players' knowledge/certainty about one another and their rationality. We begin in section 4 by appending normal-form games to the model of knowledge and certainty and we review both the equivalence between common knowledge of rationality and iterated deletion of strongly dominated strategies, and characterizations of other equilibrium concepts. Due to our concerns about the CPA and other assumptions, and their necessity in some characterizations, we conclude with dissatisfaction with these epistemic justification of certain solution concepts, such as Nash equilibrium.

Section 5 examines extensive-form games. Several papers have obtained different, apparently contradictory, conclusions concerning the implications of assuming common knowledge/certainty of rationality in extensive-form games. These conclusions include that standard models are incomplete (Binmore (1987–8), Samet (1993), Stalnaker (1994)), that common knowledge of rationality is problematic or inconsistent in extensive-form games (Basu (1990), Bicchieri (1989), Bonanno (1991)), that backwards induction is implied by common knowledge of rationality (Aumann

(1995a)), and that backwards induction is not implied by common certainty of rationality (Reny (1993), Ben Porath (1994)). We show that the model of knowledge and belief from section 3 *is* sufficient to provide a unified framework for examining these results, and we present several characterizations. These characterizations shed light on the elusive issue of what are the implications of common knowledge and certainty of rationality in extensive-form games. We feel that the most natural assumption is common certainty of rationality; this assumption characterizes the solution obtained by applying one round of deletion of weakly dominated strategies and then iterated deletion of strongly dominated strategies, which we call rationalizability with caution (see Ben Porath (1994) and Gul (1995b)).

Section 6 examines what happens if common certainty is weakened to various notions of almost common certainty. Not only does this allow us to investigate the robustness of different solution concepts and characterizations when replacing common certainty with these various notions, but it also enables us to characterize some refinements. In section 4 we observe that refinements and common certainty are inconsistent. In particular, this suggests that the idea that iterated deletion of weakly dominated strategies follows from some basic premise concerning caution and common knowledge (see, e.g., Kohlberg and Mertens (1985)) is flawed. In section 6.2 we show that the closest result one can get is that almost common certainty of caution and rationality characterizes rationalizability with caution – the same solution concept as results from common certainty of rationality in extensive-form games.

Section 6.3 uses the notion of almost common knowledge to raise concerns about refinements of Nash equilibrium, and to obtain new "refinements." We consider two notions of robustness for solution concepts. In the spirit of Fudenberg, Kreps, and Levine (1988), we first consider the following requirement. A solution concept is robust if, given a game G, the solution of G does not exclude any outcomes that it would accept if applied to *some* game in which G is almost common certainty. Rationalizability with caution is the tightest refinement of iterated deletion of strongly dominated strategies that is robust in this sense. Kajii and Morris (1995) and Monderer and Samet (1989) investigate a related notion of robustness: a solution concept is robust in their sense if any predicted outcome of G is a prediction in *all* games where G is almost common certainty. Different notions of almost common certainty lead to different conclusions concerning which solution concepts are robust. Monderer and Samet (1989) show that ε-Nash equilibrium is robust using a strong notion of almost common certainty; Kajii and Morris (1995) use a weaker notion and show that the only standard solution concept that is robust in their sense is that of a unique correlated equilibrium.

Section 7 considers models of asymmetric information where the information structure need not take the form of a partition. The connection between this and weakening the notion of knowledge is discussed, and the implications of these weakenings are explored. The section concludes with a discussion of the problems of this literature.

We need to make one final organizational point. The discussion of various interesting issues, that are related to our presentation, would be disruptive if included within the main body of the text. In addition to simply ignoring many issues, we adopt a non-standard use of footnotes to deal with this problem: we include formal statements and proof sketches within some footnotes.

Many disclaimers are appropriate; the following three are necessary. First, we describe ourselves as presenting and showing various results, but it should be understood that most of the chapter reviews existing work and these terms do not imply any originality. Second, while we make sure to reference the source of all results, and attempt to mention and cite most related research, for brevity we tend to cite the relevant work once rather than on every occasion. Finally, studying a large number of concepts and theorems formulated in different settings within a single framework, as we do here, has its costs. We cannot expect to do full justice to original arguments or hope to convey the full strength of the authors' insights. Thus, we do not expect to provide a perfect substitute for the authors' original treatment of the issues discussed below. However, a unified treatment and the comparisons it enables has benefits that hopefully will offset the inevitable loss such a treatment entails in the analysis of each individual theorem.

2 A RATIONAL VIEW OF SOME PARADOXES OF RATIONALITY

The purpose of this section is to provide a single "explanation" of some familiar and some new paradoxes of rationality. We begin with an informal review of the paradoxes, and then offer our resolution.

2.1 The paradoxes

2.1.1 The Prisoners' Dilemma

The term paradox refers either to a logical inconsistency or a counter-intuitive conclusion. For most game theorists and economists the Prisoners' Dilemma poses neither of these. Instead, it offers a simple and valid insight, perhaps the most basic insight of game theory; the conflict

between individual and group incentives and the resulting inefficiency. For non-game theorists, the Prisoners' Dilemma is apparently much more problematic (see Campbell and Sowden (1985)) and thus it serves as the ideal starting point for our analysis. The argument is the following. The obvious symmetry of the problem faced by the two agents is sufficient for anyone analyzing the problem (including the players themselves) to conclude that both agents will take the same action. Hence, player 1 knows that the outcome will be either cooperate–cooperate or defect–defect. It follows that if player 1 cooperates then cooperate–cooperate will be the outcome whereas if he defects then defect–defect will be the outcome. Since the former yields a higher payoff than the latter, rationality should lead player 1 to cooperate. This conflicts with the obvious dominance argument in favor of defecting. Most economists not working on epistemic foundations of rationality will probably dismiss the above argument for cooperating by saying that the assertion that both agents will take the "same" action is true only in *equilibrium* which according to the dominance argument specifies that both agents will defect. If player one chooses to cooperate (or contemplates cooperation) this is a deviation and hence the equilibrium hypothesis, that both agents will take the "same" action, is no longer valid.[4]

2.1.2 *Newcombe's Paradox*

Closely related to the preceding discussion is the well-known Newcombe's Paradox. Suppose that a person is faced with two boxes: box A contains $1,000 and box B contains either zero or one million dollars. The person can choose either box B or both boxes. The prizes are placed by a genie who has profound insight into the psyche of the person and thus knows whether the person will choose both boxes or just one. If the person is to choose both boxes than the genie will put zero dollars into box B. If the person is to choose only box B, then the genie will put one million dollars into box B. By the time the person makes a choice he knows the genie has already made his decision as to how much money should go into box B. Thus, as in the above analysis of the Prisoners' Dilemma, a simple dominance argument suggests that the person should take both boxes. However, the infallibility of the genie suggests that the decision to choose box B alone yields one million dollars while the decision to choose both yields $1,000. Hence the person should choose box B alone.[5]

2.1.3 *The paradox of backward induction*

In the three stage take-it-or-leave-it game (see figure 5.1), the finitely repeated Prisoners' Dilemma, and other similar games, the apparently

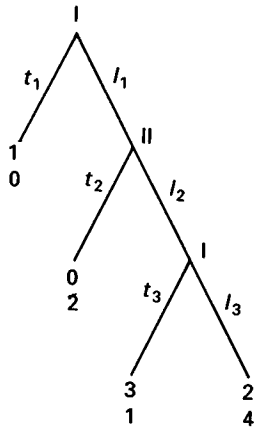


Figure 5.1 The three-stage take-it-or-leave-it game

compelling logic of backward induction suggests a solution that appears unintuitive. Yet, until recently, the strategy of always defecting in the repeated Prisoners' Dilemma was viewed to be the only "correct" solution that intuitively satisfies common knowledge of rationality. The issue of extensive-form games is discussed in more depth in section 5, where we show that a formal model of common knowledge of rationality does not yield the backward-induction outcome.

2.1.4 Bonanno's Paradox

In a recent paper (discussed further in section 5 below), Bonanno (1991) provides a more precise statement of the following paradox: Let R be the set of all propositions of the form

$$\{((P_\alpha) \to (\pi \geq x)) \text{ and } ((P_\beta) \to (\pi \leq y)) \text{ and } (x > y)\} \to \neg(P_\beta),$$

where objects in parenthesis correspond to propositions as follows: (P_α) is the proposition that "the agent chooses α"; $(\pi \geq x)$ is, "the agent receives utility no less than x;" (P_β) and $(\pi \leq y)$ are defined in an analogous manner; and, finally, $(x > y)$ is the proposition "x is strictly greater than y." Thus R is the proposition that an agent faced with these choice of α or β is rational. Suppose, for instance, that the agent faces a choice between a and b where a will yield 100 dollars and b will yield zero dollars. Thus, we have (P_a or P_b) and $\neg$(P_a and P_b). Suppose we postulate R to capture the hypothesis that the agent is rational. Then we conclude from the parameters above and the assumption of rationality that the agent does not choose b. It follows that

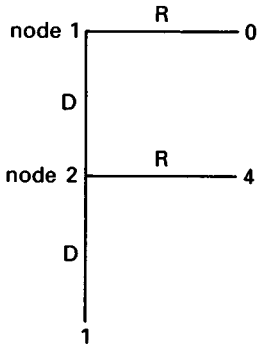


Figure 5.2 The game of the absent-minded driver

the proposition $(P_b \rightarrow (\pi \geq 1000))$ is true by virtue of the fact that (P_b) is false. Then applying R again to $(P_b \rightarrow (\pi \geq 1000))$ and $(P_a \rightarrow (\pi \leq 100))$ yields $\neg(P_a)$ which, together with $\neg(P_b)$, contradicts the fact that the agent had to choose either a or b. Thus rationality is impossible in this formulation.

2.1.5 *Piccione and Rubinstein's Paradox*

Consider the problem described in figure 5.2. A person with imperfect recall is faced with the choice of choosing R or D. If he turns out to be in node 1, R will have very unpleasant consequences. Choosing D at node 1 leads to node 2 which, owing to imperfect recall, is a situation that the decisionmaker considers indistinguishable from node 1. Choosing R at node 2 yields the most desirable outcome while choosing D a second time yields an intermediate outcome. The optimization problem faced by the decisionmaker is the following. He must choose a probability p with which to play the action D (and hence R is played with probability $1 - p$). By imperfect recall, the same p must be used at each information set. Hence the decisionmaker must maximize $4p(1 - p) + p^2$, yielding $p = 2/3$. The first observation here is that the optimal strategy is unique but not a pure strategy. Second, and this is what Piccione and Rubinstein (1995) consider paradoxical, if the decisionmaker were Bayesian and assign a probability α to the event of being at node 1, then his decision problem would be to maximize $\alpha[4p(1 - p) + p^2] + (1 - \alpha)[p + 4(1 - p)]$. It is easy to verify that, for any value of α other than 1, the solution of the second optimization yields an answer different from 2/3. Thus, we are either forced to insist that a Bayesian agent facing this problem must always assign probability 1 to being at node 1, or we must accept that the *ex ante*

solution differs from the Bayesian solution. That is, if we want to allow the probability of being at node 1 to be less than 1, we must accept the dynamic inconsistency.

2.2 The resolution

The purpose of this section is to isolate a "common cause" for each of these paradoxes. Whether this counts as a "resolution" is not at all clear. If by paradox we mean unexpected or surprising result then we may continue to believe that the results are surprising even if we agree as to what the main cause of the paradox is. On the other hand, if by a paradox we mean a logical inconsistency then understanding the cause certainly does not remove the original inconsistency. Of course, we hope that the reader will agree that the same "cause" underlies all of these paradoxes and, having identified this cause, will find the paradox less surprising or the inconsistency less troublesome as a comment on rationality.

Let us start with the Prisoners' Dilemma which is likely to be the least controversial for game theorists and economists. Most game theorists would agree that the rational course of action in the Prisoners' Dilemma is to defect. But then what happens to the claim that the two agents will choose the same action? This is still satisfied if both agents defect and will be satisfied only if agents behave as they are supposed to, i.e., in equilibrium. That is, we are investigating what outcomes are consistent with the given assumptions: that the players are rational and the outcome is symmetric. The reason that economists and game theorists are not puzzled by the paradox of the Prisoners' Dilemma as perceived by philosophers is that this kind of reasoning is very familiar from (Nash and competitive) equilibrium analysis. Whenever we investigate the implications of a set of assumptions including the assumption that each agent is rational, we are forced to justify the rationality of a given agent by comparing what he expects to receive if he behaves as we predict with what he expects to receive if he behaves otherwise. But when he contemplates behaving otherwise he cannot expect that assumptions made about his own behavior will continue to hold, even if these were very reasonable assumptions to begin with. If we insist that the rational agent will expect assumptions about his own behavior to continue to hold even as he deviates, then we will be confronted with paradoxes.

The application of this idea to the Prisoners' Dilemma is clear. The defense of the cooperate–cooperate outcome relies on showing than defect–defect cannot be the answer since by deviating an agent can conclude (by using the assertion that he and his opponent will take the same course of action) that cooperating leads to cooperate–cooperate which is better that defect–defect. But clearly, in this argument we are using the fact that player

1 knows that player 2 will choose the same course of action even as 1 contemplates deviating. Hence the contradiction.

The analysis of Newcombe's paradox is similar. If we were confronting this genie, then we would surely take both boxes. So, if the genie is as knowledgeable as claimed, he will put nothing in box B. If we were to deviate, then the genie would be wrong (but of course we would not gain from this deviation). If we hypothesize the existence of a genie that is always right, even when somebody deviates, then we will get a contradiction.

The fact that the same factor is behind the backward induction paradox is more difficult to see for three reasons. First, the dynamic nature of the strategic interaction forces us, the analysts, to discuss not only the possibility that a player may contemplate deviating, but also the fact that if he does deviate, then some other player will get a chance to observe this deviation. Hence, we are forced to analyze the deviating player's analysis of some other player's reactions to the deviation. Second, in the backward-induction paradox, unlike the remaining paradoxes discussed in this section, identifying the cause does not immediately suggest an alternative model that is immune to the problem identified. This can be seen from the fact that a number of other game-theoretic solution concepts, such as Nash equilibrium or iterative removal of weakly dominated strategies, yield the same backward-induction outcomes in the well-known examples such as the take-it-or-leave-it game or the repeated chain store paradox or the repeated Prisoners' Dilemma. Nevertheless, identifying the cause is the first step to the more sophisticated non-backward induction theories and the more elaborate arguments for backward induction that will be discussed in section 5. The task of evaluating other concepts, such as Nash equilibrium or iterative weak dominance, need not concern us here. As we will discuss in section 5, the cause of the backward-induction paradox is by now well understood. Implicit in the backward-induction argument is the assumption that, if the second information set were reached in the game depicted in figure 5.1, then player 2 continues to be certain that player 1 is rational even though this is precisely the assumption utilized in concluding that the second information set will not be reached.

Once again, the difficulty stems from insisting on apparently plausible assumptions regarding the behavior of some player even in the face of deviations. (In particular, the assumption that rationality is common knowledge and the conjecture that this implies the backwards-induction solution are upheld in the face of behavior that conflicts with the combination of these two statements.) The added difficulty of the paradox comes into play at this stage: we know where the difficulty is but the resolution – unlike those discussed above – is not agreed upon by game theorists. Some authors have concluded at this stage, that we must give up

on backward induction while others have suggested stronger notions of rationality for extensive-form games. Yet others have argued that, while we must give up on backward induction as a consequence of common knowledge of rationality, alternative plausible assumptions about behavior may (in certain games) yield backward induction. Some of the work along these lines will be discussed in section 5. Our current objective is only to note that the origin of the paradox is the same in all the cases studied in this section.

In Bonanno's Paradox we can see the same effect coming in through the fact that rationality is postulated throughout the model, and not used as a test of outcomes. Thus, even as the agent contemplates (or chooses) the irrational b, the proposition "the agent is rational" is maintained. Hence, by making the irrational choice the agent can create a falsehood. But a false proposition can imply anything, in particular it can imply that the agent receives an infeasible level of utility. Which makes the irrational action very rewarding and yields the contradiction.[6]

The Piccione–Rubinstein Paradox is more subtle but also similar. Consider the calculation that the Bayesian rational person is to undertake: choose p so as to maximize $\alpha[4p(1-p)+p^2]+(1-\alpha)[p+4(1-p)]$. Implicit in this calculation is the assumption that whatever p is chosen today will be implemented tomorrow as well, even if the choice of p reflects some sort of deviation or irrationality. Setting aside the question of whether giving a memoryless agent this implicit ability to recall past actions is a good modeling choice, we note that the absence of perfect recall presumably means that the agent cannot systematically implement different plans at the two nodes. It does not mean that the actions chosen at the two nodes are by *logical* necessity the same. To see this note that if the actions are by logical necessity the same then the agent's decision to change his mind, and choose $p \neq 2/3$, would have no meaning were he, in fact, at the second node. Alternatively put, when the agent makes a choice, since he does not know at which node he is located, his deviation must assume that whatever he planned to do originally – in this case $p = 2/3$ – stills holds elsewhere. As in the other paradoxes above, the analysis of the agent's rationality requires the analyst to assume that everything else that the agent cannot change is held constant.

This resolution, based on dropping the logical equivalence between the agent's choices at both nodes, is discussed by Piccione and Rubinstein (1995) in the section entitled "Multi-selves approaches." The analysis there proceeds as follows. First, they fix p and compute $\alpha(p)$, the long-run relative frequency of visiting the first node for an agent that exits with probability p at each node. Next, they ask what should p be so that $p' = p$ maximizes $p'[\alpha(p)p + 4(1-p) + (1-\alpha(p))] + (1-p')4(1-\alpha(p))$. They show that for

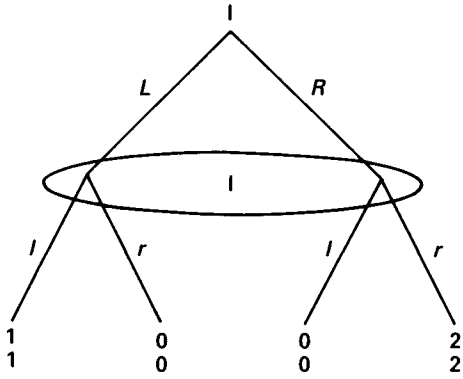


Figure 5.3

$p = 2/3$ – which is the value of p that maximizes expected utility when *ex ante* commitment is possible – any p', in particular $p' = p = 2/3$, is a maximizer. Thus, they observe that having the decisionmaker view his "other" self as a distinct agent resolves the paradox. They find this resolution unconvincing, however, noting that treating a rational decisionmaker as a collection of agents allows for the possibility that (l, L) is a solution to the single-person decision problem described in figure 5.3. They state that this way of dealing with imperfect recall ignores the fundamental difference between single-person and multi-person situations. They argue that it should be possible for a rational decisionmaker to change his strategy, not just at the current information set but also in the future. By contrast, we think that this resolution of their paradox does not require that agents are stuck with arbitrary and inefficient behavior in the future. It is only required that the agent should not assume that changing his mind now (i.e., deviating) will generate a corresponding change later in the *same* information set. To get back to the theme of this section, if at node 2, the agent cannot remember whether he has been to node 1 or not and hence what he has done at node 1, then the requirement that he take the same action at both nodes can hold only if he does what he is supposed to do; not if he deviates. This is not an endorsement of a full-blown, non-cooperative (multi-agent) equilibrium approach to the problem of imperfect recall. We may rule out (l, L) as a possible solution to the problem described in figure 5.3 because we find this to be an unacceptable plan for the agent given that $r, R)$ is also a feasible plan. It does not follow from this that a decisionmaker implementing the plan $p = 2/3$ in figure 5.2, when faced with the actual implementation, can assume that if he were to choose to go down without randomizing, he would end up with a payoff of 1 for sure. Modeling the

problem of imperfect recall in this manner (i.e., by postulating that a deviation at one node guarantees a similar deviation at a subsequent node) causes the paradox noted by Piccione and Rubinstein.

What is worth emphasizing in all of the paradoxes discussed above is that the driving force behind each of them is a statement that most of us would find a priori quite plausible (hence the paradox): it certainly makes sense that both players should take the same action in the Prisoners' Dilemma, especially since the same action is dominant for both of them; in Newcombe's Paradox the agent has a dominant action so it makes sense that a genie would know how she will behave, the simple argument for backward induction is very intuitive and compelling, postulating that a rational agent will choose the higher payoff is certainly not far-fetched, and it is almost by definition that an agent who cannot distinguish between two nodes should not be able to implement different actions at these two nodes. The purpose of this section has been to use an idea very familiar to economists to suggest a way of resolving all of these paradoxes. We certainly do not wish to claim that all logical inconsistencies can be removed in this manner. We only suggest an informal modeling idea that might help in avoiding these and other paradoxes of rationality. We feel that our resolution is implicit in the epistemic models that we discuss in sections 4, 5, and 6.[7]

3 WHAT IS A STATE OF THE WORLD AND IS THE MODEL COMMON KNOWLEDGE?

We outline here the topics of the subsections to follow. In subsection 3.1 we briefly review the basic model of asymmetric information used in economics and game theory, and discuss the connection with a model commonly used in other fields, called a Kripke model. We formally define a knowledge operator, discuss its relation with the partition model of information, and review the formal definition of common knowledge.[8] The main conclusion is that the standard interpretation of the definition of common knowledge implicitly assumes that the model itself is common "knowledge." In some contexts, where knowledge or information arises through a physical process, prior to which agents have identical information, the meaning of this assumption is clear. Moreover, in such contexts it is conceivable that the model (i.e., the description of the information-acquisition process) is commonly "known." In subsection 3.2 we take as a starting point a situation of incomplete information, where players' knowledge and beliefs are already formed and a real *ex ante* situation does not exist. In this case the choice of model to represent these beliefs and knowledge is not clear. Moreover, it is not clear a priori that a model chosen to represent these beliefs can be assumed to be common "knowledge." So in this case such an

assumption needs justification. We provide such a justification by presenting two ways to *construct* a commonly known model, using first a syntactic approach for modeling knowledge, and then Harsanyi's (1967) Bayesian approach, which is based on a probabilistic notion of knowledge that we call certainty. Subsection 3.3 discusses the relationship between knowledge and certainty. Finally, subsection 3.4 discusses the limitations of these two solutions to the problem. While they both construct a commonly known *ex ante* model, the construction does not generate a prior. Hence we argue that *ex ante* notions of efficiency and assumptions such as the common prior assumption, are not sensible in contexts where a real physical *ex ante* stage does not exist, that is, in all situations where Harsanyi's justification of the model of incomplete information is needed.

3.1 The issue

In order to understand the first issue that we will discuss, it is necessary to review the basic definition of knowledge and common knowledge in an environment of asymmetric information. An information structure is a collection $\mathcal{I} \equiv (\Omega, (\mathcal{F}_i, p_i)_{i \in N})$. The finite set Ω is the set of states of the world. Each player $i \in N$, where N is the finite set of players, has a possibility correspondence $\mathcal{F}_i : \Omega \to 2^{\Omega}$, where $\mathcal{F}_i(\omega)$ is the set of states i considers possible when the true state is ω. We abuse notation and also denote by $\mathcal{F}_i$ the set of possible information cells $\{F_i \subset \Omega : F_i = \mathcal{F}_i(\omega) \text{ for some } \omega\}$. The standard interpretation in economics is that when the true state is ω, i is informed that one of the states in $\mathcal{F}_i(\omega)$ occurred. For now $\mathcal{F}_i$ is assumed to be a partition of Ω; justifications for this assumption will be presented below, and weakenings will be presented in section 7. Finally, $p_i \in \Delta(\Omega)$ is a prior over Ω. We will usually assume that each cell in $\mathcal{F}_i$ has strictly positive probability, so that conditional probabilities $p_i(\cdot \mid \mathcal{F}_i(\omega))$ are well defined. In cases where $p_i(\mathcal{F}_i(\omega)) = 0$ for some state ω, we (implicitly) assume that the model is extended to some specification of all the conditional probabilities, $\mathcal{I} = (\Omega, (\mathcal{F}_i, p_i, p_i(\cdot \mid \mathcal{F}_i))_{i \in N})$.

In most economic models, a state ω describes something about the real world, for example, the different states might correspond to different preferences for one of the players. Thus, a model will specify, for each state, the value of the relevant parameters of the real world. For example, i's utility function at ω can be denoted by $\mathbf{u}_i(\omega)$; the event that i's utility function is some particular u_i is denoted by $[u_i] \equiv \{\omega \in \Omega : \mathbf{u}_i(\omega) = u_i\}$. This is clarified further and formalized in subsection 3.2.1 below. A model of asymmetric information, $\mathcal{I}$, combined with such a specification, will be called a Kripke model (see, e.g., Fagin *et al.* (1995, chapter 2.5)). Throughout section 3 the term model refers to a model of asymmetric information, $\mathcal{I}$,

either on its own, or appended with such a specification, i.e., a Kripke model. (The context will clarify the appropriate notion.)

We say that agent i *knows an event* $A \in \Omega$ *at state* ω, if i's information in state ω implies that some state in A must be true: $\mathcal{F}_i(\omega) \subset A$. Therefore, the set of states in which, say, 2 knows A is $K_2(A) \equiv \{\omega: \mathcal{F}_2(\omega) \subset A\}$; similarly the set of states at which 2 knows that i's utility function is u_i is $K_2([u_i])$. So, *at state* ω, 1 knows that 2 knows A if $\mathcal{F}_1(\omega) \subset K_2(A)$. Continuing in this way Aumann (1976) showed that at a state ω the event A is common knowledge – in the sense that all players know it, know they know it, etc. – if and only if there is an event F in the meet of the partitions with $\mathcal{F}_i(\omega) \subset F \subset A$. The meet of a collection of partitions is the finest partition that is a coarsening of all partitions in the collection, denoted by $\wedge_{i \in N} \mathcal{F}_i$. It is easy to see that the meet is a partition that includes all the smallest events that are self evident, where self-evident events are those that are known to have occurred whenever they occur: F is self evident if for all i and all $\omega \in F, \mathcal{F}_i(\omega) \subset F$. (Clearly the union of disjoint self-evident sets is self evident and will not be in the meet; hence the qualification to smallest events.)

To summarize and further develop the above argument, given the possibility correspondences $\mathcal{F}_i$ we have derived operators $K_i: 2^\Omega \to 2^\Omega$ which tell us the set of states at which i knows an event A; $K_i(A) \equiv \{\omega: \mathcal{F}_i(\omega) \subset A\}$. This construction then tells us, for each state of the world ω, what each player knows, what each player knows about what each player knows, etc. This operator satisfies the properties below.

T $K_i(A) \subset A$: if i knows A then A is true;

MC $K_i(A) \cap K_i(B) = K_i(A \cap B)$: knowing A and B is equivalent to knowing A and knowing B;

N $K_i(\Omega) = \Omega$: player i always knows anything that is true in all states of the world;

4 $K_i(A) \subset K_i(K_i(A))$: if i knows A then i knows that i knows A;

5 $\neg K_i(A) \subset K_i(\neg K_i(A))$, where $\neg$ denotes complements: not knowing A implies knowing that A is not known.

These properties can be used to axiomatically characterize knowledge and partitions. That is, instead of starting with partitions $\mathcal{F}_i$, and deriving a knowledge operator K_i which satisfies these properties, we could have started with such an operator K_i and derived the partitions. Formally, given any operator K_i satisfying these properties, one can define a possibility correspondence $\mathcal{F}_i: \Omega \to 2^\Omega$ by $\mathcal{F}_i(\omega) = \cap \{A \subset \Omega: \omega \in K_i(A)\}$, and such a possibility correspondence in turn would generate the same K_i according to the definition above. In fact, the only properties needed for this result are [MC] and [N]. (We allow $\mathcal{F}_i = \emptyset$; if we wanted to rule this out we

would need to add the axiom $K_i(\emptyset) = \emptyset$.) It is straightforward to check that properties [T], [4], and [5] imply that $\mathscr{F}_i$ will be a partition. We can now define $K_M(A)$ to be the event that all i in $M \subset N$ know A, $K_M(A) \equiv \cap_{i \in M} K_i(A)$, and then the event that A is common knowledge is simply $CK(A) \equiv \cap_{n=1}^{\infty} K_N^n(A)$, where K_N^n denotes n iterations of the K_N operator. In his discussion of reachability, Aumann (1976) shows that $CK(A) = \cup \{F: F \in \wedge_{i \in N} \mathscr{F}_i, F \subset A\}$; the interpretation of this result is that A is common knowledge at ω if the member of the meet of the partitions at ω is contained in A.

But, as Aumann (1976) pointed out, for this *interpretation* we must assume that 1 "knows" 2's information partition, since this is needed to interpret the set of states $K_1(K_2(A))$ as the set in which 1 knows that 2 knows A. Moreover, we will need to assume that the partitions are common "knowledge" among the players. We use quotes around the words know and knowledge since it is not the formal term defined earlier; it is a meta notion of knowledge that lies outside our formal model. But the fact that this knowledge is not formally within the model is not the main issue. Our concern is whether it is reasonable to assume that the information structure is common "knowledge," informally or otherwise.

The assumption that the information structure is common "knowledge" is easy to interpret if there is an actual *ex ante* situation and "physical" procedure that leads to asymmetric information. For example, this is the case if the underlying uncertainty Ω is the amount of oil in some tract, and each of two firms is entitled to take one soil sample, and there is a thorough understanding based on published experiments of both the prior likelihood of oil and of the distribution of possible soil samples as a function of the oil in the tract. In contrast, consider the case where the agents already have their perceptions (i.e., knowledge or beliefs) about the world, and there was no *ex ante* commonly known physical environment which generated their perceptions. Following Harsanyi (1967), we call this a situation of incomplete information. Can we model this situation of incomplete information as one that arises from a commonly "known" starting point (i.e., *as if* we are at the interim stage of a commonly "known" physical procedure such as the oil-tract story above)?[9]

Before discussing the way this issue has been addressed, it is worth clarifying why it is an issue at all. Clearly, if the model that we write down is not common "knowledge," then it is not a complete description of the players' views of the world. So, in a basic sense the model is incomplete. While this is a concern, there is a more disturbing issue. If the model is not commonly "known," we can no longer justify solution concepts, or any other conclusions that rely on the hypothesis that some event is common knowledge, such as the no-trade results. How can one interpret an

assumption that rationality, or anything else, is commonly known when there is no commonly "known" model within which to define common knowledge?[10]

Thus, our objective in the next subsection is to start with a description of an arbitrary situation of incomplete information, and develop a model and a state $\omega^* \in \Omega$, such that if the model is commonly "known," the knowledge (according to the K_i operators) of the agents at ω^* coincides with the knowledge contained in the original description of the incomplete-information situation.

3.2 Constructing a model that is commonly "known"

Aumann (1976) argued that, if the model is complete, in that each state ω is a complete description of the state of the world, then the model is common "knowledge" (at least so long as Ω is common knowledge). This is because a complete specification of a state should determine the partitions and beliefs of all the players in that state, so if the set of states is common "knowledge" then the partition cell (and beliefs corresponding to that cell) for each state will be common "knowledge."

While this seems compelling, it does not say, for example, that a complete model exists, nor does it say how, from a complete description of a situation which is not common knowledge, a commonly "known" model can be constructed. Understanding the construction is important, since, if we are going to impose assumptions on the constructed space, we need to know how to interpret these assumptions in terms of our starting point, which is a situation of incomplete information and not a commonly "known" model.

Thus, the first question that we will consider is what is a state of the world; what constitutes a complete description. Naturally, it will include two aspects: what is true about the physical world, and what is true about the epistemic world, i.e., what players know.

A preliminary example Assume that a description of the state of the world included the following: (1) a description of the true physical world, which can be either p or $\neg p$, and (2) what each player knows about p and $\neg p$. In particular, say, a state could be the specification p, player 1 knows p holds, and knows that it is not the case that $\neg p$ holds, and player 2 does not know whether it is p or $\neg p$ that holds, but 2 knows that either p or $\neg p$ holds. If we do not say anything further there are many models which would generate this knowledge. For example, consider figure 5.4, where ovals are 1's partition and rectangles are 2's partition, and ω^* indicates the true state of the world. In model (a), $\Omega = \{\omega^*, \omega\}$, $\mathcal{F}_1 = \{\{\omega^*\}, \{\omega\}\}$, $\mathcal{F}_2 = \{\{\omega^*, \omega\}\}$. In the second model (b), $\Omega = \{\omega^*, \omega, \omega', \omega''\}$, $\mathcal{F}_1 = \{\{\omega^*, \omega'\}, \{\omega, \omega''\}\}$,

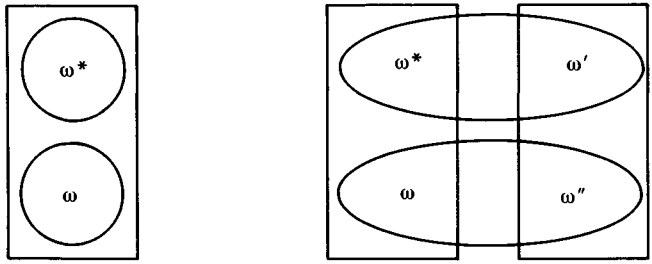


Figure 5.4 Two information structures, which at ω^*, coincide insofar as our partial description of a situation can determine, but differ in terms of the knowledge they imply if they are assumed to be common "knowledge"

$\mathcal{F}_2 = \{\{\omega^*, \omega\}, \{\omega'\}, \{\omega''\}\}$. In both cases to complete the (Kripke) model we need to specify what aspects of the real world are true in the different states: in (*a*), a property p is true in state ω^* and $\neg p$ is true otherwise, while in (*b*), p is true in states ω^* and ω', and $\neg p$ is true otherwise. We can immediately verify that in both of these models $\omega^* \in K_1([p])$ and $\omega^* \notin K_2([p])$, where $[p]$ is the set of states where p is true. Therefore, if we assume only that the players "know" the model then the state ω^* does incorporate the description above.

However, if we assume that these models are common "knowledge," then they would generate many more statements about the players' knowledge than our original description – in particular they clearly model different epistemic situations. In (*a*), in state ω^* player 1 knows that 2 does not know that p is true, while in (*b*) player 1 does not know this. So the assumption that the model is common "knowledge" would be substantive, and we could not use either model to say that given our description of a situation of incomplete information, we can construct a commonly "known" model with a state such that knowledge at that state coincides with the description. How then can we construct an information structure that may be assumed to be common "knowledge" w.l.o.g.?

It is clear what goes wrong in the example: a description needs to be complete and should specify everything possible about the physical state of the world and about the state of mind of the agents – including how they perceive the state of mind of the other agents. Thus a description will be a list of what is true of the physical world and all relevant perceptions of the agents, including perceptions about perceptions. There will also be some natural consistency conditions on states, for example, the description of a state should not say that the physical world has some property and that it does not have that property. More interesting consistency conditions arise

concerning the perceptions of the agents, for example, it seems natural to assume that the agent knows what she knows.

Thus, we are looking for a description of a situation of incomplete information that is a "complete" list of "internally consistent" statements about the agents, perceptions and the physical world. The next step is to show that this description coincides with the knowledge at a state ω^* in some commonly "known" model. This step is accomplished as follows. With the notion of a complete description in hand, we will consider the set of all such descriptions as a universal set of possible states, Ω. Then, we would like to find an information structure on Ω which, if we assume it to be common "knowledge," will generate in the state ω^* the same perceptions as we used in describing the situation of incomplete information.[11] (By generate we mean calculate the agents' perceptions using the information structure as if the structure is common "knowledge" – just like the knowledge operators K_i were derived from the possibility correspondences $\mathcal{P}_i$ above.) Then we can say that taking this information structure to be common knowledge yields nothing more nor less than our description. So, we can work with models that are commonly "known," instead of with complete descriptions.

3.2.1 The syntactic approach

The syntactic approach takes the view that a description of the world should involve a specification of the true state, each person's knowledge about the true state, each person's knowledge about the players' knowledge about the true state, etc.[12] Thus the syntactic approach starts with a countable set of symbols representing the following objects: a set of basic propositions, $X \equiv p, q, \ldots$; a set of players, $i = 1, 2, \ldots n$; conjunction $\wedge$; negation $\neg$; a constant representing truth $\top$; and knowledge of i, k_i.[13]

The set of basic propositions, X, includes, for example, "it is raining," "it is snowing," etc. Thus elements of X are not complete descriptions of the physical world. A complete description of the physical world would be $\{True, False\}^X$ – a specification for each basic proposition whether it is true or false.

The language of sentences, L, is the smallest collection of sequences of these symbols containing X that is closed under negation, conjunction, and knowledge (i.e., if $\phi, \psi \in L$ then $\neg\phi \in L$, $\phi \wedge \psi \in L$, and $k_i\phi \in L$). For example, if $X = \{p\}$ then $\neg p \in L$, as is $k_i \neg k_j p$, where the latter is interpreted as i knows that j does not know p. Since L is the *smallest* such collection, all sentences have finitely many symbols. There is an alternative approach that mirrors our construction in the next subsection; this approach builds up longer and longer sentences starting from the basic propositions, X.

Roughly speaking, the first step has two ingredients: (i) extending the set of sentences from X to the set of all sentences that is closed under conjunction and negation; and (ii) allowing knowledge to operate *once* on elements of this closure of X. Then, inductively, consider the closure under conjunction and negation of the just constructed set of sentences, and add one more level of knowledge. In this way the language would be indexed by the depth of knowledge one wants to consider. (For a precise development, see, for example, Fagin, Halpern, and Vardi (1991).) This approach has the obvious additional advantage of allowing sentences with infinitely many symbols (by extending the depth of knowledge in the sentences transfinitely). However, the notational complexity does not warrant an explicit development, and we will informally discuss sentences with infinitely many symbols where relevant.

The following axioms will be used to impose consistency in the definition of a state.

T $k_i(A) \to A$, where the symbol $\phi \to \psi$ stands for $\neg(\phi \wedge \neg\psi)$: if i knows A then A is true;

MC $k_i(A) \wedge k_i(B) \leftrightarrow k_i(A \wedge B)$: knowing A and B is equivalent to knowing A and knowing B;

N $k_i\top$: agent i knows the truth constant;

4 $k_i(A) \to k_i(k_i(A))$: if i knows A then i knows that i knows A;

5 $\neg k_i(A) \to k_i(\neg k_i(A))$: not knowing A implies knowing that A is not known.

Note the similarity to the assumptions on K_i above. We will see that these assumptions on k_i generate partitions over states, just like the same assumptions on K_i generated partitions $\mathscr{F}_i$ in subsection 3.1. The difference is that K_i is an operator on exogenously given states, *assuming* that there is some commonly "known" model. On the other hand, k_i is a primitive symbol for describing the players' knowledge in some situation of incomplete information; we will now show how it is used to *construct* a commonly "known" model of a set of states and partitions.

In constructing the commonly "known" model the first step is to define the set of states as all complete and consistent descriptions within our language. To formalize this, a useful notion is that of a theorem: these are all sentences that are true in every state. Formally, these include $\top$ and all sentences that can be derived from the five axioms above using two rules of inference: [MP] if ϕ and $\phi \to \psi$ are theorems then so is ψ; and [RE] if $\phi \leftrightarrow \psi$ is a theorem then so is $k_i\phi \leftrightarrow k_i\psi$. A complete and consistent description of a state of the world is a subset of sentences in L that includes all theorems, includes ϕ if and only if it does not include $\neg\phi$, and is closed under the usual rule of logic, [MP], namely if the formulas ϕ and $\phi \to \psi$ are in the

state then so is ψ. A state of the world is thus a list of sentences that are interpreted as true in that state, where the sentences fall into two categories: sentences concerning only basic propositions in X and epistemic sentences concerning knowledge (i.e., involving k_i). Such states are artificial constructs describing imaginable situations of incomplete information.[14] How then do we construct a commonly "known" model? Consider the set of all states of the world as just constructed. In each state ω we can identify the set of sentences that each individual knows: $\mathbf{k}_i(\omega) = \{\phi: k_i\phi \in \omega\}$. The natural information structure is that at each state ω, i cannot distinguish between states where his knowledge is the same, so $\mathcal{F}_i(\omega') = \{\omega: \mathbf{k}_i(\omega') = \mathbf{k}_i(\omega)\}$. Thus far, we have constructed a standard model of asymmetric information: we have a state space Ω and partitions $\mathcal{F}_i$.[15] Finally, we associate with each state in the model the basic propositions which are true in that state. Thus, in addition to a standard model of asymmetric information, this construction yields a function specifying for each state which propositions in X are true.[16] In this constructed Kripke model we ignore the specification of which epistemic sentences are true.

There is a formal equivalence between the Kripke models that we have just constructed, and the complete description of states in our language. The main point is that we can forget about the complete description of the state and derive from the model – which includes only a set of points Ω, partitions of Ω, and a list of those *basic propositions in X* that are true at each $\omega \in \Omega$ – what each player knows, what each player knows about what each one knows, etc., at a particular state of the world, using the assumption that the model is common "knowledge." The sentences we derive in this way will be exactly those in the complete description of the state. Formally, let $[\psi]$ be the set of states at which ψ is true, $[\psi] = \{\omega: \psi \in \omega\}$; the result is that ψ is known according to the language – i.e., the sentence $k_i(\psi)$ is part of the description of ω, $k_i(\psi) \in \omega$ – if and only if ψ is known according to the model – $\omega \in K_i([\psi])$.[17] Hence we have constructed a commonly "known" model from a complete description of a situation of incomplete information.[18]

The construction and results above suggest that one can work with standard partition models, define knowledge in the standard way – i knows A at ω if $\mathcal{F}_i(\omega) \subset A$ – and assume the partition and state space is informally common "knowledge" w.l.o.g. However, this is not precisely the case. There are two issues concerning the richness of the language L. Our original model was based on this language, so only sentences ϕ could be known according to k_i. So only events of the form $[\phi]$ could be known.[19] But in a standard model of asymmetric information any event $A \subset \Omega$ can be known. So assuming the model is common "knowledge" will enable us to say more then we could using our language. For instance, if the model of figure 5.4b, is commonly "known" then we could deduce the following:

$\omega^* \in K_1(\neg CK_N(p))$, 1 knows that p is not common knowledge. But there is no sentence expressing this in our language, as our language only had finite sentences (and common knowledge requires infinitely many conjunctions).[20] Nevertheless, there is no other model that agrees with the model of figure 5.4b on all finite sentences.[21] Thus the finite language uniquely determines the model here. So, while there are statements that the Kripke model can imply that the language cannot even express, there is no doubt that this model is "correct" and, assuming it is common "knowledge" w.l.o.g., in that if we were to extend the language to infinite sentences we would get exactly the same conclusions as arise from the model.

There is, however, a second, more worrisome, problem. In general the model is *not* uniquely determined by what we called a complete description. Moreover, there is no way to a priori bound the depth of statements about knowledge that is needed to obtain a complete description.

Consider the two models in figure 5.5.[22] Player 1's partitions are ovals, 2's are rectangles, and 3's are diamonds. Assume the true state is any state of the world ω except ω^* and that p is a proposition that is true, say, only in state $\omega_{0,0}$. Then, formally, in the model of figure 5.5a, $\omega \in K_3(\neg CK_{1,2}(p))$, while, in figure 5.5b, this is not true. If we assume, informally, that the model is common "knowledge," then clearly all three players' knowledge coincide in both models except that in 5.5b player 3 does not know that p is not common knowledge among 1 and 2, and, in 5.5a, 3 does know this. However, such a sentence would not be included in ω since such sentences were not part of our finite syntactic framework.

Thus, the syntactic framework cannot distinguish between these two models while our informal common "knowledge" assumption does enable a distinction. So we have not fully achieved our objective of constructing a model that is common "knowledge" w.l.o.g. If we take as a complete description all sentences about finite levels of knowledge, then in the example of figure 5.5, assuming common "knowledge" of the information structure *is* a substantive assumption providing more information than originally contained in the, so-called, "complete" description.

One might think that the problem can be solved by allowing for a richer language, namely one in which not only finite conjunctions are permitted, but also conjunctions of sets of formulas of higher cardinality. Similarly, perhaps the problem can be addressed by introducing a syntactic symbol c_M analogous to CK_M for $M \subset N$.[23] However, there is no conjunction, and no collection of symbols, which would be large enough – Heifetz (1995c (see also 1995b)) shows that an example with the properties of figure 5.5 can be derived no matter how many conjunctions and symbols are introduced.[24]

This shows that the sense in which the constructed model can be assumed to be common "knowledge" w.l.o.g., depends crucially on the epistemic

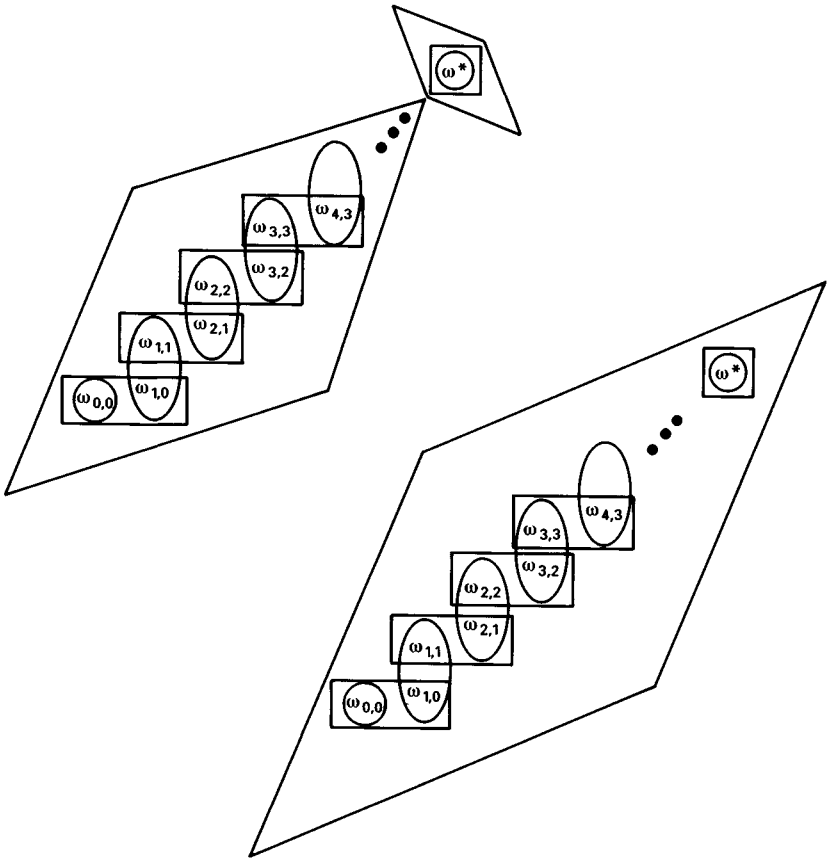


Figure 5.5 Two information structures, which, at any $\omega_{k,k-1}$, coincide insofar as our, so-called, complete description of a situation of incomplete information can determine, but differ in terms of the knowledge they imply if they are assumed to be common "knowledge." Thus a "complete" description does not distinguish between these two models, hence is not truly complete.

sentences which we want the model to incorporate. It is impossible to construct a model which allows us to interpret *all* conclusions that can be made with the model using the assumption that the model is common "knowledge." Only those conclusions that are meaningful in our original syntax are permitted. Alternatively put, while in figure 5.5a we could derive a statement about $K_3(\neg C K_{1,2}(p))$, we are not justified in interpreting this as player 3 knowing that p is not common knowledge among 1 and 2. Moreover, however large a language we start with, the model we construct

will still have conclusions of this form – i.e., using K_is and CK_Ms, etc. – that seem to have an intuitive interpretation, but whose interpretation is not valid since they are not expressible in the syntax.

On the other hand, it *might* be that we do not care about results where, say, CK_Ms appear transfinitely many times.[25] If this is the case, i.e., if we can a priori place a bound on the depth of knowledge about knowledge that is of interest to us, then we can construct a model such that all the "interesting" conclusions derived using the assumption that the model is commonly "known" will be correctly interpreted. To do this we would simply use the syntax that includes as deep an iteration as we are interested in for our results.

In conclusion, the syntactic approach does construct a commonly "known" partition model. Each state of the world is constructed from a "complete" and consistent list of sentences, and the derived model specifies only an abstract set of states, a specification of which elements in X are true in each state, and partitions. Consider any sentence about the knowledge of the individuals, that lies in some state of the world, say $k_i(\phi) \in \omega$. Using the K_i operators, we can derive an analogous result from the constructed commonly "known" partition model: $\omega \in K_i([\phi])$. Assuming the constructed model is common "knowledge" we thus obtain the same conclusions from the partition model as from the syntactic sentence. The problem we saw was that the converse is false: in the constructed model there are statements about players' knowledge that could not arise in the sentences of the syntactic framework. (That means that there may be formal results which can be proven using models, and which can be interpreted using the assumption that the model is common "knowledge," that could not be derived in the given syntactic framework.) No particular syntactic language enables a construction of a model which can be assumed to be common "knowledge" in this stronger sense. The extent to which this problem should concern us is not clear. First, instead of asking whether there is a language that provides complete descriptions for all possible models, we could reverse the order of the question. In fact, given any model that is assumed to be common "knowledge," there is some syntactic language, possibly a language that allows for "very many" conjunctions, that justifies and provides a complete description of that model.[26] For example, a language which would distinguish between the models of figure 5.5 would require either infinitely many conjunctions and/or a symbol for common knowledge among a subset of players, c_M. Second, as a pragmatic matter, if there exists a sufficiently rich language, in the sense that it incorporates every sentence that we might ever care about, then again there is no problem as we could just construct the models generated by that language.

3.2.2 The Bayesian approach

Harsanyi (1967) argued that a complete description of a situation of incomplete information would involve a specification of each person's beliefs, beliefs about beliefs, etc., and that the constructed *ex ante* state space is one where Ω is equal to the set of all such infinite hierarchies of beliefs. Mertens and Zamir (1985) explicitly constructed such an Ω, as did Ambruster and Böge (1978) and Böge and Eisele (1979), using a less familiar framework; these authors focused on Harsanyi's concern with games of incomplete information. Basically, they have shown that any situation of incomplete information that is completely described by a hierarchy of beliefs is equivalent to a state of the world in a standard, commonly "known," model of asymmetric information. Thus, in the context of a Bayesian model, the problem we ran into above with a syntactic model seems to be solved.[27]

Consider a basic space of uncertainty, S, which is commonly known to be the basic set of physical uncertainties of interest for agents 1 and 2.[28] Let $X_0 = S$, and for any $n > 0$, let $X_n = [\Delta(X_{n-1})]^N \times X_{n-1}$. Player i's (first-order) beliefs over S are an element of $\Delta(X_0)$, denoted by t_1^i; i's (second-order) beliefs about S and about j's beliefs over S are an element t_2^i of $\Delta(X_1)$, and so on.[29] Thus, a complete specification of i's beliefs is an element of i's type space, $T_0^i = \Pi_{n=1}^{\infty}\Delta(X_n)$. This generates an expanded space of uncertainty that appears to include all uncertainties: an element of $\Omega \equiv S \times T_0^1 \times T_0^2$ specifies the true physical state as well as all of i's beliefs and all of j's beliefs.

However, there are two related problems. First, this construction just begs the question of what are i's beliefs over j's types, i.e., over T_0^j. Second, the beliefs just constructed may be incoherent; for example i may fail to know his own beliefs (we have allowed him to have non-degenerate beliefs over his own beliefs), and i's beliefs may fail to uniquely specify his own beliefs, for example, i's belief about S calculated from i's second-order beliefs t_2^i, $\text{marg}_{X^0} t_2^i \in \Delta(S)$ may differ from i's first-order beliefs, $t_1^i \in \Delta(X_0)$. Since we want to assume that i's beliefs are coherent we make two assumptions: i's beliefs over his own beliefs are his actual beliefs, i.e., i knows his own beliefs (see [4] and [5] above), and i's beliefs on any set calculated using any order belief coincide. Thus, it is assumed that i's beliefs are an element of $T_1^i \subset T_0^i$ which denote those beliefs which are coherent.[30] It turns out that restricting attention to coherent beliefs also solves the first problem. This follows from Kolmogorov's theorem which implies that a complete specification of beliefs for i, $\tau_0^i = (t_1^i, t_2^i, \ldots) \in T_0^i$ is coherent, i.e., is in T_1^i if and only if there is a corresponding belief for i over S and over all of j's possible types, namely a $\mu \in \Delta(S \times T_0^j)$ such that the beliefs given by μ on any measurable set A in X_n

coincide with the beliefs given by τ_0^i (equivalently, the beliefs given by t_n^i) on A. That is, a coherent belief for i, which is a belief over S and over j's beliefs, and j's beliefs over S etc., also determines a belief over j's possible types. But i's beliefs, even if coherent, are not guaranteed to determine a belief for i over j's beliefs over i's types. This will happen if and only if i assigns probability zero to types of j that are incoherent. If we define knowledge as probability 1 – this will be discussed further in section 3.3 below – then we can say that if i knows that j is coherent then i can derive beliefs over j's beliefs over Ω. Note the similarity with partitions: if i "knows" j's partition and that j knows i's partition, then i can calculate j's belief over i's beliefs; while here, if i knows j is coherent then i can calculate j's beliefs over i's beliefs. More generally, there is a similar correspondence between i knowing that j knows that . . . i is coherent and i knowing that j knows . . . i's partition. Thus, common knowledge of coherency is a formalization within the model of the informal assumption that the partitions and beliefs are common knowledge; however it seems less controversial.

Thus, by assuming common knowledge of coherency we will generate spaces $T^i \subset T_0^i$ which have the property that each type in T^i is a complete and consistent description: each type is a belief over S, and over the other players' belief over S, etc., moreover each such type generates a belief over $\Omega \equiv S \times T^1 \times T^2$. So we have created an *ex ante* space Ω, and a possibility correspondence where each player i is informed only of his type in T^i, i.e., for $\bar{\omega} = (\bar{x}, \bar{t}^1, \bar{t}^2)$, we have $\mathscr{F}_i(\bar{\omega}) \equiv \{\omega \in \Omega : \omega = (x, t^1, t^2) \text{ s.t. } t^i = \bar{t}^i\}$. This information structure generates a belief over the space that coincides with the belief described by the state of the world. So, this structure can be taken to be common "knowledge" w.l.o.g.

How have we achieved this result which was unattainable before? The richer and continuous structure of countably additive beliefs is crucial here. Consider the example in figure 5.5 again. As before, for any strictly positive probability measure, at any state other than ω^*, 3 does not know $\cap_{m=1}^n K_{1,2}^m(p)$ for any n. In the syntactic approach we say no more, so we cannot specify whether 3 knows $\neg CK_{1,2}(p)$ or not. In the Bayesian approach, if, say 3 does not know $\cap_{m=1}^n K_{1,2}^m(p)$, for all m, then there exists a decreasing sequence, $p_n < 1$, of the probabilities which 3 assigns to this sequence of events. The countable intersection of these events is exactly $CK_{1,2}(p)$. So the probability of this limit event is given by the limit of the sequence. If the limit is 0, then 3 knows that p is not common knowledge among 1 and 2, while if it is positive 3 does not know that p is not common knowledge among 1 and 2. And this conclusion is true regardless of 3's partition, i.e., for both models in figure 5.5, since knowledge here is defined as belief with probability 1.[31]

3.3 Belief with probability 1 or knowledge?

The usual notion of knowledge requires that when a player knows something it is true. This is property [T], which in the partition model of section 3.1 results from the assumption that ω is an element of $\mathcal{F}_i(\omega)$, and which is assumed directly in the syntactic framework of section 3.2.1 and built into the construction in section 3.2.2. However, the probabilistic notion of certainty used in the Bayesian model need not have such an implication. This subsection discusses the relationship between knowledge and certainty and briefly shows how to adapt the presentations above to a notion of certainty. The distinction between certainty and knowledge, and the combined development of both notions within one model in subsection 3.3.3 below, turn out to be very useful in discussing extensive-form games (see section 5).

3.3.1 A certainty operator

Given an information structure $(\Omega, \mathcal{F}_i, p_i)$, we can derive, in addition to K_i, a belief-with-probability-one operator $B_i: 2^\Omega \to 2^\Omega$ given by $B_i(A) = \{\omega: p_i(A \mid \mathcal{F}_i(\omega)) = 1\}$. Recall that we use the term certainty as an abbreviation for belief with probability one. This certainty operator is equivalent to the following: at any state ω you are certain of (any superset of) the intersection between your information at that state $\mathcal{F}_i(\omega)$ and the support of your beliefs (denoted by S). The operator B_i satisfies the properties below.

D $B_i(A) \subset \neg B_i \neg A$: if i is certain of A then i is not certain of the complement of A;

MC^B $B_i(A) \cap B_i(C) = B_i(A \cap C)$: being certain of A and C is equivalent to being certain of A and of C;

N^B $B_i(\Omega) = \Omega$: i is always certain of anything that is true in all states of the world;

4^B $B_i(A) \subset B_i(B_i(A))$: if i is certain of A then i is certain that i is certain of A;

5^B $\neg B_i(A) \subset B_i(\neg B_i(A))$, being uncertain of A implies being certain that one is uncertain of A.

As in subsection 3.1, given such a certainty operator B_i we can define an information structure on Ω: let p_i be any probability on Ω with support $S \equiv \cap \{E: B_i(E) = \Omega \text{ and } E \text{ is closed}\}$, and let $\mathcal{F}_i(\omega) = \{\omega': \forall E, \omega' \in B_i(E) \Leftrightarrow \omega \in B_i(E)\}$.[32] Moreover, this information structure will generate (in the way defined above) the same B_i as we started with.

3.3.2 *A syntactic symbol for certainty*

How is the syntactic notion of knowledge weakened to certainty? Start with a language as in subsection 3.2.1, and introduce a symbol b_i for certainty instead of k_i for knowledge. Consider modifying the axioms [D, MC^B, N^B, 4^B, and 5^B] as we did in going from the operator K_i to the language symbol k_i. That is, replace B_i with b_i, $\supset$ with $\rightarrow$, sets A with propositions ϕ, etc., and add [RE] and [MP]. We can then create a consistent and "complete" description of situations of incomplete information, which will generate a state space and an operator B_i satisfying the axioms. Since we have just seen that such a B_i is equivalent to an information structure we again have an equivalence between the syntactic approach and the asymmetric information model.

3.3.3 *Knowledge and certainty combined*

For characterizing solution concepts in games we will want a model that allows for both knowledge and certainty: for example, a player should know what her own strategy is, but can at most be certain, but not know, what her opponent's strategy is. Therefore, we now present a unified treatment of K_i and B_i.[33] Given an information structure we can generate K_i and B_i as above. In addition to the properties derived above, these will satisfy the following.

BK $K_i(A) \subset B_i(A)$: you are certain of anything that you know;
4^{BK} $B_i(A) \subset K_i(B_i(A))$: you know when you are certain of something;
5^{BK} $\neg K_i(E) = B_i(\neg K_i(E))$.

Similar to the equivalencies in subsections 3.1 and 3.3.1, given a B_i and a K_i satisfying [T, BK, 4^{BK}, 5^{BK}, MC, MC^B] we can construct a partition and a non-unique prior which in turn generate K_i and B_i.[34] In the future we will need a notion analogous to common knowledge for the case of beliefs. We say that an event E is common certainty at ω if everyone assigns probability 1 to E, and to everyone assigning probability 1 to E, etc.

3.4 The implications of constructing a commonly "known" model

We have discussed the constructions of Harsanyi's and Aumann's expanded state spaces. These constructions show that it is without loss of generality to work with models where we assume (informally) that the information structure is common "knowledge." Alternatively put, we *can* take any situation of incomplete information that is completely specified and consistent, and view it as a state in a model that is commonly "known."

This will be useful in formalizing assumptions such as common knowledge or rationality, and in solving games where we are given some situation of incomplete information. The main purpose of this subsection is to argue that these constructions do *not* justify acting as if all circumstances derive from a commonly known *ex ante* model. The most obvious reason for this is that the constructed models of subsection 3.2 do not contain a prior: only probabilities conditional on a players' information are constructed.[35] Moreover, it seems problematic to argue that, say, Savage's (1954) framework justifies assuming that players have a prior on the constructed state space. This state space includes states that the individual views as impossible. So forming preferences over acts would require contemplating preferences conditional on knowledge that conflicts with knowledge that the player actually has, which seems conceptually problematic (and in violation of the behavioral objectives of Savage (1954)). We have a second type of concern with assumptions that are based on an *ex ante* model; these concerns follow from the difference between justifying assumptions and solution concepts for the artificially constructed model and justifying them for a true physical *ex ante* model.

The argument that no prior was constructed immediately leads to the conclusion that we ought to be cautious in using concepts that must be defined in terms of priors. So, for example, *ex ante* efficiency and the CPA are both questionable notions in true situations of incomplete information. Naturally, in contexts where a real physical *ex ante* stage does exist, there is no problem: then a Savage (1954) approach would justify assuming priors on the space at the *ex ante* stage.[36] But here we focus on situations of incomplete information where the commonly "known" information structures, as constructed in section 3.2, are needed.

In the remainder of this subsection we elaborate further on the limitations of these constructions. First we highlight the artificial nature of the construction. Then we suggest that this raises doubts about other notions, such as interim and *ex post* efficiency, as well as raising additional concerns with justifications for the CPA. To clarify the artificial nature of the state space constructed in subsection 3.2, we consider a simple version of this construction, one which does not yield a space of all possible situations of incomplete information, but does have a particular infinite hierarchy of beliefs that is captured by a standard model. First, assume that there is some fact about the world and the agents are concerned as to whether it is true or false. Denote this fact by p, and assume that we are in a situation where for $i = 1, 2$, and $j \neq i$, i knows p but does not know if j knows it, and believes that j does not know if i knows it, and more generally i's hierarchy of beliefs is that there is common knowledge that i does not know if j knows it and j does not know if i knows it. This can be captured by the following Kripke

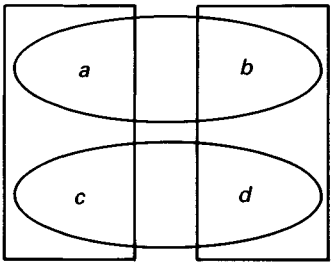


Figure 5.6

model: $\Omega = \{a, b, c, d\}$; p is true in states a, b, and c, and false in state d; $\mathscr{F}_1 = \{\{a, b\}, \{c, d,\}\}$; $\mathscr{F}_2 = \{\{a, c\}, \{b, d\}\}$ (figure 5.6).

The true situation is described by state a. All the states other than a are artificial constructs. While they represent situations the agents can imagine, they do not represent reality in any way. Recall that we are assuming that we are given the situation of incomplete information described above, which did not arise from any particular commonly known *ex ante* physical situation. In particular both agents as well as the analyst know that state d is false, and that it is an artificial construct.

What does this artificiality imply? First, we think it raises doubts about the use of other efficiency notions, such as interim and *ex post* efficiency. These notions seem to us to be based on giving the artificially constructed states more meaning then they have. However, we have not been able to develop this argument. Therefore, we should emphasize that our concerns with interim and *ex post* efficiency notions are on less solid grounds than the clear-cut argument – based on the lack of a prior – against *ex ante* efficiency.

A second consequence of the artificial nature of the *ex ante* state space can be found in Bhattacharyya and Lipman (1995). They provide an example of trade with a common prior, despite the no-trade theorem. Trade occurs because *ex ante* utilities are not well defined, since the *ex ante* utility functions are unbounded. But the interim utility functions can be bounded without changing the essence of the example. While it seems plausible to argue that utility functions are bounded, does this argument apply to an *ex ante* artificial construct? We would say no, raising doubts about no-trade theorems in such contexts, over and above any concerns about the common prior assumption.[37]

The final problem that results from the artificial nature of the state space concerns the main argument in favor of the CPA. This problem is of interest because justifications for standard equilibrium concepts require a common prior on such an artificial state space. (We say that an artificial state space is necessary, because in the model that justifies solution concepts, the *ex ante*

stage is one where the player does not know his own strategy, and may not even know whether or not he is rational. The assumption that there actually exists such a stage for decisionmakers seems very questionable.) The problem focuses on Aumann's (1987) argument that a common prior is plausible because players with identical information would have the same beliefs. But how could player 1 receive the information that corresponds to 2's artificial information set $\{a, c\}$, which is constructed as the set where player 1 both knows p and does not know p? (see also Gul (1995a)).

While we have said this several times, it is probably worth repeating here. If we set aside the question of justifying solution concepts, then for some economic applications the discussion of this subsection is not of direct interest. If there is a real *ex ante* situation, then the notions are meaningful since they do not rely on tradeoffs over artificial constructs. For example, auction theory, while often citing Harsanyi's work as a justification for the incomplete-information model, could be based on an oil-sample story such as was described in subsection 3.1, where intuitively there is a real physical *ex ante* commonly known framework. Alternatively, in auctions with private values, an *ex ante* foundation exists by arguing that bidders are drawn from a population with a commonly known distribution of preferences. On the other hand, while a private-values with correlated beliefs model could be imagined, it does not seem to correspond to any plausible *ex ante* story, in which case any research examining *ex ante* efficiency of various mechanisms in such a context needs to be motivated much more carefully.

Having argued that both the notion of efficiency in games of incomplete information, and the assumptions underlying standard solution concepts, are not plausible in the artificially constructed space, one might think that the whole exercise was vacuous: can any sensible assumptions be made on the constructed model? Assumptions that do not refer to the constructed state space, but rather are assumed to hold in the true state are on a solid footing. For example, the assumption that at the actual state rationality is common knowledge, is sensible. Such a statement only uses the artificially constructed states the way they originated – namely as elements in a hierarchy of beliefs. This obviously contrasts with assumptions that essentially require the artificial constructs in order to be interpreted, such as the CPA.

3.5 Conclusion

When there is a real commonly known *ex ante* stage then clearly it is appropriate to model the situation with an asymmetric information model that is commonly "known." The constructions in subsection 3.2 justify the

assumption of a commonly known asymmetric information model in all contexts where the players' views of the world are complete. These two justifications of (almost) the same model differ in that only the former has a well-defined notion of a prior and only in the former are all the states truly feasible. This distinction argues against using *ex ante* notions in cases where the second model is deemed necessary.

However, these notions are used heavily in information economics and – in the case of the CPA – in the characterizations of solution concepts and the analysis of their robustness. For these reasons we will emphasize results that avoid the CPA, but we will still review other results, such as Aumann's provocative characterization of correlated equilibrium, as well. Moreover, we will use both the real *ex ante* stage and the artificially constructed commonly known model interchangeably with the hope that by now the reader understands the important differences between them. In particular, we will continue to describe a model as $(\Omega, \mathscr{F}_i, p_i)$, rather than $(\Omega, \mathscr{F}_i, p_i(\cdot \mid \mathscr{F}_i))$, even though the latter, rather than the former, is what we justified in this section. We will leave it to the reader to decide on the usefulness of the various results, and we try to minimize repetition of our concerns in the remainder of the chapter.

4 A STATE-SPACE MODEL FOR KNOWLEDGE, CERTAINTY, AND RATIONALITY – CHARACTERIZING NORMAL-FORM SOLUTION CONCEPTS

We now introduce the notion of a model for characterizing solution concepts, which specifies an information structure and a function from states to normal-form games and to strategy profiles of the game in that state.[38] (A model is the same as the interactive belief systems used by Aumann and Brandenburger (1995)), and closely related to the framework used in Aumann (1987) to characterize correlated equilibrium (see also Stalnaker (1994)). For simplicity we assume that strategy spaces of the games are the same in every state; Σ_i denotes i's strategy set, and $\Sigma \equiv \Pi_i \Sigma_i$ is the set of strategy profiles. Thus, for the remainder of the chapter, a model is a collection $\{\Omega, \mathscr{F}_i, p_i, \Sigma_i, \mathbf{u}, s_i\}$, where, $s_i: \Omega \to \Sigma_i$ specifies i's actions in each state $\omega, \mathbf{s}(\omega) = (\mathbf{s}_1(\omega), \ldots, \mathbf{s}_n(\omega))$, and $\mathbf{u}: \Omega \to \{(u_i)_{i \in N} \mid \forall i, u_i: \Sigma \to \Re\}$ specifies the payoff functions. Thus $\mathbf{u}_i(\omega')(\sigma)$ is the payoff to i in state ω' if the strategy profile σ is played. We also assume that the partitions of i have the property that in each cell a common strategy of i is specified at all states: i knows his own action, $\forall F \in \mathscr{F}_i, \forall \omega, \omega' \in F, \mathbf{s}_i(\omega) = \mathbf{s}_i(\omega')$. Finally, we assume that each player knows his own payoff function (u_i is the same in each cell of i's partition). This last assumption is substantive, see, e.g., footnote 4. Note

that if we ignore the function **s** specifying the strategy profile in each state, then a model is just a game with a move by Nature: it specifies an information structure, and, for each state of the world, a game. Under this interpretation **s** is a strategy profile for the game with a move by Nature. This interpretation will be useful later.

We denote by $[u]$ the event in Ω that the payoff functions are $u = (u_1, \ldots, u_n)$, $[u] \equiv \{\omega \in \Omega: \mathbf{u}(\omega) = u\}$ by $[\sigma]$ the set of states where the strategy profile is σ, $[\sigma] \equiv \{\omega \in \Omega: \mathbf{s}(\omega) = \sigma\}$, and similarly for $[\sigma_i]$, $[\sigma_{-i}]$, etc. This notation simplifies our assumptions above: i knows his own action is simply $\forall \sigma_i, K_i[\sigma_i] = [\sigma_i]$; and i knows his payoff becomes $\forall u_i, K_i[u_i] = [u_i]$. At each state ω, each player has an induced belief over the opponents, which we denote by $\text{marg}_{\Sigma_{-i}} p_i(\cdot \mid \mathcal{F}_i(\omega))$; the event that these beliefs equal some particular distribution $\phi_{-i} \in \Delta(\Sigma_{-i})$ is denoted by $[\phi_{-i}]$. Following Aumann and Brandenburger (1995) we use the term conjectures of i as an abbreviation for i's induced beliefs over Σ_{-i}. Finally, we denote by $\mathcal{S}$ the operator on games of deleting one round of strongly dominated strategies for all players; similarly $\mathcal{W}$ denotes deletion of weakly dominated strategies. Since the strategy spaces are held constant we abuse notation and write $\mathcal{S}^\infty(u)$ to denote the operation of infinite deletion of strongly dominated strategies in the game $G = (\Sigma, u)$.

Definition Player i is rational in state ω if given i's beliefs at ω, his action maximizes his expected utility

$$\sum_{\omega' \in \mathcal{F}_i(\omega)} p_i(\omega') \mathbf{u}_i(\omega')(\mathbf{s}_i(\omega), \mathbf{s}_{-i}(\omega')) \geq \sum_{\omega' \in \mathcal{F}_i(\omega)} p_i(\omega') \mathbf{u}_i(\omega)(\sigma_i, \mathbf{s}_{-i}(\omega'))$$

for all $\sigma_i \in \Sigma_i$.

The set of states at which all players are rational is the even $[rationality]$.

Proposition 1 (Bernheim (1984), Pearce (1984))[39] *$CB([u] \cap [rationality]) \subset [S^\infty(u)]$. Moreover, there exists a model such that $CB([u] \cap [rationality]) = [S^\infty(u)]$.*

If at a state ω, rationality and the game is common certainty, then at ω each player is choosing an action that survives iterative deletion of strongly dominated strategies. The idea of the proof is well known – rationality is equivalent to players choosing only strategies in $\mathcal{S}^1(u)$; the fact that rationality is known implies that they only choose best replies to $\mathcal{S}^1(u)$, so only strategies in $\mathcal{S}^2(u)$ are chosen, etc.

What is the relationship between equilibrium concepts and $\mathcal{S}^\infty$? Brandenburger and Dekel (1987a) show that the strategies and payoffs resulting

from $\mathscr{S}^\infty(G)$ are the same as the strategies and payoffs in the interim stage of an a posteriori subjective correlated equilibrium of G.[40] An a posteriori subjective correlated equilibrium of a game G is essentially a Nash equilibrium of the game where an information structure, interpreted as a correlating device about which players may have different priors, is observed before G is played. (Aumann (1974) introduced this solution concept.) The interim stage of such an equilibrium is the stage after receiving the information of the correlating device. A correlated equilibrium is the same as an a posteriori equilibrium except that there is a common prior over the correlating device. A correlated equilibrium distribution is the probability distribution over Σ induced by the correlated equilibrium.

Proposition 2 *Fix an information structure* $(\Omega, \mathscr{F}_i, p_i)$ *and a game* $G = (\Sigma, u)$. *Consider a game* G', *where before* G *is played, the players observe their private information concerning* Ω. *Strategies for* i *in* G' *are functions* $\mathbf{s}:\Omega \to \Sigma_i$ *that are constant on an information cell for* i. *Consider a Nash equilibrium* $\bar{\mathbf{s}}$ *of* G', *where* $\bar{\mathbf{s}}_i(F_i)$ *is optimal for all* F_i *(even those with zero prior probability). The interim strategy choices (and expected utilities) are rationalizable:* $\bar{\mathbf{s}}(\omega) \in \mathscr{S}^\infty(G)$ *for any* F_i *in* $\mathscr{F}_i$.

Conversely, given any strategy $\sigma \in \mathscr{S}^\infty(G)$ *there is an information structure and a Nash equilibrium as above where* σ *is played in some state of the world.*[41]

Thus, common certainty of rationality justifies equilibrium analysis so long as the equilibrium allows for differing priors. To get more traditional equilibrium concepts one typically needs to assume a common prior on Ω.[42] Aumann's characterization of correlated equilibrium was the first characterization of an equilibrium concept within a formal state-space model describing common knowledge of players' rationality.

Proposition 3 (Aumann (1987)) *If there is a common prior* p *in a model, and the support of* p *is a subset of* $[\textit{rationality}] \cap [u]$, *then the distribution over actions induced by the prior* p *is a correlated equilibrium distribution of* $G = (\Sigma, u)$.

Intuitively, this follows from propositions 1 and 2: common certainty of rationality is the same as $\mathscr{S}^\infty(G)$ which is the same as subjective correlated equilibrium; imposing a common prior in addition to common certainty of rationality should then be the same as objective correlated equilibrium. This is not precise because propositions 1 and 2 focused on the players' actual beliefs at a state of the world, not on the *ex ante* constructed model and an overall distribution of actions.[43]

Proposition 4 (Aumann and Brandenburger (1995)) *In a two-person game if the events* [*rationality*], [u] *and* [ϕ_{-i}] *for* $i = 1, 2$, *are mutually certain at* ω (*i.e., each player assigns conditional probability* 1 *to these events*), *then* (ϕ_1, ϕ_2) *is a Nash equilibrium of* $G = (\Sigma, \mathbf{u}(\omega)) \equiv (\Sigma, u)$.

The idea of the proof is as follows. First add the assumption that the players are mutually certain that they are mutually certain that the payoffs are u. If i is certain that j's conjecture is $\phi_{-j} \in \Delta(S_i)$, and that j is rational and that j is certain his payoffs are u_j, and i assigns positive probability to σ_j, then σ_j must be a best reply for j given j's conjecture about i's actions, ϕ_{-j}, and given u_j. So, under the additional assumption the result that (ϕ_1, ϕ_2) is a Nash equilibrium is obtained. In fact, since we assume that players know their own payoffs, Aumann and Brandenburger show that one only needs to assume that the payoffs are mutually known. This is because if i assigns probability 1 at ω to $[u_j]$, [*rationality*] and $[\phi_{-j}]$, and positive probability to σ_j, then there must be a state $\omega' \in [u_j] \cap$ [*rationality*] $\cap [\phi_{-j}] \cap [\sigma_j]$. At ω', j is rational, j's conjecture is $[\phi_{-j}]$, j's payoffs are $[u_j]$ and j knows this by our assumption, and j is choosing σ_j. This completes the proof.[44]

It is worth emphasizing that the statement that $[\phi_i]$ are mutually certain is significantly stronger than saying that i is certain j's conjecture is $[\phi_j]$. Since players have beliefs about their own conjecture, and naturally their beliefs about their own conjecture are correct, assuming that conjectures are mutually certain implies that the beliefs about the conjectures are correct (see Aumann and Brandenburger (1995, lemma 4.2)). By contrast, for an arbitrary event E, players 1 and 2 could be mutually certain of E but be wrong.

Aumann and Brandenburger (1995) also characterize Nash equilibrium in games with $n > 2$ players, using the CPA, common knowledge of the conjectures, and mutual knowledge of payoffs and rationality. They also provide a series of examples to show the necessity of these assumptions. They discuss a second characterization of Nash equilibrium, where the CPA and common certainty of the conjectures are replaced with independence, which they find less attractive (see also Brandenburger and Dekel (1987a) and Tan and Werlang (1988)). Given our concerns with the CPA we find the characterization using independence no less palatable.

In conclusion, we have concerns with applications of game theory whose conclusions rely on Nash or correlated equilibrium. The role of the CPA raises doubts about the use of correlated equilibrium; the necessity of mutually certain conjectures raises some doubts about Nash equilibrium in two-person games; and the peculiar combination of the CPA and common certainty of conjectures, or independence and mutual certainty of conjectures, raises serious doubts about the interpretation of Nash equilibrium in

games with more than two players. Naturally, other foundations for these solution concepts may exist, and some are currently being developed in the context of evolutionary and learning models. However, until such foundations are obtained we must be cautious in interpreting applications that rely on Nash equilibrium.

We now discuss the application of the ideas above to games with moves by Nature and games of incomplete information. As we know from section 3 there is a difference between these two environments. In one, there is a real *ex ante* stage at which an analysis can be carried out, at the other *ex ante* stage is an artificial construct used to think about the situation of incomplete information. Obviously, the way we would analyze the case of a real *ex ante* stage is by embodying the whole game with a move by Nature into a state. To clarify this, recall that in the analysis above the state ω determined the game by specifying the strategies and the payoffs. In the case of a game with a move by Nature these strategies are functions from the players' private information into their choices, and the payoffs are the *ex ante* payoffs. Carrying out the analysis at a state ω will mean providing an *ex ante* solution of this game, i.e., specifying what functions from private information about Nature into Σ will be played. So, in the case of a game with a move by Nature, i.e., with a real *ex ante* stage, the model is unchanged and the results above can be meaningfully applied to the *ex ante* stage of the game. In the case of a game of incomplete information, the space Ω will capture all the incomplete information, and carrying out the analysis at a state ω will mean specifying a strategy in Σ (not a function from private information about Nature into Σ). The model for this case will be the same as the one used throughout this section, except that now $[u]$ is no longer mutual or common certainty, so the results obtained earlier are not meaningful. We now examine the implications of this in the context of the characterizations of $\mathscr{S}^\infty$ and of Nash equilibrium.[45]

First consider proposition 1: in this case common certainty of rationality implies that players will choose actions that are iteratively undominated in the interim sense in the game of incomplete information. (The distinction between *ex ante* and interim dominance can be seen, e.g., in the example in Fudenberg and Tirole (1991, p. 229).) On the other hand, if a game with a move by Nature is played at ω then the *ex ante* payoff function is commonly known and iterated deletion of *ex ante* dominated strategies will be the consequence of common certainty of rationality.

Next consider proposition 4. If we consider the case of a game with a move by Nature there are no problems: if we assume that the game and rationality are mutually certain at ω, as are the players' conjectures – which are over functions from private information into Σ – then we have characterized Nash equilibria of the game with a move by Nature.

However, for the case of true games of incomplete information, where there is no *ex ante* model, there seems to be no obvious characterization of "Bayesian Nash equilibrium at ω." That is, we are not aware of an equilibrium analog to iterated deletion of interim-dominated strategies.

The solution concepts above do not include any refinements. One might want to see the extent to which concepts from the refinements literature can be characterized using assumptions about knowledge, certainty, and rationality. In particular, one might want to add an assumption of *caution* as a requirement, namely that player i is never certain about j's strategy; the event where this holds, [*caution*], is formally $\cap_i \cap_{\sigma_{-i}} \neg B_i \neg [\sigma_{-i}]$. On this event i's conjecture over his opponents has full support on Σ_{-1}. Clearly this is inconsistent with common knowledge or common certainty of rationality since these full support beliefs do not allow a player to know or be certain about anything concerning their opponents' strategies. Thus there is no way to assume common certainty of rationality and that players are cautious, since caution conflicts with certainty.[46] We return to this issue in section 6 below.

5 CHARACTERIZING SOLUTION CONCEPTS IN EXTENSIVE-FORM GAMES AND COMMON KNOWLEDGE/CERTAINTY OF RATIONALITY

5.1 The model

Much of the research on rationality in the extensive form falls in one or more of the following categories: criticisms of backward induction and analysis of the problematic nature of common knowledge of the rationality assumption for extensive-form games, identification of weaker, non-backward induction theories as the consequence of rationality, and, finally, alternative axioms or formulations of common knowledge of rationality that yield backward induction. Below we will utilize the knowledge and belief structure described in section 3 to discuss and summarize these results.

Let $\Gamma = \{Y, Z, \prec, N, I, A, (u_i)_{i=1}^n\}$ be an extensive-form game of perfect information. The set $X \equiv Y \cup Z$ denotes the set of nodes and Z is the set of terminal nodes. The binary relation $\prec$ on X is transitive and irreflexive and hence has a minimal element. Furthermore, $\prec$ satisfies arborescence: $x \prec v, y \prec v$ implies $x \prec y$ or $y \prec x$ or $x = y$. The function I determines for each non-terminal node in Y the player $i \in N$ who moves at that node. The mapping A associates with each non-terminal node in Y, a non-empty set of nodes (the set of actions for player $I(v)$), $A(v) = \{y \mid v \prec y$ and $v \prec y'$ implies $(y = y'$ or $y \prec y')\}$. A strategy s_i for player i is a collection actions, one from

each node that is not precluded by and earlier action of player i. The set S_i is the set of all strategies for player i, s denotes a strategy profile, and S is the set of all strategy profiles. Each u_i associates with every terminal node a payoff for agent i.

As in section 4, a model specifies an information structure and for each state the game to be played at that state and the strategies chosen at that state. Here we assume for simplicity that the game is the same, Γ, in every state. A model is then $\mathbf{M} = (\Omega, \mathbf{s}, \mathscr{F}_i, p_i, \Gamma)$. As in section 3.3.3 the model generates knowledge and certainty operators K_i and B_i satisfying Axioms $[\mathrm{BK}, \mathrm{T}, 4^{BK}, 5^{BK}, MC, MC^{BK}]$. Much of the following will be stated using these operators, rather then using the partitions, $\mathscr{F}_i$, and conditional probabilities $p_i(\cdot \mid \mathscr{F}_i)$. In order to identify the implications of common knowledge of rationality or common certainty of rationality in various settings, we need to identify nodes with subsets of Ω: $[\Lambda] = \bigcup_{x \in \Lambda}[x]$ for $\Lambda \subset S$, S_i, or S_{-i}.

It is often said that in discussing rationality in the extensive form there is a need to incorporate counterfactuals or nearby worlds or hypotheticals or at least a different kind of implication than the standard material implication of propositional logic into the analysis (see Binmore (1987–8), Samet (1993) and Aumann (1995b)). There are two separate issues that necessitate counterfactuals. The first arises even in normal-form games. To justify the rationality of action α for a given player, the other players and the modeler have to argue that, given what he knows and believes, choosing α is a good as any other action the agent might have undertaken. But this requires discussing what would happen if this player were to choose β at some state in which he is specified to choose α. To be sure, this is a counterfactual of some sort but not one that requires an elaborate theory of nearby worlds. As in section 4, such counterfactuals will play a role only in the definition of rationality and only implicitly. The second source of need for counterfactuals or hypotheticals is present only in extensive-form games. The rationality of a particular action α may rest on the predicted reaction of the opponent to the alternative course of action β which itself may or may not be consistent with rationality. Furthermore, the prescribed rational course of action α or the contemplated alternative β may or may not be anticipated by the opponent of the player. Thus, the modeler is forced to discuss not only contemplated deviations by a rational player but also predicted responses to deviations or surprises. In our approach to rationality in the extensive form we will utilize the distinction between knowledge and certainty to do away with this second source of need for counterfactuals. Specifically, an event E will be considered a surprise by player i at any state in $\neg K_i \neg E \cap B_i \neg E$. Moreover, we will often require that agents are, at most, certain about their opponents' strategies and cannot know the

strategies their opponents are using. Thus we will not need to consider what players do when faced with situations they *know* to be impossible (i.e., counterfactuals). Therefore, the only counterfactuals will be those implicit in the definition of rationality; strategy β will be irrational in some state at which the agent is to choose α since had he chosen β he would receive a lower payoff.

Definition (extensive-form rationality): For any model $\mathbf{M}$ let (a) $\mathbf{D}(s_i) = \{E \subset \Omega \mid \exists \beta \in S_i \, s.t. \, u_i(\beta, \mathbf{s}_{-i}(\omega)) > u_i(s_i, \mathbf{s}_{-i}(\omega)) \forall \omega \in E\}$; (b) $\neg R^v(s_i) = \neg K \neg [v] \cap \bigcup_{E \in \mathbf{D}(s_i)} K_i([v] \to E) \bigcup \neg B \neg [v] \cap \bigcup_{E \in \mathbf{D}(s_i)} B_i([v] \to E$; (c) $R^v = \{\omega \mid \mathbf{s}_{I(v)}(\omega) \in R^v(\mathbf{s}_{I(v)}(\omega))\}$, $R_i = \bigcap_{I(v)=i} R^v$, $R = \bigcap_i R_i$.

The first item of the definition above identifies the collection of events in which the strategy s_i is strictly dominated. The second item defines the set of states of the world in which the strategy s_i is irrational at a particular node v. The strategy s_i is irrational at $[v]$ if $[v]$ is possible ($\neg K_i \neg [v]$) and i knows that s_i is dominated at v or v is plausible ($\neg B \neg [v]$) and i believes s_i is dominated at v. Note that, if i knows that v will not be reached, then he is rational at v regardless of what he plans to do there. The final part of the definition states that i is rational at v in state ω if and only if the prescribed strategy for i is not irrational at $[v]$. The event "i is rational" corresponds to all states at which i is rational at every v.

5.2 Common knowledge of rationality

Since knowledge implies truth (Axiom T of section 3), a standard model of knowledge may fail to incorporate even the most basic requirement of extensive form rationality. Consider the game Γ_1 in figure 5.7.

Player 2's knowledge that player 1 will not choose β renders any analysis of what *might* happen *if* Player 1 chooses β irrelevant for the analysis of Γ_1 given our definition of rationality. Various ways of dealing with this issue have been studied in the literature. Following the lead of the refinements literature and extensive-form rationalizability (see Pearce (1984)), Ben-Porath (1994), Reny (1993), Gul (1995b) require rationality in the extensive form to imply that a player at an information set chooses a best response to some conjecture that reaches that information set. Hence common knowledge of rationality rules out player 1 choosing α in Γ_1. By contrast Samet's (1993) formalism does not yield the result that common knowledge of rationality implies player 1 chooses α. Without further assumptions, any Nash equilibrium outcome is consistent with common knowledge of rationality in his model. The same is true of the standard model defined above. If player 2 knows that player 1 will not choose β then any analysis of

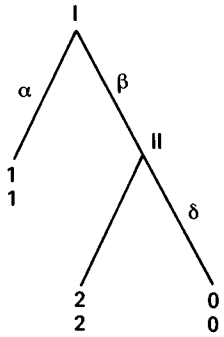


Figure 5.7

what 2 would do when faced with a surprise is moot. This suggests that an additional assumption is needed if the standard model is to incorporate surprises: in an epistemic model of extensive-form rationality, convictions regarding opponents should be limited to beliefs if this model is to adequately deal with surprises and the issue of credibility. This is a weakening of the assumption of caution introduced at the end of section 4, which required that a player not be certain about an opponent's action, here we allow for certainty but rule out knowledge of an opponent's action. This idea is formalized in the definition below. Proposition 5 summarizes the implications of common knowledge of rationality for the standard model when no such axiom is imposed.

Proposition 5 *Let s be a Nash equilibrium of the perfect information extensive-form game Γ. Then, there exists a model* **M** *such that* $CK[R] = [s]$. *Moreover, if Γ is generic, then for any model* **M**, $CK[R] \subset \bigcup_{s \in \mathbf{NE}}[s]$ *where* **NE** *is the set of Nash equilibrium profiles of Γ.*

It can be verified that the set of outcomes consistent with common knowledge of rationality in Samet's model is identical to the set of outcomes corresponding to strategies associated with common knowledge of rationality in the standard model of proposition 5 above. Samet's model incorporates a theory of hypotheticals that enables him to impose additional restrictions regarding the "beliefs" at unreached information sets. In our setting the distinction between knowledge and belief was developed to handle just such concerns. As noted above, an additional principle is needed in the current setting to rule out incredible threats as in the imperfect equilibrium of Γ_1. We will call this principle the Weak Caution (WC). We view WC to be an essential element of extensive-form rationality.

Definition (WC) $WC = \bigcap_i \bigcap_{s_{-i}} \neg K_i[\neg[s_{-i}]]$.

5.3 The problematic nature of extensive-form rationality

Consider the game in figure 1 from section 2. This game, or a closely related alternative, is studied by Basu (1990), Ben-Porath (1994), Bicchieri (1989), Binmore (1987–8), Rosenthal (1981), Reny (1993), (1992), and Gul (1995b). The standard argument of backward-induction is as follows: if 2 is rational then he will choose t_3 at his final information set if given a chance. Therefore if 2 is given a chance then he knowing that 1 is rational and being rational himself will choose t_2. Hence, being rational, knowing that 2 knows he is rational and that 2 is rational, 1 should choose t_1 at his first information set. This is the well-known argument of backward induction. All of the authors listed above note the following flaw in this argument: the claim that 2 knows that 1 is rational when she has a chance to move is legitimate only if 2 has no reason to doubt the rationality of 1 when she is reached. The backward induction argument outlined above, shows that when reached, 2 can no longer maintain all the assumptions that yield the backward induction prediction. Basu (1990) and Reny (1993) provide two rather different ways of formalizing this criticism of backward induction. Reny (1993) defines formally what it means for rationality to be common certainty at every relevant information set and proves that except for a rare class of games, rationality cannot be common certainty at every relevant information set. In particular, he shows that rationality cannot be common certainty at player 2's information set in Γ_2. In related work, Basu (1990) shows that there cannot be any solution concept for extensive-form games that satisfy the following apparently plausible restrictions. (1) Rational agents take, at each information set, actions that maximize their payoff. (2) Agents start off certain of the rationality of their opponents and remain certain until actions that are inconsistent with *any* rational strategy are observed. (3) If rational agents observe their opponents take an irrational action then they can no longer rule out any possible action of the irrational opponents. (4) Finally, any course of action which can lead to a payoff at least as high as a possible payoff according to the theory, must be an allowed action for the theory. To see the inconsistency, note that in Γ_2 backward induction cannot be the theory, since if it were and 1 were to choose l_1 then 2 would have to believe that 1 is irrational and hence by (3) might choose l_3 if given a chance. Hence 2 might choose to give him this chance. But this means that 1 might do better by choosing l_1 then t_1. But this contradicts (4) since t_1 is prescribed by the theory and l_1 is not. Thus, any theory satisfying Basu's requirements must allow for player 2's information set to be reached by some rational strategy. But, then (1) and (2) imply that player 2 must choose t_2 which contradicts the rationality of l_1.

Proposition 6 *If v is not on the unique backward induction path of a generic extensive-form game and $[v] \cap CB[R] \neq \emptyset$ in some standard model of this game* **M**, *then there exists $[v']$ such that $[v'] \cap CB[R] \cap \bigcup_i B_i \neg [v'] \neq \emptyset$.*

Proposition 6 is related to Reny (1993). If there is an information set not on the unique backward induction path then it is either common certainty that this information set will not be reached (hence rationality cannot be common certainty once the information set is reached) or there is some other state consistent with the common certainty of rationality where some player is being surprised. That is, one player is choosing a strategy that another player believes will not be chosen. The possibility of this type of surprise is ruled out by Basu (1990) since his axiom (4) implicitly rules out the possibility that the opponent may believe that the information set $[v']$ will not be reached even though reaching it is not inconsistent with $CB[R]$ (i.e., the theory). Note that proposition 6 above is not vacuous. We can construct standard models for generic games in which $[v] \cap CB[R]$ is non-empty even when v is off the backward induction path.

5.4 The resolution: weak caution and common certainty of rationality

Gul (1995b) proposed the solution concept described next as the weakest solution consistent with extensive-form rationality. For any extensive-form game Γ, let $\mathcal{S}_i^e \subset S_i$ be the set of strategies of i that are not strongly dominated at any information set that they reach against conjectures that also reach that information set, and let $\mathcal{S}^e = \Pi_{i \in N} \mathcal{S}_i^e$. Gul's solution concept is $R^e = \mathcal{S}^\infty \mathcal{S}^e$. For generic extensive-form games, the set of strategies $\mathcal{S}_i^e$ corresponds to the set of weakly undominated strategies in the normal-form representation, G, of Γ, and hence $R_i^e = \mathcal{S}^\infty \mathcal{W}(G)$, the strategies that are left after one round of removal of weakly dominated strategies and then iterative removal of strongly dominated strategies. The main result of Ben-Porath (1994) establishes in an axiomatic framework, that in generic perfect information games a strategy profile s is consistent with common certainty of rationality at the beginning of the game if and only if it is in $\mathcal{S}^\infty \mathcal{W}(G)$.

Proposition 7 below is closely related to the ideas of Ben-Porath (1994) and Gul (1995b):

Proposition 7 *In any standard model* **M**, $WC \cap CB[R] \subset [R^e]$. *Moreover, for every extensive game of perfect information there exists some standard model* **M** *such that* $CB[R] = [R^e]$.

5.5 Backward induction revisited

Three recent papers Aumann (1995a), Bonanno (1991), and Samet (1993) have provided formal logical systems for backward induction in perfect information games. Aumann identifies, with each state a behavioral strategy and defines rationality as not knowing that the behavioral strategy is dominated at any information set. He proves that common knowledge of rationality implies and is consistent with backward induction. Bonanno augments the standard model of propositional logic with a description of the given extensive-form game (as a collection of propositions) and a rule of inference that states that if it can be shown that an i permissible hypothesis implies that the choice α is sure to yield a payoff greater than the payoff associated with the choice β, then choosing β is irrational for agent i. (An i permissible hypothesis is one that does not utilize any reference to player i's rationality or utilize in its proof any propositions that invokes i's rationality.) Bonanno proves that adding the axiom "all agents are rational" to this extended propositional logic, yields backward induction as a theorem in certain games, Bonanno also shows that without restriction to i permissible hypothesis, the new system of logic yields a contradiction (i.e., is inconsistent). Samet proves that an assumption that he calls common hypothesis of node rationality yields backward induction. The significant difference between these three formulations and the models discussed above is the fact that the key assumption of rationality in these models has an "ignore the past" feature. To see this note that in Aumann's formulation the relevant objects of analysis are behavioral strategy profiles; rationality of i at node v has force even at a node v that is inconsistent with the rationality of i. Similarly, in Bonanno (1991), the restriction to i permissible hypotheses ensures that when making deductions about the behavior of agent i at a node v only the behavior of agents at successor nodes are relevant since deductions about the behavior of predecessors will typically involve conclusions about i's own behavior. The same effect is achieved by Samet's common hypothesis of the node rationality assumption. In the reasoning of each agent or the reasoning of an agent about the reasoning of any future agents, observed past deviations are assumed to be irrelevant for future expectations. As suggested by Selten (1975), implicit in the idea of backward induction is this notion that past deviations are "mistakes" that are unlikely to be repeated in the future. Such a flavor is present in all three of the results mentioned above. In our setting, we can utilize the distinction between knowledge and belief to achieve this effect. Specifically, we will characterize backward induction by the requirement that agents have doubts about behavior at predecessor nodes and hence are (weakly) cautious. On the other hand, we will assume that rationality of successors is common

knowledge. Our aim is not to strengthen the case for backward induction but to provide a unified framework for evaluating both the criticisms of backward induction and its axiomatizations.

We need the following notation. Let s_M denote a subprofile of strategies, one for each $i \in M \subset N$, and let S_M and s_M be defined in a similar fashion. Finally, let $R_M = \cap_{i \in M} R_i$.

Definition (CK of rationality of non-predecessors) $CK[R_{np}] = \bigcap_{E \in \Psi} E$ where $\Psi = \{CK_M[R_{M'}]$ for all $M, M' \subset N$ such that M' contains no player who moves at a successor node to some node owned by a player in $M\}$.

Definition (Weak caution regarding predecessors) $WC_p^i = \bigcap_{s_M} \neg K_i \neg [s_M]$ where M is the subset of players that own an information set preceding an information set owned by i; $WC_p = \cap_i WC_p^i$.

Proposition 8 *If* **M** *is a standard model for a generic extensive-form game in which each player moves once, then* $WC_p \cap CK[R_{np}] \subset [s^o]$, *where* s^o *is the unique backward induction strategy profile. Moreover, for any such game there exists some standard model such that* $WC_p \cap CK[R_{np}] = [s^o]$.

6 WEAKENING THE NOTION OF CERTAINTY: THE BAYESIAN APPROACH

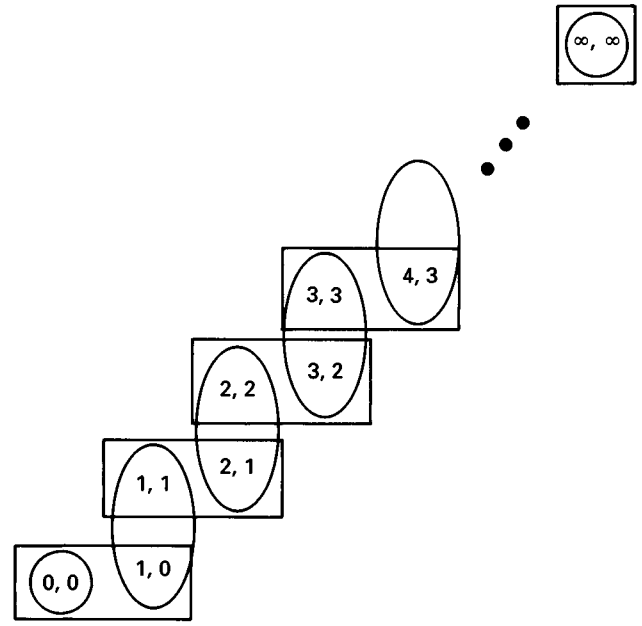
Rubinstein (1989) provides a provocative example of a game with a move by Nature, similar to the information structure and payoffs in figure 5.8 below. Rubinstein's story for this information structure is that either game a or b is played, and it is a with probability greater than 1/2. If it is b, 1 is so informed and a signal is sent from 1 to 2, and a confirmation from 2 to 1, and a (re)confirmation from 1 to 2, etc. Confirmations are lost with probability ε, and once a confirmation is lost the process ends. Players observe how many signals they have sent (which for player 1 equals the number received plus one, and for player 2 is equal to the number received). One can readily verify Rubinstein's result that with $\varepsilon > 0$ the only Nash equilibrium in which (U, L) is played when the game is a, has (U, L) played when the game is b as well. (More precisely, the only Nash equilibrium in which player 1 plays U when not informed that the game is b, has the players choosing (U, L) in every state of Nature except a state that has zero probability when $\varepsilon > 0$, namely $\omega = (\infty, \infty)$.) By contrast consider the case where there are no doubts about which game is being played, e.g., if $\varepsilon = 0$ there are no doubts in the only two states that have positive probability: $(0, 0)$ and (∞, ∞). In this case there is an equilibrium where (U, L) is played when the game is b, and (D, R) if it is a.

Clearly, it is common certainty at state (∞, ∞) that the game is b; similarly, when $\varepsilon = 0$ the probability that the game will be commonly certain is 1. Moreover, $(n, n+1) \in B_N^n([b])$, i.e., at $(n, n+1)$ the players are certain that . . . [n times] . . . that they are certain the game is b. Therefore, it might appear that when ε is small and/or at states $\omega = (n, n+1)$ for n large, there is "almost" common certainty of the game.[47]

This suggests that when there is "almost" common certainty that the game is G, the equilibria may be very different from the case when it is commonly certain that the game is G.

The example above generated a literature examining the robustness of solution concepts to weakening common-certainty assumptions.[48] In the current section we review this literature and relate it to the characterizations in the preceding sections. While previously we assumed Ω is finite, clearly we now need to allow Ω to be infinite; however, we restrict attention to countably infinite Ω.

In subsection 6.1 we define various notions of almost common certainty, and in subsection 6.2 we provide characterizations results. First, we examine the robustness of the normal-form characterizations of section 4. The main message here is that for results to be robust, common certainty can be weakened to "almost common 1 belief" (defined below), which is not satisfied by the example of figure 5.8. Next we use a notion of almost common certainty – which avoids the conflict between certainty and caution discussed at the end of section 4 – to characterize refinements. The main result here is that almost common certainty of caution, rationality, and of the game yield the same solution concept we saw in section 5: $\mathscr{S}^\infty \mathscr{W}$. Subsection 6.3 examines refinements from a different perspective from the rest of this chapter. Instead of axiomatically characterizing solution concepts, we examine directly the robustness of solution concepts to a particular perturbation of the game. Given a game G, we consider applying the solution concept to a game with a move by Nature, $\tilde{G}$; we focus on those games $\tilde{G}$ in which it is almost common certainty that the game is in fact G. For example, $\tilde{G}$ could be the game of figure 5.8, G could be the game b, and we could look at states in $\tilde{G}$ where G is "almost" common certainty. We then ask what is the relationship between the solution of G and the solution of $\tilde{G}$ in those states. That is, we are asking whether the solution concept is robust to assuming the game is almost common certainty, instead of common certainty as is typically assumed. Our interest in these refinements issues is enhanced by the fact that both the methods and the results are closely related to the characterizations in section 6.2. There are two types of questions we consider when examining solution concepts. First, in subsection 6.3.1, we ask whether, for a given solution concept, an outcome that is excluded by that solution concept in game G would still be excluded by the



	L	R
U	1, 1	0, –3
D	–3, 0	0, 0

Game *a*

	L	R
U	0, 0	0, –3
D	–3, 0	1, 1

Game *b*

Figure 5.8 Rubinstein's e-mail game. In state (0, 0) game *a* is played, in all other states game *b* is played

same concept in $\tilde{G}$ when the game is almost common certainty. The main conclusion here is that rationalizability with caution is the tightest refinement of rationalizability that is robust in this sense.[49] In subsection 6.3.2 we ask the converse. Will an outcome that is accepted by a solution concept for a particular game, continue to be accepted by the same solution concept if the game is only almost common certainty? Here we show first that strict Nash equilibrium (and ε Nash equilibrium) are robust in this sense, but, as in subsection 6.2, only with the notion of "almost common 1 belief"; and, second, other solution concepts are robust to weaker notions of

almost common certainty. Thus there will be several notions of almost common certainty, and various questions which can be asked using each of these notions. We restrict attention to a few of the most interesting results rather than provide an exhaustive analysis of each notion and question.

6.1 Notions of almost common certainty

One possible definition is based on the number of levels for which mutual certainty holds. That is, E is nth order mutual certainty at ω if $\omega \in B_N^n(E)$, and E is almost ∞ certain if n is large. One might well expect that almost ∞ certainty is quite different from common certainty, since no finite n is similar to ∞. As Rubinstein (1989) observed this is familiar from repeated games where large finite repetitions is quite distinct from infinite repetitions.

All the other notions we use weaken the certainty aspect of the notion of common certainty. Instead of requiring probability 1, we require only probability close to 1. For example, common certainty of an event E is when everyone assigns probability 1 to . . . everyone assigning probability 1 to E. So Monderer and Samet (1989) (see also Stinchcombe (1988)), consider common q belief of E – everyone assigns probability at lest q to . . . everyone assigning probability at least q to E. To formalize this let $B_i^q(E) = \{\omega: p_i(E \mid \mathcal{F}_i(\omega)) \geq q\}$ denote the set of states in which i assigns probability at least q to E.[50] Then E is common q belief at ω if $\omega \in \cap_{n=1}^{\infty}(B_N^q)^n(E)$.[51] The event that E is common q belief is denoted $C^q(E)$.[52] We will say that E is almost common 1 belief at ω, if E is common q belief at ω, for q close to 1. This is a very demanding notion: in the example of figure 5.8 for no state $\omega \neq (\infty, \infty)$ and not event E that is a strict subset of Ω is there almost common 1 belief of E at ω. However, we will see that it is the appropriate notion for obtaining robustness of our characterizations.

Both preceding notions directly weaken the definition of an event being common certainty at a state ω; the former weakens the number of iterations, the latter the degree of certainty. Thus they do not require a prior, only conditional probabilities, as we argued for in section 3. Therefore, they will not be useful for examining solution concepts that are based on common prior, or more generally, for considering solution concepts from an *ex ante* perspective. One possible *ex ante* notion involves a strengthening of the notions above. In particular, continuing with our approach of assuming probability close to 1 instead of equal to 1, we may want to assume that almost surely an event, say E, is almost common certainty. Formally, consider the assumption that everyone assigns probability at least q to the event that E is common q belief: $p_i(C^q(E)) > q$. When this is true for q close to 1 we say that E is almost certainly almost common 1 belief.

An *ex ante* notion that is weaker is used by Kajii and Morris (1995). If there is a common prior, then $p(E) = 1$ implies that $p(C(E)) = 1$: if E is certain then it is certain that E is common certainty. Then one could say that E is almost certain if $p(E)$ is close to 1, and one could explore this as an alternative notion of being close to common certainty (since it is close to a sufficient condition for common certainty). However, this is not a notion of "almost common certainty" when there is no common prior, since without a common prior even the extreme case of $p_i(E) = 1$ for all i does not imply that E is commonly certain (so naturally $p_i(E)$ close to 1 for all i need not look at all like common certainty). To see the problem consider the following information structure; $\Omega = \{a, b, c, d\}$, $\mathscr{F}_1 = \{\{a\}, \{b\}, \{c, d\}\}$, $\mathscr{F}_2 = \{\{a, b\}, \{c\}, \{d\}\}$, $p_1^n = (0, 1 - 1/n, 0, 1/n)$, $p_2^n = (1/n, 0, 1 - 1/n, 0)$, and let $E = \{b, c\}$. While 1 and 2 are both almost certain that E occurs, each is almost certain that the other is almost certain that E does not occur. So, in the limit (using the conditional probabilities given by the limit of the conditionals), E is subjectively certain but not common certainty. Nevertheless, for examining robustness of conclusions, one might be interested in the notion of (subjective) almost certainty. This is because there may be occasions where we think our assumptions (about the payoffs, rationality, etc. of the players) are almost certainly correct, but we may not be sure that they are almost common 1 belief.[53]

The relationship among the last two, *ex ante*, notions is worth clarifying. Clearly if E is almost certainly almost common 1 belief then E is almost certain, and in general the reverse implication is false. However, if there is a common full support prior and Ω is finite, then the reverse implication does hold; see Fudenberg and Tirole (1991, theorem 14.5) and Kajii and Morris (1995, equation 7.1).[54]

6.2 Robustness of characterizations and introducing caution

We begin by examining the robustness of the characterizations in section 4. First we formally state that proposition 1 is robust to weakening common certainty to almost common 1 belief.[55]

Proposition 9 *Fix a game $G = (S, u)$. There is a $\bar{q} \in (0, 1)$ such that given a model $\{\Omega, \mathscr{F}_i, p_i, \mathbf{s}, \mathbf{u}\}$ if $q > \bar{q}$ then $C^q([u] \cap [\text{rationality}]) \subset [\mathscr{S}^\infty(u)]$.*

The proof of this result is the same as the iterative part of the proof of proposition 12.

Aumann's characterization of correlated equilibrium is also robust: a common prior and almost certainty of rationality and of the game characterize almost correlated equilibrium.[56]

Proposition 10 *Consider a sequence of models, which differ only in the common prior:* $\{\Omega, \mathcal{F}_i, p^n, \mathbf{s}, \mathbf{u}\}_{n=1}^{\infty}$. *Assume that* $\mathbf{u}$ *is bounded, i.e.,* $\mathbf{u}(\omega)(s) < b \in \Re$ *for any* ω *and all* s.[57] *Let* E^n *be the event that the payoffs are* u *and the players are rational in model* n.[58] *If* $p^n(E^n) \to 1$, *then the limit of any convergent subsequence of the distribution on actions induced by* p^n *and* $\mathbf{s}$ *is a correlated equilibrium distribution.*

This is an upper hemi continuity result which can be proven along lines similar to those used by Kajii and Morris (1995, theorem 3.2) (see also the end of subsection 6.3.2 below).[59]

It is interesting that the results suggest that the characterization of correlated equilibrium does not really rely on common certainty of rationality or of the game. In the context of section 4, the characterization can be stated using the assumption either that rationality and the game have prior probability 1, or, that at every non-null ω the event $[rationality] \cap [u]$ is common certainty; the two assumptions are equivalent with a common prior. However, here we saw that we can make do with almost certainty of rationality and of the game, and we do not need the assumption of almost certainty that $[rationality] \cap [u]$ is almost common 1 belief at each non-null ω. As we mentioned at the end of subsection 6.1, the latter assumption *is* stronger than almost certainty of $[rationality]$ and of $[u]$.[60]

We conclude our analysis of robustness of characterizations of solution concepts by considering Nash equilibrium. We show below that the characterization of Nash equilibrium for two-person games in proposition 4 is robust if mutual certainty at ω is weakened to conditional probability close to 1 at ω. The characterizations of Nash equilibrium for games with $n > 2$ players are similarly robust, if common certainty of the conjectures at ω is weakened to almost common 1 belief of the conjectures at ω.

Proposition 11 *Consider a sequence of models* $\{\Omega, \mathcal{F}_i, p_i^n, \mathbf{s}, \mathbf{u}\}_{n=1}^{\infty}$, *assume* $N = \{1, 2\}$, $\mathbf{u}$ *is bounded and fix an* ε. *There is a* δ *such that if the utility functions, rationality, and the conjectures are almost certain at* ω, *i.e.,* $p_i([u] \cap [\phi_1] \cap [\phi_2] \cap [rationality] \mid \mathcal{F}_i(\omega)) = 1 - \delta$, *then* $(\phi_1, \phi_2))$ *is an* ε *Nash equilibrium in the game* $G = (S, \mathbf{u}(\omega))$.

The notion here of an ε-Nash equilibrium is the standard one; ϕ_i is within ε of a best reply to ϕ_j.[61] The idea of the proof is the same as for the case of common certainty, proposition 4.[62]

To summarize, all the characterization results are robust to weakening certainty to probability close to 1, e.g., weakening common certainty at ω to common almost 1 belief at ω.

We now introduce caution into the model. Recall that $[caution]$

$= \cap_i \cap_{s_{-i}} \neg B_i \neg [s_{-i}]$; thus if a player is cautious at ω then her conjectures assign strictly positive probability to every action of her opponents: ϕ_{-i} has support S_{-i}. The main result of this section is that, if rationality, the game G, and caution are almost common 1 belief, then players choose strategies in $\mathscr{S}^\infty \mathscr{W}(G)$; we called this solution concept rationalizability with caution. That is, to the extent that "admissibility" can be made common knowledge, it does not characterize iterated deletion of weakly dominated strategies, $\mathscr{W}^\infty$; rather it characterizes rationalizability with caution. Similarly, applying caution to Nash-equilibrium characterizations yields Nash equilibrium in strategies that are not weakly dominated, not any stronger refinement.[63]

Proposition 12 (Börgers (1994)) *Fix a game G. There is a $\bar{q} \in (0, 1)$ such that, given a model $\{\Omega, \mathscr{F}_i, p_i, \mathbf{s}, \mathbf{u}\}$, if $q > \bar{q}$ then $C^q([u] \cap [caution] \cap [rationality]) \subset [\mathscr{S}^\infty \mathscr{W}(u)]$.*

The idea of the proof is as follows. Each i's strategy is a specification of what to do conditional on each information cell. We consider two types of cells – the cell including ω, denoted by F_i and the other cells. At ω each player i knows her payoffs are u. Caution then implies that everyone chooses a strategy that, conditional on F_i, specifies something that is not weakly dominated in $G = (u, S)$. If q is large enough, then i believes with high probability that player j is choosing a strategy that survives one round of deletion of weakly dominated strategies in G. Being rational at ω, we deduce that i chooses a strategy that is a best reply to such a belief. For q close enough to 1, conditional on F_i, $\mathbf{s}(\omega)$ cannot specify that i choose something that is strongly dominated in the restricted game $\mathscr{W}(G)$. (If it did then i is not rational at ω since i could do better by avoiding the dominated strategy.) Now iterate on the last step.[64]

6.3 Robustness of solution concepts

In this subsection we view a model as a game with a move by Nature, which we solve using various solution concepts. With this interpretation, the game is given by $(\Omega, \mathscr{F}_i, p_i, \mathbf{u})$, and $\mathbf{s}$ will be a strategy profile, of the players in this game, satisfying some solution concept. We assume for the remainder of this section that $\mathbf{u}$ is bounded. We consider sequences of games parametrized by probabilities, p_i^n, and we denote the solution of the nth game by $\mathbf{s}^n$. The question we ask is what solution concepts are robust to the game being almost common certainty, and for which notions of almost common certainty does this robustness hold?

How would the different notions of almost common certainty be used to

examine this question? Let $\tilde{G}$ be a game with a move by Nature: $(\Omega, \mathscr{F}_i, \tilde{p}_i, \mathbf{u})$, and $G = (S, u)$. If $\omega \in C^q([u])$ for q close to 1, then in $\tilde{G}$ at ω there is almost common 1 belief that the game is G. Similarly, if $\tilde{p}_i(C^q([u])) > q$ for q close to 1, then in $\tilde{G}$ the game G is almost certainly almost common 1 belief. Finally, if $\tilde{p}_i([u]) > q$ for q close to 1, then we say that in $\tilde{G}$ the game G is almost certain.

Next we need to ask how we formally compare the solution of $\tilde{G}$ with the solution of G. Strategies for i in $\tilde{G}$ are not elements of S_i; they are functions from i's private information $\mathscr{F}_i$ into S_i. What we compare is, for example, the specification of the equilibrium strategy in $\tilde{G}$ *in state* ω, at which G is almost common 1 belief, with the equilibrium of G. More generally, we compare the distribution over S given by the equilibrium conditional on an event where u is almost common certainty. (More formal specifications will be given below.)

In addressing the question of this subsection, there is an issue concerning what one means by comparing the case of almost common certainty with the common certainty situation. Fixing a model where some event is common certainty, there are many "nearby" models in which that event is almost common certainty. Should we ask: What is the outcome that survives in all these "nearby" models, or in some? More specifically, one might ask whether strategy profiles that are excluded by a solution concept when the game G is common certainty are also excluded for all games $\tilde{G}$ within which G is almost common certainty. Alternatively one might wonder if strategy profiles that are accepted by our solution concept are also accepted in all possible games $\tilde{G}$ in which G is almost common certainty. The former perspective – which amounts to comparing the solution of G with the union of the solution concept over all possible games $\tilde{G}$ in which G is almost common certainty – is appropriate for questioning refinements: we should not exclude an outcome in a given, commonly certain, game G if that outcome is plausible in some "nearby" game when G is almost common certainty. This is the motivation underlying the research initiated by Fudenberg, Kreps, and Levine (1988).[65] The second perspective – which amounts to comparing the solution of G with the intersection of the solutions of all games $\tilde{G}$ in which G is almost common certainty – is more in the spirit of obtaining new solution concepts. We should not feel comfortable accepting a prediction of an outcome in a particular environment if there is some "nearby" environment where the game is almost common certainty and all the assumptions underlying our solution concept are satisfied, but the prediction in this "nearby" environment is very different from the original prediction. This is closer to the motivation of Monderer and Samet (1989) and Kajii and Morris (1995).

We will be applying the solution concepts explicitly here, rather than

axiomatically. This gives us greater flexibility in considering solution concepts, but involves a different interpretation: we are not trying to justify the solution concepts, but are investigating whether outcomes are excluded or accepted in a robust manner. An example of the significance of this different interpretation of results will be discussed in subsection 6.3.2 below.[66]

6.3.1 *Almost common certainty and excluding outcomes robustly*

Following Fudenberg, Kreps, and Levine (1988), the first version of this robustness question argues that it is unreasonable to exclude particular outcomes if our solution concept when applied to G yields very different outcomes than when it is applied to a game with a move by Nature in which G is almost common certainty. For example, we might think that any strict Nash equilibrium is a reasonable prediction when a game is common certainty among the players. However, since we the analysts might only know that some game G is almost common certainty, it would be unreasonable to rule out any outcomes which are strict Nash equilibria in a game of incomplete information where G is almost common certainty (and the game of incomplete information is assumed to be common certainty among the players).

While the focus is different there is an obvious connection between this question and the characterization results of the previous section. There we also asked whether our axioms excluded behavior which would not be excluded if the game were almost common certainty, rather than common certainty. For example, a corollary to proposition 9 is that $\mathscr{S}^\infty$, applied to a game $\tilde{G}$ and evaluated at a state ω at which G is almost common 1 belief, yields a subset of $\mathscr{S}^\infty$ applied to G. (Moreover, taking the union over all such games $\tilde{G}$ yields all of $\mathscr{S}^\infty(G)$, since we can always take the game $\tilde{G}$ to be equal to G.) Thus the previous subsection showed that $\mathscr{S}^\infty$ is robust (to almost common 1 belief) in the first sense considered in this subsection.[67]

A similar connection to the previous section is via proposition 12. That result suggests that if we are given a game $\tilde{G}$, in which a particular game G is almost certainly almost common 1 belief, then applying the solution concept of rationalizability with caution to $\tilde{G}$, evaluated at those states where G is almost common 1 belief, yields a subset of the same concept applied to G; roughly speaking this says that $\mathscr{S}^\infty\mathscr{W}(\tilde{G}) \subset \mathscr{S}^\infty\mathscr{W}(G)$. Moreover, once again, it is trivial that there is a game $\tilde{G}$, in which G is almost certainly almost common 1 belief, such that, evaluated at those states where G is almost common 1 belief, $\mathscr{S}^\infty\mathscr{W}(\tilde{G})$ yields exactly $\mathscr{S}^\infty\mathscr{W}(G)$ – simply take $\tilde{G} = G$.

Thus, rationalizability with caution is robust in the sense that, when applied to a game G, it only excludes strategy profiles that would be excluded when applied to any game in which G is almost certainly almost common 1 belief. One might also ask if there are any tighter solution concepts that are robust in this sense. Dekel and Fudenberg (1990) show that this is essentially the tightest non-equilibrium solution concept. Formally they show that the union of solutions of all games in which G is almost certainly almost common 1 belief using iterated deletion of weakly dominated strategies yields rationalizability with caution. The proof is similar to the arguments already given.

Dekel and Fudenberg (1990) also show how this robustness test corresponds to issues that have come up in analyzing extensive-form games. The robustness question examines what happens if payoff are almost common 1 belief. Thus, in particular, in an extensive-form game, it allows players to change their beliefs about the payoffs in a game if they are surprised by an opponent's action. This formally corresponds to the issues raised in criticism of backwards induction: if a player is surprised and observes non-backward-induction play, what should she believe? The robustness test allows her to believe that the payoffs are different than she previously thought. (Or, equivalently, revise her view of the rationality of the opponents. These are equivalent since there is no way to distinguish irrationality as it is used here from different payoffs.) This explains why the solution concept $\mathscr{S}^\infty\mathscr{W}$, which was introduced when Dekel and Fudenberg (1990) analyzed the robustness of solution concepts, is the same as the one achieved using common certainty of rationality in extensive-form games (see section 5).

6.3.2 *Almost common certainty and accepting outcomes robustly*

We now turn to the second, and opposite, notion of robustness: an outcome should be accepted as a prediction for game G only if it would be accepted in all games $\tilde{G}$ in which G is almost common certainty. For brevity and simplicity, we restrict attention to strict Nash equilibria of G in this discussion. As before, consider a game G, and a sequence of games G^n in which G is, in one of the senses formally defined earlier, almost common certainty. Roughly speaking, we want to know if equilibria of G are also equilibria in all such sequences G^n.

We first consider the case in which G is common q^n belief at ω, where $q^n \to 1$. This is the case considered by Monderer and Samet (1989). Their results imply that a strict Nash equilibrium s of G is robust in that for any G^n there is always a Nash equilibrium which plays s on states where G is almost common 1 belief. Moreover, the interim payoffs are the same and if almost

common 1 belief of G is almost certain then the *ex ante* payoffs converge as well.[68]

But are strict Nash equilibrium robust when all we know is that G is almost certain? This is the question asked by Kajii and Morris. They show that the answer is no – there exists a generic game G with a *unique* strict Nash equilibrium s, and a sequence G^n of games in which G is almost certain, $p^n([u]) \to 1$, with a unique rationalizable strategy profile $\mathbf{s}$, s.t. the distribution of actions generated by $\mathbf{s}$ on the event $[u]$ is not s.[69] Kajii and Morris go on to examine sufficient conditions for a Nash equilibrium of a game G to be robust in this strong sense. They obtain two types of sufficient conditions. The first is that a unique correlated equilibrium is robust. This follows from an upper-hemi continuity argument related to the one we gave following proposition 10. The limit of the distribution of actions in the game G^n, conditional on the event $[u]$, where the game G^n is solved by Nash equilibrium, must be a correlated equilibrium distribution of G if G is almost certain in G^n. Then, if there is a unique correlated equilibrium of G, these limits must all be converging to that correlated equilibrium distribution, so it is robust.

The preceding paragraph may help clarify the relationship between this subsection, which takes solution concepts as given, and subsection 6.2, which looks at characterizations. Both the characterization of correlated equilibrium, and the solution concept itself are robust: this is the message of proposition 10 and the Kajii and Morris result discussed above. But proposition 11 suggests that the characterization of Nash equilibrium is robust, while the cited example of Kajii and Morris suggests that Nash equilibrium is not robust to allowing almost common certainty. Why is it that their example does not bear a similar relationship to proposition 11 as does the relationship of their robustness of unique correlated equilibrium to proposition 10? The answer is complex since the analysis of robustness of characterizations and of the equilibrium concept are different in many ways. However, by considering how one could modify the models to make them more similar, it appears that there is a basic incompatibility between the lack of robustness of Nash equilibrium as in Kajii and Morris on the one hand, and the robustness of characterizations of Nash equilibrium. This incompatibility may raise doubts about the robustness test used by Kajii and Morris for Nash equilibrium.[70]

Kajii and Morris have a second sufficiency condition for robustness of Nash equilibrium. While the general version requires too much setup for this survey, a very special case of their second result is useful for drawing connections to the literature and to examples above. The restriction of their results to two-person two-strategy games with two strict Nash equilibrium, implies that the robust outcome is the risk-dominant one. This is exactly

what happens in the Rubinstein example at the beginning of the section, and a similar result of this kind was obtained by Carlsson and van Damme (1993) for 2×2 games when players observe their payoffs with noise.[71]

6.4 Necessity of almost common 1 belief for robustness

The final issue we consider in this subsection is the question of necessity. We saw that almost common 1 belief is sufficient for robustness of strict Nash equilibrium, while almost certainty is sufficient for robustness of a unique correlated equilibrium. Is almost common 1 belief necessary for the former result? Monderer and Samet (1990), and subsequently, Kajii and Morris (1994a), show that one cannot in general weaken the hypothesis of almost common 1 belief.[72] As noted, the interest in this result is that it shows that the sufficiency of almost common 1 belief for robustness presented throughout this section is, in some sense, necessary for lower hemi continuity of the equilibrium strategies and payoffs.

7 WEAKENING THE NOTION OF KNOWLEDGE: THE SYNTACTIC APPROACH

7.1 Generalized knowledge operators and possibility correspondences

In section 3 we observed that any knowledge operator $K_i: 2^\Omega \to 2^\Omega$ satisfying [MC, N] generates a possibility correspondence $\mathcal{F}_i: \Omega \to 2^\Omega$, and conversely. More precisely, using the definitions in section 3.1, starting from either K_i (or $\mathcal{F}_i$), going to $\mathcal{F}_i$ (or K_i), and then going back, one ends up with the same operator as one started with. Moreover, we saw that if the knowledge operator satisfies [T],[4], and [5] as well, then $\mathcal{F}_i$ is a partition.

Many authors have questioned the appropriateness of [T] and of [5] for modeling knowledge and decision making.[73] As we discussed in subsection 3.3 above, [T] is often weakened to [D] to model belief rather than knowledge, and the importance of this weakening for modeling weak caution in extensive-form games is explained in section 5.[74] In this section we focus on dropping [5].

Alternatively, some authors have directly questioned the appropriateness of partition for modeling information. They consider various properties of the possibility correspondence, all of which are satisfied by partitions, and examine the effect of weakening them and the relationship between these properties and the axioms on K_i. In this section we will be discussing these two approaches; in the current subsection we present the basics. In subsection 7.2, as a first view of these models, we show the implication of

weakening the knowledge axioms, and the partition structure of information, in single-person decision theory. We consider games in subsection 7.3; we show that without a common prior there is a sense in which these weakenings have no effect, therefore the only insights these weakenings may yield are in the context of assuming the CPA. In subsection 7.4 we present the most widespread application of these weakenings, namely to no-trade results. (Readers for whom the no-trade results are new may want to look at the appendix, where we review the analogous results for the case where partitions are used.) We conclude in subsection 7.5, where we discuss severe problems with the motivation for most of these results, especially applications such as those in subsection 7.4.

The remainder of this subsection presents the relationship between weakening axioms on a knowledge operator and weakening the partition structure of information. The following are some properties of possibility correspondences, $\mathcal{F}_i$, which have appeared in the literature.

P1 (reflexive) $\omega \in \mathcal{F}_i(\omega)$

P2 (nested) For F and F' in $\mathcal{F}_i$, if $F \cap F' \neq \phi$ then either $F \subset F'$ or $F' \subset F$.

P3 (balanced) For every self-evident E, i.e., E s.t. $\omega \in E \Rightarrow \mathcal{F}_i(\omega) \subset E$, there exists $\lambda_E : \mathcal{F}_i \to \Re$ s.t. $\forall \omega \in \Omega, \Sigma_{F:\omega \in F \in \mathcal{F}_i} \lambda(F) = 1$ if $\omega \in E$ and 0 otherwise. We say that $\mathcal{F}_i$ is positively balanced and satisfies [P3$^+$] if λ_E can be taken to be non-negative.

P4 (transitive) $\omega'' \in \mathcal{F}_i(\omega')$ and $\omega' \in \mathcal{F}_i(\omega) \Rightarrow \omega'' \in \mathcal{F}_i(\omega)$.

P5 (Euclidean) $\omega' \in \mathcal{F}_i(\omega)$ and $\omega'' \in \mathcal{F}_i(\omega) \Rightarrow \omega'' \in \mathcal{F}_i(\omega')$.

Possibility correspondences satisfying several of these properties can be seen in figure 5.9. The connection among the axioms on K_i and properties of $\mathcal{F}_i$, where K_i and $\mathcal{F}_i$ are related as discussed above, are presented in proposition 13 below. The most interesting parts are (i)–(iii), which relate K_i to $\mathcal{F}_i$; the rest are useful properties for understanding the results below.

Proposition 13

(*i*) K_i *satisfies* [T] *if and only if* $\mathcal{F}_i$ *satisfies* [$P1$].
(*ii*) K_i *satisfies* [4] *if and only if* $\mathcal{F}_i$ *satisfies* [$P4$].
(*iii*) K_i *satisfies* [5] *if and only if* $\mathcal{F}_i$ *satisfies* [$P5$].
(*iv*) *If* $\mathcal{F}_i$ *satisfies* [$P1$] *and* [$P2$] *or* [$P1$] *then it satisfies* [$P3$].
(*v*) *If* $\mathcal{F}_i$ *satisfies* [$P1$, $P2$, *and* $P4$] *then it satisfies* [$P3^+$].
(*vi*) *If* $\mathcal{F}_i$ *satisfies* [$P1$] *and* [$P5$] *then it satisfies* [$P4$].
(*vii*) $\mathcal{F}_i$ *satisfies* [$P4$] *if and only if* $\omega' \in \mathcal{F}_i(\omega) \Rightarrow \mathcal{F}_i(\omega') \subset \mathcal{F}_i(\omega)$.

For proofs see Chellas (1980) and Geanakoplos (1989).

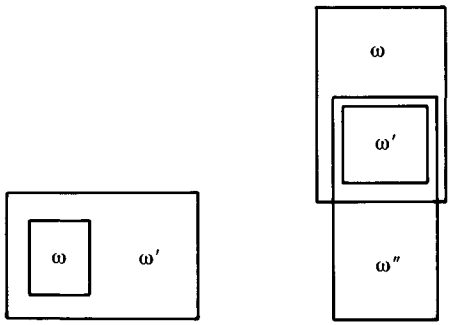


Figure 5.9 Examples of possibility correspondences that are not partitions.
(*a*) $\mathcal{F}(\omega) = \{\omega\}$, $\mathcal{F}(\omega') = \{\omega, \omega'\}$. [P1]–[P4] are all satisfied, [P5} is not.
(*b*) $\mathcal{F}(\omega) = \{\omega, \omega'\}$, $\mathcal{F}(\omega') = \{\omega'\}$, $\mathcal{F}(\omega'') = \{\omega', \omega''\}$. [P1], [P3], and [P4] are satisfied, [P2} and [P5] are not.

7.2 An application to single-person decision theory: When is "information" valuable?

A single-person decision problem with information is determined by an information structure $\{\Omega, p_i, \mathcal{F}_i\}$, a strategy space S_i, and utilities $u_i: S_i \times \Omega \to \Re$. Given any information $F \in \mathcal{F}_i$ the player can choose any element of S_i, so a decision rule is a function $\mathbf{s}_i: \mathcal{F}_i \to S$, which determines a function from Ω into S_i that we also (with abuse of notation) denote by $\mathbf{s}_i$, where $\mathbf{s}_i(\omega) \equiv \mathbf{s}_i(\mathcal{F}_i(\omega))$. Given any $F \in \mathcal{F}_i$, the player's interim expected utility is naturally defined as $Eu(\mathbf{s}_i \mid F) \equiv \Sigma_{\omega' \in F} u(\mathbf{s}_i(\omega), \omega') p_i(\omega')/p_i(F)$. However, if $\mathcal{F}_i$ is not a partition then it is not clear how to define *ex ante* utility. The *true ex ante* utility is $Eu(\mathbf{s}_i \parallel \mathcal{F}_i) \equiv \Sigma_{\omega \in \Omega} p_i(\omega) Eu(\mathbf{s}_i \mid \mathcal{F}_i(\omega))$. On the other hand, the expectation of the interim utilities, $\Sigma_{F \in \mathcal{F}_i} p_i(F) Eu(\mathbf{s}_i \mid F)$, may be different. But this latter notion does not make sense: if a person thought she had a non-partitional $\mathcal{F}_i$, then she would refine it to a partition. For example, if she thought that in state $\omega' \in \mathcal{F}_i(\omega)$ the information received were different from $\mathcal{F}_i(\omega)$, which must happen for some ω and ω' if $\mathcal{F}_i$ is not a partition, then she would not think that ω' is possible when ω happened. So $\mathcal{F}_i(\omega)$ would not be the possibility set for ω. For this reason, when working with non-partitional possibility correspondences, $\mathcal{F}_i$, it is often said that the players do not know their information structure. Therefore, the notion of *ex ante* expected utility should be thought of as something the analyst considers and not the player, and as such we focus on the true *ex ante* utility $Eu(\mathbf{s}_i \parallel \mathcal{F}_i)$.

A partition $\mathscr{F}_i$ is more informative than $\mathscr{F}'_i$ if the former is a refinement of the latter, i.e., if for all $\omega, \mathscr{F}_i(\omega) \subset \mathscr{F}'_i(\omega)$. It is well known that the *ex ante* expected utility increases when a partition becomes more informative: if $\mathscr{F}_i$ is a more informative partition than partition $\mathscr{F}'_i$, then $Eu(\mathbf{s} \parallel \mathscr{F}_i) \geq Eu(\mathbf{s} \parallel \mathscr{F}'_i)$. Does this result hold if information is not given by a partition? Equivalently, does this result hold if some of the axioms of knowledge are dropped?

Proposition 14 (Geanakoplos (1989)) *If a possibility correspondence $\mathscr{F}_i$ satisfying* [P1, P2, *and* P4] *is a refinement of a partition $\mathscr{F}'_i$ then $Eu(\mathbf{s} \parallel \mathscr{F}_i) \geq Eu(\mathbf{s} \parallel \mathscr{F}'_i)$. Moreover, if $\mathscr{F}_i$ fails* [P1, P2, *or* P4] *then there is a partition $\mathscr{F}'_i$ that is less informative but yields higher ex ante expected utility.*

Thus [P1, P2, P4] are necessary and sufficient for information to be valuable.[75]

7.3 Solution concepts with general possibility correspondences

In order to examine the effect of allowing for non-partitions in multi-person environments, we need to extend and re-examine our definitions of common knowledge and of equilibrium. As before, the basic definition of common knowledge of E at ω is that all players know E, know that they know E, etc., and a person knows E at ω if $\mathscr{F}_i(\omega) \subset E$.[76] A (Bayesian) Nash equilibrium in a game with a non-partitional information structure, will also be defined as in the usual case.

Definition Given an information structure $(\Omega, (\mathscr{F}_i, p_i)_{i \in N})$ and a game $G = (S_i, u_i)_{i \in N}$, where $u_i: \Omega \times S \to \mathfrak{R}$, a decision rule for each player is, as above, a function $\mathbf{s}_i: \mathscr{F}_i \to S_i$. A profile $\mathbf{s} = (\mathbf{s}_i)_{i \in N}$ is a Nash equilibrium if for all i, and for all ω, i prefers $\mathbf{s}_i(\omega)$ to any other s_i: for all i, $F \in \mathscr{F}_i$, and s_i.

$$\sum_{\omega' \in F} u(\mathbf{s}(\omega), \omega') p_i(\omega')/p_i(F) \geq \sum_{\omega' \in F} u(s_i, \mathbf{s}_{-i}(\omega), \omega') p_i(\omega')/p_i(F).$$

What are the implications of these weakenings of knowledge for solution concepts? This line of research has only been explored to a limited extent; the main result is that a posteriori equilibria continue to correspond to $\mathscr{S}^\infty$, even when the information structures of the correlating devices violate all of the properties above, so long as each player has a well-defined "conditional" probability given each information she can receive.

Proposition 15 (Brandenburger, Dekel, and Geanakoplos (1992)) *Consider a Nash equilibrium, $\mathbf{s}: \Omega \to S$, as defined above, of a game $G = (S, u)$ with*

possibility correspondences $\mathscr{F}_i$ *and priors* p_i *on* Ω, *where* $u: S \to \Re$. (*Note that* u *is not a function of* Ω.) *For all* i *and all* $\omega, \mathbf{s}(\omega) \in \mathscr{S}^\infty(G)$.

Why is it that weakening the assumption that $\mathscr{F}_i$ is a partition has no effect on the set of strategies and payoffs that can arise in a posteriori equilibria? The definition of Nash equilibrium requires that for every i and each $F_i \in \mathscr{F}_i$, $\mathbf{s}(F_i)$ is a best reply to beliefs over the set $\{s_{-i} \in S_{-i}: \exists F_{-i} \in \mathscr{F}_{-i} \text{ s.t. } \mathbf{s}(F_{-i}) = s_{-i}\}$. Therefore, the product of the sets $\{s_i \in S_i: \exists F_i \in \mathscr{F}_i \text{ s.t. } \mathbf{s}(F_i) = s_i\}$ are best-response closed sets (cf. Pearce (1984)) and hence in $\mathscr{S}^\infty$. Thus, weakening the assumptions does not enlarge the set of a posteriori equilibrium. We saw in proposition 2 that $\mathscr{S}^\infty$ corresponds to a posteriori equilibrium with partitions. Therefore, allowing for non-partitions has no effect on this solution concept. On the other hand, this equivalence between a posteriori equilibria with and without partitions breaks down if a common prior is imposed. That is, the set of correlated equilibria without partitions but with a common prior, allows for strategies and payoffs that cannot arise in any correlated equilibrium with a common prior and with partitions. This is demonstrated by the betting example below. Thus one can interpret weakening the partition structure as weakening the CPA. The relation between propositions 1 and 2, suggests that there ought to be a reformulation of proposition 15 that directly states that common knowledge of rationality and the game, even when knowledge violates [T, 4, and 5], will also continue to characterize $\mathscr{S}^\infty$.

Proposition 16 *If at* ω *both* [u] *and* [*rationality*] *are common knowledge, where for all* i, K_i *satisfy* [MC] *and* [N], *then* $\mathbf{s}(\omega) \in \mathscr{S}^\infty(u)$.

7.4 No-trade theorems and general possibility correspondences

In the appendix we review the no-trade results used in this subsection for the case where $\mathscr{F}_i$ are partitions. The propositions below show that results, which in the context of partitions seem quite similar, are no longer similar when non-partitions are considered. The results differ in the extent to which partition properties, equivalently the axioms on knowledge, can be weakened. For example, Aumann's (1976) agreeing-to-disagree result is more robust than the no-bets-in-equilibrium and no-common-knowledge-of-bets results. (Moreover, the last two also differ in their requirements on knowledge.)

Proposition 17 (Geanakoplos (1989)) (*see also Rubinstein and Wolinsky* (1989) *and Samet* (1990)). *Consider two agents,* $i = 1, 2$, *who share a common*

prior p, on a state space Ω, *and each has a possibility correspondence* $\mathcal{F}_i$ *that satisfies* [P1] *and* [P3] (*a fortiori* [P1] *and* [P4]).

If in some state of the world ω, *the agents' posterior beliefs concerning an event* $A \subset \Omega$, *namely the values* $\bar{p}(A \mid \mathcal{F}_i(\omega))$, *for* $i = 1, 2$, *are common knowledge, then these posteriors are equal.*

Conversely, if for some i, $\mathcal{F}_i$ *violates* [P3], *then there exists a space* Ω, *partitions* $\mathcal{F}_j$ *for all* $j \neq i$, *a common prior p and an event A such that the value of the posteriors are common knowledge but they differ.*

For simplicity we present a sketch of the proof for the first part of the proposition only, and assuming [P1, P4]. There are two main aspects to the proof. First, posterior probabilities have a "difference" property: if $F'' \subset F$ and $p(A \mid F) = p(A \mid F'') = \bar{p}$, then $p(A \mid [F - F'']) = \bar{p}$ also. Second, using (vii) of proposition 13, [P1] and [P4] imply that $\mathcal{F}_i$ also has a "difference" property described below. The rest follows from the definitions and the two difference properties just mentioned. Since it is common knowledge that *i*'s posterior probability of A is $\bar{p}_i$, there is a self-evident event G such that, for all ω in G, $p(A \mid \mathcal{F}_i(\omega)) = \bar{p}_i$. The difference property of $\mathcal{F}_i$ alluded to above implies that G is the union of disjoint sets F_i such that each F_i is either in $\mathcal{F}_i$ or is the difference of two sets in $\mathcal{F}_i$. Using the difference property of posterior probabilities, for each such F_i, $p(A \mid F_i) = \bar{p}_i$. But this implies $p(A \mid G) = \bar{p}_i$.

Proposition 18 (Geanakoplos (1989)) *Consider two agents,* $i = 1, 2$, *who share a common prior p, on a state space* Ω, *and each has a possibility correspondence* $\mathcal{F}_i$ *that satisfies* [P1, P2, *and* P4]. *Assume in addition that these two agents are considering saying yes or no to a bet* $X: \Omega \to \mathfrak{R}$ *and the agents' utility is their expected gain.*

In any Nash equilibrium the expected payoffs of the agents are zero. Conversely, if a players' possibility correspondence violates one of the assumptions, there is a betting environment with all other players having partitions, and with a Nash equilibrium where the expected utilities would not be zero.

There is also an analog to the no-common-knowledge-of-bets result.

Proposition 19 (Geanakoplos (1989)) *Consider two agents,* $i = 1, 2$, *who share a common prior p, on a state space* Ω, *and each has a possibility correspondence* $\mathcal{F}_i$ *that satisfies* [P1] *and* [$P3^+$]. *Assume in addition that these two agents are considering saying yes or no to a bet* $X: \Omega \to \mathfrak{R}$, *and an agent says yes if and only if her expected utility is non-negative.*

If it is common knowledge at ω *that the agents are saying yes to the bet, then their expected payoffs equal zero.*

Conversely, if a player's possibility correspondence violates one of the assumptions, there is a bet against players with partitions and a state of the world at which it would be common knowledge that all the players accept the bet, but their expected utilities would not be zero.

To understand why the agreeing-to-disagree result requires fewer assumptions than the two no-betting results, recall the proof of the agreeing-to-disagree result (see the appendix for a review). The "difference" property does not hold for inequalities: $F'' \subset F$ and $p(A \mid F) \geq 0$ and $p(A \mid F'') \geq 0$ do not imply $p(A \mid [F - F'']) \geq 0$. For example, let $\Omega = \{a, b, c\}$, $p(\omega) = 1/3$, $\mathcal{F}_1(\omega) = \Omega$ for all ω, $\mathcal{F}_2(a) = \{a, b\}$, $\mathcal{F}_2(b) = \{b\}$, $\mathcal{F}_2(c) = \{b, c\}$, and 1's payments to 2 are $X(a) = X(c) = -\$5.00$, $X(b) = \$7.00$. Clearly in every information cell both players will say yes if they expect their opponent to always say yes to the bet, and this is a Nash equilibrium, but their conditional expected utilities are strictly positive. Note that 2's *ex ante* utility is negative despite her interim expected utility always being strictly positive.[77]

7.5 Motivation

Weakening the assumptions on knowledge and on information seems appealing. People are not perfect reasoners and an axiomatic approach to weakening knowledge should clarify which aspects of their reasoning ability we are weakening. In fact, in contrast to the almost common certainty approach of the previous section, the syntactic approach seems to address more fundamental problems with our assumptions on knowledge. Similarly, partitions do not seem appropriate with agents who are boundedly rational, if our notion of bounded rationality allows for errors in processing information.

In this subsection we review the motivations for the weakenings of K_i and $\mathcal{F}_i$ discussed earlier, and argue that, while these weakenings might seem a priori interesting, the framework to which they are applied and the results obtained are problematic. Our concerns are based on three related issues.

> The ideas of bounded rationality and reasoning ability used to motivate non-partitions do suggest, in some *examples*, that we drop the assumptions needed to get the partition structure. However, there are very few results explaining why we should be willing to assume, say, [P1]–[P4] but not [P5] or be interested in possibility correspondences that satisfy [P1–P3], etc. The examples no more imply that we should drop [P5] in a general

analysis of games and of trade, than does the fact that juries are supposed to ignore a defendant's decision not to testify on his own behalf. Perhaps the decision making by such jurors in court may be fruitfully modeled using a particular non-partition (which violates [P5]). However, these people ought not be modeled using non-partitions in other aspects of their lives. Moreover, even if people carry this mistake outside the courtroom, it is not clear why it generalizes to other contexts as a failure of [P5]: while the non-partition that captures the ignoring of someone's decision not to provide information does violate [P5], there are many non-partitions where [P5] is violated that do not have the same easy interpretation.

The most compelling motivations for particular weakenings often argue strongly for making other modifications in the framework; but the analysis typically relies heavily on keeping the rest of the framework (e.g., the CPA and Nash equilibrium) that underlies most applications.

1 For example, the use of priors and of Bayesian updating seems to be questioned by the very same arguments used to motivate dropping [P5]. As we said at the beginning of this section, *ex ante* analysis without partitions is questionable. On the one hand, players are assumed not to know what they know; on the other hand, they are assumed to have a prior and use it in Bayesian updating based on information cells. There are essentially no *results* that justify stapling traditional frameworks together with non-partitions.
2 The motivations for non-partitions make the use of equilibrium especially questionable: if people are that boundedly rational, what justifications are there for standard equilibrium notions? Moreover, if one is inclined to use a solution concept that has stronger, albeit in this context not very strong, foundations, namely $\mathscr{S}^{\infty}$, then propositions 15 and 16 show that allowing for non-partitions has no effect. Clearly, as Geanakoplos (1989) observes, justifications of Nash equilibrium that are based on introspection and assumptions about knowledge of rationality and strategies seem in conflict with an environment where players do not know their own information structure. Drew Fudenberg (personal communication) has argued that it will be hard to obtain a learning environment in which players will learn enough to play a Nash equilibrium but not enough to

learn their own information structure, so that learning-based justifications of equilibrium with non-partitions are also questionable.

3 In an analysis using equilibrium and common knowledge, when non-partitions are involved, each player is assumed to correctly know the opponents' possibility correspondences, but not her own. Moreover, each player i knows what each opponent j thinks i's possibility correspondence is, but i thinks j is wrong despite j being correct. This seems like an odd assumption. There seems to be no reason to restrict their mistakes to their own information, and moreover to take the form of weakening [5].[78]

To clarify our concerns we will review the motivations for this research that appear in the literature. These arguments fall into two categories: explanations for why $\mathcal{F}_i$ need not be a partition when players are boundedly rational; and motivations for why K_i need not satisfy all the axioms [T, MC, N, 4, 5] when players have limited reasoning ability.

Provability and complexity *Modeling knowledge as something that agents deduce satisfies* [4] *and violates* [5].

Geanakoplos (1989) informally proposed, and independently Shin (1993) formally considered, the case where knowledge comes from proofs; we call such knowledge *provable knowledge* in this paragraph. Clearly, not having a proof for something, say fact p, implies that p is not provably known. On the other hand, not having a proof does not imply that one has a proof that a proof does not exist, so one would not provably know that p is not provably known. We find this motivation for results along the lines of subsection 7.4, questionable. Why is provable knowledge a useful notion for decision theory? The mathematician who has not proven Fermat's last theorem, and who certainly does not have a proof that such a proof is impossible, would not say that he does not know that he does not have a proof of the theorem. So, while Shin (1993) demonstrates that provable knowledge satisfies [P1] and [P4] and may violate [P5], we do not view this as a compelling motivation for considering knowledge and information that violates [P5].

Similarly, Samet (1990) motivates keeping [T] and [4], while dropping [5] by proposing axioms on the complexity of sentences that a person can use or "produce." He considers a sentence, say p, that has complexity greater than the bound on the person's ability, and argues that the person will not know that she does not know this sentence. (The axioms imply that,

not knowing p, and knowing that one does not know p, should have greater complexity than p.) Nevertheless, it is possible, even with these complexity bounds, for a person to know whatever they do know; so [4] can be satisfied. Here again, it is not clear that the violation of [5] is relevant.

Forgetfulness *Since people often forget information, the representation of what they ultimately remember may fail to be a partition.*

Geanakoplos (1989) formalizes this idea and proves that nestedness corresponds to a condition on memory. Assume that each state is a list of n statements concerning the truth or falsity of an ordered list of n facts. That is, $\Omega = \{T, F\}^n$, where the kth element of ω, denoted $\omega(k)$ says whether the kth fact is true or false.[79] Assume that information satisfies [P1] and that information about facts is remembered in the ordered sequence. That is, given any possibility cell $\mathcal{F}_i(\omega)$ there is an l s.t. $\mathcal{F}_i(\omega) = \{\omega' : \omega(j) = \omega'(j), j = 1, \ldots l\}$. This could happen if people remember the truth and falsehood only of recent facts, or of important facts. If, in addition, $l < n$ is possible, then $\mathcal{F}_i$ need not be a partition but it will satisfy [P3]. Moreover, given any nested $\mathcal{F}_i$ there is a memory-based justification of this sort.

Forgetfulness is an important limitation to incorporate into decision making. However this form of forgetfulness implies that whether or not you forget depends on the sequence of truths and falsehoods and not on the facts which are true or false. So, to get a non-partition it is necessary to forget, say, whether statements $2, \ldots, n$ are true when statement 1 is false but to remember differently when statement 1 is true. It would be interesting to see what other conditions can be motivated by more realistic models of forgetfulness. In this light it is worth noting that if people forget things in a less orderly manner, then any $\mathcal{F}_i$ satisfying [T] can be obtained. That is, if what causes people to forget can depend on particulars of the precise state of the world, then it would be equally plausible to assume that in each state ω you remember only the information received corresponding to $\mathcal{F}_i(\omega)$. This would then justify looking at any non-partition (satisfying [P1]).

Moreover, to the extent that forgetfulness provides a motivation for nestedness, it is worth noting that none of the results in the literature rely on nestedness alone. So, the sufficiency results are of limited interest unless the other conditions can be motivated as well. The necessity results fare better – they imply that bets can arise so long as there is one person with a memory problem of this ordered type, and that information may be "bad" for anyone with this memory problem.

Dynamic models *The information provided by a dynamic process may be best summarized by a non-partition.*

Shin (1989) shows that [P1, P3, P4] are necessary and sufficient for information to come from a particular dynamic model. This motivation for an interesting class of non-partitions is then as compelling as is this dynamic model. The reader is referred to the paper for details, but we do not think the decision problem portrayed by this dynamic model is a common situation. However, in contrast to Geanakoplos' forgetfulness result, Shin (1989) characterizes exactly those information structures for which information is valuable. So, to the extent that the dynamic representation is compelling, it succeeds in motivating both the sufficiency and necessity results concerning bets and valuable information.

Preferences and axioms *Different axioms on preference relations may yield behavior that is equivalent to that arising from non-partitions.*

Morris (1996) proposes a Savage-like model of preferences, where in each state the player's preferences over acts may differ. (This is not state-dependent utility; the whole preference relation on acts may depend on the state.) Knowledge is defined subjectively – similar to certainty, where a person knows if she (subjectively) assigns probability 1 – and may depend on preferences at each state. Naturally one can assume that preferences at each state correspond to expected utility preferences conditional on information received at that state, where information is given by some partition; this will correspond to a partition model. Instead Morris considers axioms on preferences that do not assume any particular information structure and shows how different axioms correspond to different information structures. For example, some of the axioms have natural interpretations concerning the value of information. The problem is that the basic axiom tying together the preferences at different states is hard to interpret, and in fact highlights the problem with non-partition models: there is no sensible way for the individual to have an *ex ante* view of the world that corresponds to a non-partition model.

This just reinforces a general problem raised earlier: if there is no *ex ante* view corresponding to a non-partition model, then there is no justification for modeling the decisionmaker as having a prior which is updated in a Bayesian like manner on non-partition information cells. But in that case, why look at *ex ante* equilibrium notions and the value of information?

Imperfect information processing *Mistakes in processing information may be modeled by non-partitions.*

These motivations do not have formal results relating them to non-partitions. Nevertheless, they do seem to justify the use of certain non-partitions. For example, Geanakoplos (1989) discusses a situation where *selection bias* – a situation where individuals select which information to use, e.g., ignoring bad news but taking good news into account – will lead to non-partitional $\mathcal{F}_i$s. However, no connection is developed between selection bias and the properties used in results on non-partitions. So, while this might motivate dropping the partition structure of information, it does not motivate results based on the particular weakenings discussed in the previous sections.

The argument in the previous paragraph seems to apply also to *implicit information*, a related cause for non-partitional information. It is argued that when reading the newspaper few people make deductions based on what is not there. But it remains to be seen that this motivates dropping [P5] in general.

Unawareness: an example and a reappearance of syntax *Axiom* [5] *implicitly requires awareness of all states and hence should be dropped.*

The most commonly cited argument for dropping [5] is unawareness. The argument that only partitions make sense starts by assuming that a person believes $\mathcal{F}(\omega)$ is the set of possible states, and that for $\omega' \in \mathcal{F}(\omega), \mathcal{F}(\omega') \neq \mathcal{F}(\omega)$. Then the person would conclude that when the state is ω' she would think the possible states are $\mathcal{F}_i(\omega')$, and differ from what she actually thinks at ω, so it cannot be that ω' is possible. But, the counterargument says that if the individual is unaware of ω', she could not "know" the set $\mathcal{F}(\omega')$.[80] Non-partition possibility correspondences may be plausible for this reason.

But this raises another question: What does it mean to say that, since $\omega' \in \mathcal{F}(\omega)$, when ω occurs the person thinks ω' is possible, even though ω' is not conceivable. The following example of unawareness, from Geanakoplos (1989), suggests a potential answer. A scientist, who is unaware of γ rays, and a fortiori is unaware of the fact that γ rays indicate that the ozone layer is disintegrating, is interested in knowing whether or not the ozone layer is disintegrating. When it is not disintegrating, the scientist can not use the non-existence of γ rays to deduce that it is not disintegrating (since she is unaware of γ rays she can not use their absense to update); when it is disintegrating she discovers γ rays and then deduces the connection with ozone, concluding that the ozone layer is disintegrating. If we describe the

state space as $\{d, \neg d\}$, where d denotes disintegrating, the scientist information is non-partitional: $\mathscr{F}(d) = \{d\}, \mathscr{F}(\neg d) = \{d, \neg d\}$. The answer this suggests to our question concerning ω' above seems to be that the scientist does not have a complete description of the state. Thus what the scientist thinks is possible in state $\neg d$ has nothing to do with the γ rays of which she is unaware; rather she only thinks about the payoff – relevant events and views *disintegration* and *not disintegration* as possible.[81]

But this introduces another concern: we have not provided a correct description of the model, neither from the scientist's limited perspective nor from ours. In the state corresponding to d the scientist is aware of γ rays. Could the model then have $\{d, \gamma\}, \neg d$ as the two states? Not if we want a model of the world as the scientist perceives it, since then in state $\neg d$ the scientist's perception is no longer that either of those two states are possible. Could this be a model of the world from the analyst's perspective? The non-partitional model is an incomplete shorthand for a more complete model with *partitions* and a prior for the person that differs from the "actual" prior. The true state space is $\{\gamma, \neg d\} \times \{d, \neg d\}$, the information *is* partitional: the scientist observes either $\{\gamma\} \times \{d, \neg d\}$ or $\{\neg\gamma\} \times \{d, \neg d\}$, and the "correct" prior on this state space has probability zero on $\neg\gamma, d$ while the scientist behaves as if this state had positive probability. (Once again, this may clarify why allowing for non-partitions is like allowing for different priors in a game.)

So what of our unawareness motivation for non-partitions satisfying [P4] but violating [P5]? It seems to be having a hard time. At best, it is a shorthand for a partitional model and it does not really justify weakening [P5]. It suggests that certain "wrong" priors can be modeled as one very special form of non-partitions; it fails to motivate general results along the lines in the previous subsection. Like our discussion of jurors' decision making earlier, the model of the scientist can at most suggest using particular non-partitions in particular examples, not exploring the general implications of non-partitions that violate [P5].

We conclude that the intuitive example of unawareness fails to motivate results where [P5] is dropped and other axioms retained. The main remaining argument for dropping [P5] in general is based on the syntactic interpretation of [5] (which we saw is equivalent to [P5]). Assumption [5] states that if an individual does not know something, then she knows she does not know it. If she does not know it because she is not aware of it, then presumably we should not insist that she knows that she does not know it. But if one is adopting the syntactic approach, one should do so with care. Modica and Rustichini (1993, 1994) argue that to model unawareness, the notion should be introduced into the language and examined. They then accept the definition that a person is unaware of ϕ if she does not know ϕ

and she does not know that she does not know ϕ, formally they introduce the symbol $\mathscr{U}$ and define it by $\mathscr{U}(\phi) \leftrightarrow \neg k(\phi) \wedge \neg k(\neg k(\phi))$. But, they argue that in a state in which you are aware of γ rays then you should also be aware of *no* γ rays; that is, they impose the symmetry axiom, [S], $\neg \mathscr{U}(\phi) \leftrightarrow \neg \mathscr{U}(\neg \phi)$. They develop various interesting results on the relationship of this symmetry axiom with [T, N, MC, 4, 5]. In particular they show that [T, N, MC, 4, S] characterizes partitions: symmetry of awareness is equivalent to [5] so leads to no real unawareness in this context.[82] Dekel, Lipman, and Rustichini (1996) make a related point. They say that to be unaware of ϕ at ω the person should not know that (she does not know)^n ϕ for all n (since if there is a sentence where there is positive knowledge of ϕ she must be aware of ϕ). They prove that requiring this for all integers n allows for unawareness, but extending it (as one should) transfinitely, results in the impossibility of real unawareness. In conclusion, these papers show that non-partitions are not an appropriate way to model unawareness because [N] and [MC] must be dropped, so a knowledge operator that allows for unawareness cannot come from a partition.[83]

APPENDIX

No-trade results

We review here three simple and basic "no-trade theorems."[84] The earliest explicit result of this kind is in Aumann (1976).[85]

Result 1 (Aumann (1976)) *Consider two agents, $i = 1, 2$, who share a common prior p, on a state space Ω, and each has information given by a partition $\mathscr{F}_i$.*

If in some state of the world ω, the agents' posterior beliefs concerning an event $A \subset \Omega$, namely the values $\bar{p}_i$ of $P_i(A \mid \mathscr{F}_i(\omega))$, for $i = 1, 2$, are common knowledge, then these posteriors are equal.

The basic property used in the proof is a version of the sure-thing principle: If $p_i(A \mid F) = p_i(A \mid F') = \bar{p}_i$ for $F, F' \in \mathscr{F}_i, F \cap F' = \phi$, then $p_i(A \mid F \cup F') = \bar{p}_i$. This property and the formal definition (see below) that at ω the events $\{\omega' : p_i(A \mid \mathscr{F}_i(\omega')) = \bar{p}_i\}$ for $i = 1, 2$, are common knowledge, immediately imply the result. More directly interesting for economics is the version of this result that speculation and trade cannot be a consequence of private information alone. If private information cannot explain speculation, other reasons must be the driving force in trade and speculation.

Result 2 (Sebenius and Geanakoplos (1983)) *Consider the framework in the result above. Assume in addition that these two agents are considering saying*

yes or no to a bet, which is a specification of how much agent 1 *will pay agent* 2 *if a certain state occurs. That is, the bet is a random variable* $X: \Omega \to \Re$. (*Negative amounts indicate payments from* 2 *to* 1.) *Assume also that, at some state of the world* ω, *both agents are willing to accept the bet if and only if in such a state of the world each agent has a non-negative expected payoff*: $E(X \mid \mathcal{F}_i(\omega)) \geq 0$.

If it is common knowledge that the agents are willing to bet, then their expected payoffs are equal to zero.

Given the first result, at a formal level this second result is not surprising: the first result says that if it is common knowledge that a particular conditional expectation for both players – namely their conditional probabilities of an event A – equal $\bar{p}$, then these conditional expectations are equal. The second says that for any conditional expectation for both agents, if it is common knowledge that they lie in the intervals $[0, \bar{p}_1]$ and $[-\bar{p}_2, 0]$ then both conditional expectations equal 0. However, the second result is, in a sense that is made precise in section 7, less robust than the first.

The final preliminary result is an equilibrium version of no-speculation. This result shows that the conclusion obtained above, by assuming a common prior and common knowledge that players are optimizing, is also obtained when we use a common prior and solve the game using correlated equilibrium. This foreshadows Aumann's (1987) characterization result, see proposition 3, that correlated equilibrium is equivalent to the assumptions of common knowledge of rationality and a common prior.

Result 3 *Consider the betting framework of the result above. In any correlated equilibrium* (*a fortiori any Nash equilibrium*) *of the game where players say YES or NO to the bet as a function of their private information and the bet is in effect if both say yes, the expected payoffs of the agents are zero.*

It is worth noting that a solution concept based solely on common knowledge of rationality – namely rationalizability – does not yield the same no-trade result: even if there is a common prior on Ω, rationalizability implicitly introduces non-common priors since it is equivalent to subjective correlated equilibrium (see proposition 2).

Notes

We are grateful to the NSF and Alfred P. Sloan for financial support, to Aviad Heifetz, Bart Lipman, Giacomo Bonanno, Steve Morris, and Aldo Rustichini for very helpful and insightful comments and conversations, and to our colleagues for their patience while we agonized over this chapter.

1 Some additional sources that include an overview or a more complete presentation of material related to this survey are Aumann (1995b), Bacharach and Mongin (1994), Binmore and Brandenburger (1990), Bonanno (1993), Brandenburger (1992), Brandenburger and Dekel (1993), Chellas (1980), Fagin *et al.* (1995), Fudenberg and Tirole (1991), Geanakoplos (1992), Lipman (1995b), Lismont and Mongin (1994a), Morris (1995a), Myerson (1991), Osborne and Rubinstein (1994), Reny (1992), Rubinstein and Wolinsky (1989), and Tan and Werlang (1984).
2 Stephen Morris has suggested that an equivalent way to view our argument is that the paradoxes arise from confusing the (exogenous) states and the (endogenous) choice of actions. This interpretation fits in better with the topic of the next section, which is concerned with constructing the state space.
3 Alternatively, in accordance with the view of footnote 1, the player is confusing the opponent's action which from her perspective is exogenous, with her own choice.
4 A game theorist reading this analysis might be inclined to dismiss it entirely as absurd and the resolution offered below as obvious. A survey of the philosophy literature on the analysis of the Prisoners' Dilemma will offer ample evidence that to many the analysis is not absurd and the resolution is not trivial. Consider the following passage from Campbell and Sowden (1985) regarding the first two paradoxes discussed in this section: "Quite simply, these paradoxes cast in doubt our understanding of rationality and, in the case of the Prisoners' Dilemma, suggest that it is impossible for rational creatures to cooperate. Thus, they bear directly on fundamental issues in ethics and political philosophy and threaten the foundations of social science." For our purposes, the most important thing to note is that the "resolution" is the same as those offered for the less obvious paradoxes below.
5 We suspect that most economists would dismiss Newcombe's Paradox as uninteresting since the source of the difficulty appears to be arising from the infallibility of the genie which in itself would appear to be problematic and perhaps not a relevant modeling hypothesis. However, as is often noted, the paradox persists even if the genie is correct with high probability, since expected-utility calculations would still yield that it is better to take box B alone rather than both. Thus, it is argued that the problem is not one postulating an all-knowing genie and hence is genuine. The "resolution" that we will offer is equally applicable to this probabilistic version.
6 Bonanno's resolution to the paradox is instructive. As discussed in section 5, he replaces the rationality axiom above with a rule of inference that allows inferences about agent i's behavior only if the hypothesis does not involve any reference to i's behavior or to any proposition that utilizes a hypothesis involving i's behavior in its proof. This "solves" the paradox since, effectively, the new rationality hypothesis overcomes the problematic feature of maintaining the truth of a player's rationality while he contemplates deviating. However, the solution goes too far by not allowing i to use deductions made about the behavior of j which relied on j's belief of i's rationality. As a consequence, Bonanno's modified model

of rationality will fail to eliminate even those strategies that require two rounds of deletion of strategies that are strictly dominated by pure strategies, in a two-person simultaneous-move game.

7 The paradoxes discussed in this section are of a different nature than the discussions of Anderlini (1990), Binmore (1987–8), and Canning (1992). These authors model players as machines and obtain impossibility and inconsistency results, closely related to, and following from, Gödel's famous work, concluding that rationality is a problematic assumption.

8 The review and some of the development below is rather terse, one source for a more thorough presentation of the material in 3.1 and 3.2.1 is Fagin *et al.* (1995, chapters 2 and 3); Chellas (1980) is a very comprehensive reference.

9 Thus Aumann's (1976) discussion of the assumption that the information structure is commonly "known" is closely related to Harsanyi's (1967) development of a commonly "known" game for situations of incomplete information.

10 Of course justifications of solution concepts that do not impose common-knowledge assumptions – such as those using an evolutionary approach, or Aumann and Brandenburger's (1995) characterization of Nash equilibrium in two-person games, see proposition 4 below – do not require a commonly "known" model.

11 In fact, we will find an information structure that, if it is assumed to be common "knowledge," generates in every state ω' the perceptions that are described by state ω'.

12 Bacharach (1985, 1988) and Samet (1990) present different, but related, approaches. As in most of the literature, the models that we will construct from this framework will involve partitions only, and no probabilities.

13 The symbol for conjunction should not be confused with the same symbol used earlier for the meet of partitions. Since in both cases $\wedge$ is the standard symbol, we abuse notation in this way. Similarly, $\neg$ will represent both set complements and syntactic negation as will be clear from the context.

14 Excepting, of course, our original description of the actual situation of incomplete information which is not a construct, but a given primitive of our analysis.

Aumann (1995b) and Hart *et al.* (1995) show that the constructed state space has the cardinality of the continuum.

15 The fact that the $\mathcal{F}_i$s constructed in this way are partitions is a result; see, e.g., Aumann (1995b), Chellas (1980), and Fagin *et al.* (1995), for proofs of such results.

16 This is a more formal version of the Kripke models described in the preceding subsection. Kripke models are usually defined using binary relations, e.g., ω *is considered possible at* ω', rather than possibility correspondences $\mathcal{F}_i$. The two methods are easily seen to be equivalent so we use the one more familiar to economists.

17 For proofs of this type of result, and more general connections between results in the language, called syntax in the literature, and those in the model, called semantics in the literature, see references cited at the beginning of this section.

18 There is an issue here which requires some caution. We have just constructed a set Ω where each ω is a list of true sentences in L. Our main claim is that this construct is equivalent to the following: view Ω as a set of states, with each state specifying only what sentences in X are true; append to Ω a partition (which we derived) for each player and using this commonly "known" information structure derive which epistemic sentences are true (just as we constructed K_i from the partitions). However, we could imagine a slightly different question. Start with a set Ω and functions $\mathcal{F}_i: \Omega \to 2^{\Omega}$, and a function from Ω into 2^X representing the subset of X that is true in each ω. Use this to determine which epistemic sentences are true. (Where $k_i(\phi)$ is true at ω if $[\phi] \in \mathcal{F}_i(\omega)$.) Check if the epistemic sentences satisfy [T, MC, N, 4, and 5]. If so, is it necessary that $\mathcal{F}_i$ is a partition? *No.* Consider the following example: $X = \{p\}$, $\Omega = \{\omega, \omega'\}$, $\mathcal{F}_i(\omega) = \mathcal{F}_i(\omega') = \{\omega\}$, and the unique basic sentence in X is p, and p is true at both ω and ω'. It is easy to see that all the axioms are satisfied concerning knowledge about p, but $\mathcal{F}_i$ is not a partition. To conclude that any Kripke structure that satisfies [T, MC, N, 4, and 5] must be partition we either need to verify the axioms on $K_i: 2^{\Omega} \to 2^{\Omega}$ or, if we want to verify it on sentences $k_i(\phi)$, then we must allow for all possible assignments of truth valuations from Ω into X. For more on this see, e.g., Fagin *et al.* (1995).

19 Maruta (1994) develops the notion of events that are *expressible* in the syntax, and, among other results, characterizes partitions using a weaker axiomatic structure but assuming that many events are expressible.

20 Actually, common knowledge can be defined without infinitely many conjunctions, by adding symbols to the language and providing a fixed-point definition of common knowledge. Even then, one might be concerned with the notion of common knowledge in a syntactic framework, since it seems to require working with sentences that involve infinitely many conjunctions. Thus, it might seem questionable to model decisionmakers who, in some vague sense, need to conceptualize, let alone verify, infinitely many sentences to know whether a sentence in their language is true.

Aumann's (1976) state-space characterization of common knowledge suggests that this should not be a concern. While common knowledge can be defined as everyone knowing that everyone knows, it is equivalent to define common knowledge in terms of the meet of the information partitions, which are simple to construct and have a simple interpretation as self-evident events. This is often called the fixed-point characterization of common knowledge. Since it seems obvious that certain events, e.g., publicly announced events, are common knowledge, it seems natural that we can verify common knowledge without getting into "infinities" (see, e.g., Milgrom (1981)).

A related issue has been addressed in syntactic models by asking whether common knowledge can be characterized with finitely many finite axioms. That is, add a symbol c_M for common knowledge among $M \subset N$ to the language of subsection 3.2.1. The question is whether there are assumptions that can be added to [T, MC, N, 4, and 5] which define c_M and have the property that, when we construct the state space and partitions (analogously to the construction

above), we get an equivalence not only between k_i and K_i, as in subsection 3.2.1, but also between c_M and CK_M? Intuitively, the answer is yes if we allow for infinitely many conjunctions and define c_M that way. But it can also be done with finite axioms using the fixed-point approach, allaying any concerns of this type. (See, Bonanno (1995), Bonanno and Nehring (1995), Fagin *et al.* (1995), Halpern and Moses (1992), and Lismont and Mongin (1993, 1994a, b). See also Barwise (1988) for a different perspective.)

21 To be more precise, there is no other model without redundant states, or put differently, any other model that agrees with that of figure 5.4b on finite levels also agrees with it on all levels.

22 This is related to Fagin, Halpern, and Vardi (1991), Fagin *et al.* (1995) and Rubinstein (1989). Carlsson and van Damme (1993) describe a realistic environment which generates an information structure with similar properties, see footnote 6.

23 The relationship between the approach of allowing more conjunctions and that of adding symbols is discussed briefly in subsection 3.2.1.

24 Heifetz (1995c) constructed a more general two-person example which has two advantages: first all the models have the property that in every state the players' knowledge coincide except concerning sentences outside the original syntax – this is not true at state ω^* in the example above, and, second, he shows that there are as many such models as partitions of Ω. See also the second part of subsection 3.2.2.

25 By contrast we definitely should care about the model of figure 5.5: A straightforward extension of Rubinstein (1989), by adding a third player, yields those types of models; and, as pointed out in footnote 6, Carlsson and van Damme (1993) provide an economically interesting model with the same properties as Rubinstein's model. The point here is that examples with even greater transfinite depth of knowledge may not have similar motivations.

26 See theorem 3.1 in Heifetz and Samet (1995).

27 Tan and Werlang (1984) are concerned with the reverse implication: they show how any standard model of asymmetric information can be mapped into (a subspace of) the Mertens and Zamir (1984) model. Brandenburger and Dekel (1993) re-examine the Mertens and Zamir formalization, focusing on Aumann's concern with common knowledge of the information structure, in particular showing which assumption in Mertens and Zamir plays the role of the assumption that the information structure is common "knowledge."

28 We present the case where there are only two players; the extension to more players is straightforward. In this discussion S is assumed to be complete, separable, and metric. Heifetz (1993) generalizes the result to the case where S is Hausdorf.

29 It is necessary to allow i to have joint beliefs over S and over j's beliefs over S since the true state in S may naturally influence j's beliefs, e.g., if i thinks j has private information about S.

30 Formally, $T_1^i \equiv \{(t_1^i, t_2^i, \ldots) \in T_0^i : \text{marg}_{X_{n-1}} t_{n+1}^i = t_n^i$ and the marginal of t_{n+1}^i on the ith copy of $\Delta(X_{n-1})$ assigns probability 1 to $t_n^i \in \Delta(X_{n-1})\}$.

31 Heifetz (1995a) shows that there is an additional implicit assumption of countable additivity: not only are we assuming that the beliefs τ_k are countably additive, but also the meta assumption is adopted that the beliefs generated by the sequence τ_k, over types of the opponent, are countably additive. This is a meta assumption because it concerns how we extend beliefs from the hierarchy to the state space, and it turns out that many finitely additive beliefs over T exist which yield the correct marginals, and only one of these is countably additive. So the uniqueness, as in subsection 3.2.1, is not completely w.l.o.g. On the other hand, Heifetz also shows that there is an alternative way to generate a Bayesian model, using non-well founded sets, which does not require this meta assumption.

There is a mathematical feature that might be helpful in understanding the difference between constructions in subsection 3.2.1 and 3.2.2. The cardinality of the set of beliefs over a set can be the same as the cardinality of the set itself. Therefore, it is feasible to construct a model where every possible belief over the state space is represented by some state. On the other hand, the cardinality of the set of partitions of a set is always bigger than the set itself. Therefore, one cannot construct a state space in which each state incorporates every possible partition of the set. So, we should expect that it will not be the case that each state can express every possible sentence about the players' knowledge (see also the discussion following figure 5.5, and Gilboa (1988) and Heifetz and Samet (1995)).

32 The extension of this subsection to infinite Ω is delicate as it requires attention to various topological and measure-theoretic details, such as why E must be closed. We maintain the finiteness assumption for simplicity.

33 Battigalli and Bonanno (1995), Brandenburger and Dekel (1987b), Lamarre and Shoham (1994), and Nielson (1984) also consider the relationship between knowledge and certainty.

34 Note that property [N] follows from the other assumptions, and that any prior with the same support, S, will generate the same B_i and K_i.

35 However, the model of certainty and knowledge at the end of subsection 3.3 does create a support, even though it does not create a unique prior. This suggests that perhaps a weaker assumption, such as common supports, can be justified. There is very little exploration of this issue. Assumptions like this have been used, together with additional assumptions, to characterize Nash equilibrium in generic games of perfect information (Ben Porath (1994)). Stuart (1995) shows that common support and common certainty of rationality yields Nash equilibrium in the finitely repeated Prisoners' Dilemma.

36 Morris (1995a) compellingly argues in favor of dropping the CPA even when an *ex ante* stage does exist.

37 The reader will note that in section 6 some results rely on the uniform boundedness of the utility functions and should therefore be treated with similar scepticism.

38 The reader might wonder why we care about characterizations of solution concepts. The common view of characterization results such as those in this chapter is that they explain how introspection alone can lead agents to play in

accordance with various solution concepts. We would like to emphasize two closely related roles: characterizations provide negative results and they suggest which solution concepts are more appropriate in different environments. For example, there may be results which provide very strong sufficient conditions for a particular equilibrium notion, such as Nash equilibrium. This cannot constitute a proof that the equilibrium notion is unreasonable, since it is possible that an alternative model with appealing sufficient conditions exist (such as, for example, recent evolutionary models). Nevertheless, most results are tight in that if assumptions are weakened then play could be different from the characterized solution concept. Thus, these results do suggest that various solution concepts, in particular Nash equilibrium and backwards induction, are implausible in various contexts. An example of the second type is the following: for some solution concepts it is shown that common certainty of rationality rather than common certainty of beliefs is sufficient. Such a result may be important in a learning or evolutionary environment where common certainty of rationality is more plausible than common certainty of beliefs. Similar insights can be obtained by examining the robustness of characterizations to weakenings of the common-certainty assumptions. For example, different characterizations are robust to replacing common certainty with different notions of "almost" common certainty. Therefore, since common certainty assumptions are unlikely to be satisfied in any real context, the appropriate concept should depend on which form of "almost" common certainty is most likely to be satisfied in that context.

39 See also Stalnaker (1994) and Tan and Werlang (1988).

40 That $\mathcal{S}^\infty$ is equivalent to an interim solution concept is natural: it is characterized by an assumption that is made about a particular state of the world, not an *ex ante* assumption about the constructed states of the world.

41 That any such equilibrium uses strategies that survive $\mathcal{S}^\infty$ follows from arguments similar to Bernheim's and Pearce's characterization of rationalizability: each strategy used in such an equilibrium is a best reply to a belief over the set of those opponents' strategies that are used; therefore these sets of strategies are best reply sets, hence survive iterative deletion. To construct an a posteriori subjective correlated equilibrium that uses any strategy that survives iterative deletion we simply construct an information structure where the state space is the set of strategy profiles, each player is informed of a recommendation for her own strategy, and the players' conditional probabilities on the state space are, for each possible recommendation to i, say σ_i, the belief – which is over the opponents' strategies that survive iterative deletion – which makes σ_i a best reply. (Such a belief exists since the strategies are rationalizable.) For more detail see Brandenburger and Dekel (1987a).

42 For two-person games the CPA is not needed.

43 Clearly the hypothesis that each ω in the support of p is also in $[rationality] \cap [u]$ is equivalent to the assumption that at each such ω there is common certainty of $[rationality]$ and of $[u]$. Thus it might appear that common certainty of rationality is necessary for correlated equilibrium. However, in

subsection 6.2 we will see that the appropriate statement of this *ex ante* result is as in proposition 3 and not with common certainty; see the text preceding footnote 6.2.

44 It is important that j *knows* the payoffs are u_j; if j were only certain of the payoffs then j could be wrong, and then the players would not necessarily be playing a Nash equilibrium of the game $(\Sigma, \mathbf{u}(\omega))$. Moreover, they would not even necessarily be playing a Nash equilibrium of the game they believe they are certain they are playing, as each player i could believe that j has wrong beliefs about j's own payoffs. Thus, while we agree that certainty "of one's own payoff is tautological" (Aumann and Brandenburger (1995, section (7c)), knowledge of one's own payoff is not tautological, and seems necessary for proposition 4.

45 Applying this analysis to the case of correlated equilibrium may shed light on which of the various definitions of correlated equilibrium, see, e.g., Cotter (1991) and Forges (1992), are appropriate and in which contexts.

46 These and related concerns with characterizing refinements are discussed, e.g., by Börgers and Samuelson (1992) and Pearce (1982), see also Cubitt (1989) and Samuelson (1992).

47 Throughout this section we use the term almost common certainty for the general idea only – each particular notion will go by a different label. Thus, while the term common certainty is formally equivalent to the term common 1 belief (defined below), almost common certainty is not a formal precise term, and is not equivalent to almost common 1 belief, formally defined below.

There are other models of noise, which may appear more natural, that lead to similar information structures; for example, Carlsson and van Damme (1993) use a generalization of an information structure where players are interested in the value of a parameter x, but each player is informed of the true value of x plus some i.i.d. noise with support $(-\varepsilon, \varepsilon)$. Then a player may believe that the true value is close to zero, and that the other believes the true value is close to zero, but given any value y as far from zero as you want, there is a chain that one believes 2 believes . . . 1 believes y is possible. Formally, neither Carlsson and van Damme's (1993) model, nor Rubinstein's (1989) model, have a non-empty strict subset which is common q belief at any state for q close to 1. See below for a definition of common q belief.

48 A related question would be to ask what notions of convergence of probabilities yield (lower hemi) continuity results. For example, consider a coin tossed infinitely often (including infinity) and let p be the probability of heads. Consider the game where a player can choose to play and get 2 if the coin never falls on heads, 0 otherwise, or not to play and get 1 for sure. Clearly she should choose to play for $p = 1$ and not to play for $p < 1$. Consider $p_n \to 1$. The induced probability distributions on the state space of the first time the coin falls on heads, $\Omega = \{1, 2, \ldots\} \cup \{never\}$, converge weakly but the *ex ante* expected payoffs are 1 along the sequence and 2 in the limit, and the strategy choice is not to play in the sequence and to play in the limit. Clearly, the standard notion of weak convergence is not sufficient. The notions of convergence of probability that yield lower hemi continuity in this context are developed in Engl (1994).

Kajii and Morris (1994a) develop other notions of convergence more closely related to almost common certainty. Both these papers discuss the relationship among these notions of convergence and almost common certainty.

49 Fudenberg, Kreps, and Levine (1988) obtained the analogous result for refinements of Nash equilibrium, essentially that the only robust refinement is that of Nash equilibrium in strategies that are not weakly dominated (see also remark 4.4 in Dekel and Fudenberg (1990)).

50 In the case that $\mathcal{F}_i(\omega)$ has zero probability, choose any version of a conditional probability.

51 Here again the subscript N denotes the intersection of B_i^q over all i in N, and the superscript n denotes n iterations of the operator B_i^q.

52 There is a useful characterization of common q belief similar to that of common knowledge, simplifying the iterative definition to a fixed point definition. It is analogous to the result that A is common knowledge at ω if there exists a self-evident event that contains ω and is in A. Say that an event E is *evident q belief* if for all $\omega \in E$, it is true that $p_i(E \mid \mathcal{F}_i(\omega)) \geq q$, i.e., whenever E happens everyone assigns probability at least q that E happened. Monderer and Samet show that E is common q belief at ω if and only if there is an evident q belief set F, with $\omega \in F \subset B_N^q(E)$. (An extension of common q belief to uncountable Ω is in Kajii and Morris (1994b).)

53 It is worth emphasizing that this motivation is based on the *analyst's* prior and doubts about her own assumptions.

54 The rough idea for this fact can be seen as follows. If *ex ante i* thinks E is likely then it is *ex ante* likely to be the case that after receiving her private information i still thinks E is likely. (After all, i's *ex ante* belief in E is the weighted average, using the prior, of her conditional beliefs [which are bounded]. So if *ex ante* E is likely, "most" conditional beliefs are that E is likely.) Using the common prior, this implies that everyone thinks it is likely to be the case that everyone's conditional beliefs are that E is likely. But this is just the statement that E is likely to be evident q belief for q large, i.e., E is almost certainly almost common 1 belief.

55 As one might expect, since subjective certainty of an event is not at all like common knowledge, weakening our common knowledge requirement to the assumption that the game and rationality is almost subjectively certainty does affect our conclusions. In particular this would only characterize $\mathcal{S}^2(\pi, S)$. Moreover, here the assumption that it is common knowledge that everyone knows their own payoffs is important; in its absence the characterization is weakened to $\mathcal{S}^1(\pi, S)$.

Perhaps surprisingly, it turns out that almost certainty of the game and rationality is nevertheless sufficient to characterize iterated dominance. However, since almost certainty requires a common prior it does not make sense in this context where we are not assuming a common prior. We state this result when we discuss the incomplete-information-game interpretation below; see subsection 6.3.1.

It is also easy to see that for any finite game, weakening common certainty to

almost ∞ certainty, i.e., requiring iteration to arbitrarily high (but finite) levels, also does not effect the conclusion of the results on $\mathscr{S}^\infty$ because only finitely many iterations are necessary in any finite game. Lipman (1994) examines infinite games, and this robustness does not extend to discontinuous infinite games.

As we should expect, proposition 9 above does not have anything to say about the *ex ante* payoffs in the model: it could be that while [*rationality*] is almost common certainty at ω, the state ω can be unlikely (*ex ante*). Nevertheless, it is an immediate corollary to proposition 9, and proposition 2, that if [*rationality*] and [u] are almost certainly almost common 1 belief, then the *ex ante* payoffs are also close to expected payoffs that survive $\mathscr{S}^\infty$.

56 An advantage of introducing the CPA as an event, [CPA], is that one can use it to evaluate the implications of assuming that it is almost common certainty that there is a common prior. Thus, one could examine the robustness of various results, including a local version of proposition 3 that holds at a state of the world ω instead of globally, but we have not done so.

57 This is a substantive assumption – see also section 3.4.

58 The event [u] is the same in each model, but, since the probabilities are changing, the optimal strategies may be changing, hence the event [rationality] may change in the sequence.

59 We introduce some simplifying assumptions that will make it possible to present a brief sketch of a proof. Consider a particular convergent subsequence of distributions on actions, say, $\{\phi^n\}_{n=1}^\infty$, where $\phi^n \in \Delta(S)$ is defined by $\phi^n(s) \equiv \Sigma_{\{\omega: \mathbf{s}(\omega)=s\}} p^n(\omega)$. (We will denote the subsequence by n rather than the more precise n_k to simplify notation.) Our first simplifying assumption is that p^n converges; we denote its limit by p. We argue that $\mathbf{s}: \Omega \to S$ is a correlated equilibrium with the information structure $(\Omega, \mathscr{F}_i, p)$. If not then some player i has a profitable deviation. Our second simplifying assumption is that this implies that there is a state ω, with $p(\omega) > 0$, at which playing something other than $\mathbf{s}_i(\omega)$, say $\tilde{s}$, is a better reply against i conjecture at ω, $\phi_{-i}(\omega) \in \Delta(S_{-i})$, where $\phi_{-i}(\omega)(s_{-i}) \equiv \Sigma_{\{\omega': \mathbf{s}_{-i}(\omega')=s_{-i}\}} p(\omega' \mid \mathscr{F}_i(\omega))$. (This is simplifying since for infinite Ω we should consider the case where each ω is null, and look at deviations on measurable sets with positive probability.) But then, since $p(\omega) > 0, \phi_{-i}(\omega) = \lim \phi^n_{-i}(\omega)$, where $\phi^n_{-i}(\omega)(s_{-i}) \equiv \Sigma_{\{\omega': \mathbf{s}_{-i}(\omega')=s_{-i}\}} p^n(\omega' \mid \mathscr{F}_i(\omega))$. So, $\tilde{s}$ must be better than $\mathbf{s}_i(\omega)$ against ϕ^n_{-i} for n large, i.e., $\mathbf{s}_i(\omega)$ is not optimal for i in the nth model. On the other hand, since $p(\omega) > 0, \omega \in E^n$ for n large enough, which means that $\mathbf{s}_i(\omega)$ should be optimal, leading to a contradiction.

60 See also footnote 4.

61 Proposition 10 could also be stated in this way; we chose not to do so because the notion of an ε-correlated equilibrium is not standard.

62 There we noted that if i assigned positive probability to j playing s_j, and probability 1 to $[u] \cap [rationality] \cap [\phi_j]$ then there is a state $\omega' \in [u] \cap [rationality] \cap [\phi_j] \cap [s_j]$ which implies that s_j is a best reply against ϕ_j given payoffs u_j. Now we simply note that the total probability that i assigns to strategies s_j that are not best replies is bounded by $1 - \varepsilon$, since at any

$\omega' \in [u] \cap [rationality] \cap [\phi_j], s_j(\omega')$ is a best reply to $[\phi_i]$ given payoffs u_j. Also, as in the discussion after proposition 4, one could drop the assumption that players know their own payoffs, but this seems to require strengthening the hypothesis by adding the assumption that at ω there is almost mutual certainty that $[u]$ is almost mutually certain, i.e., $p_i([u] \mid \mathcal{F}_i(\omega)) \geq 1 - \delta$ and $p_i(\{\omega'': p_j([u] \mid \mathcal{F}_j(\omega'')) \geq 1 - \delta, j \neq i\} \mid \mathcal{F}_i(\omega) \geq 1 - \delta$.

63 This result follows the same arguments as in preceding results, so will not be repeated. The result does imply that the only standard refinement with a somewhat appealing epistemic characterization is that of trembling-hand perfection, and then only for two-person games.

64 Because the strategy spaces are finite, there is a $\bar{q}$ that works uniformly throughout the iteration.

65 This is also the approach implicit in subsection 6.2. Fudenberg, Kreps, and Levine focus on the case where almost certainly there is almost common certainty that the payoffs are *almost as in G*. Formally, for a sequence $u^n \to u$ and $q^n \to 1$ they assume that with probability at least q^n, u^n is common q^n belief. By contrast we assume here $u^n = u$. For examining robustness of solution concepts theirs is a very sensible weakening of the assumption that the analyst is almost certain that the game G is almost common certainty, by adding the adjective *almost* before the game G as well. Since it turns out that this change only complicates the statement of the results, without contributing significantly to understanding the issues with which we are concerned here, we do not consider this additional weakening in this chapter. Dekel and Fudenberg (1990) precisely analyze the role of allowing for this additional "almost."

66 To be precise, it appears in the fourth paragraph of that section.

67 Not only is $\mathcal{S}^\infty$ robust in this first sense to weakening common certainty to almost common 1 belief; it is also robust to weakening it to almost certainty. Formally, assume that $s^n \in \mathcal{S}^\infty(G^n)$. This implies that if $\lim p^n([s]) > 0$ then $s \in \mathcal{S}^\infty(G)$. If the analyst believes that any uncertainties about the payoffs are reflected by a common prior of the players, but that strategic uncertainty is not captured by a common prior, the analyst will want to know if $\mathcal{S}^\infty$ is robust to weakening common certainty to almost certainty. This makes sense if the uncertainties about the payoffs arise from some physical structure about which players have asymmetric information (e.g., the success of player i's firm); but the uncertainties concerning opponents strategies are subjective.

On the other hand, if the uncertainties over payoffs do not come from a common prior, the appropriate question is whether $\mathcal{S}^\infty$ is robust to weakening common certainty to almost subjective certainty. As we noted above – see that text preceding section 6.2 – it is *not*. So, even though $\mathcal{S}^\infty$ is a very coarse solution concept, it is not robust in this very demanding sense.

68 To see all this fix a strict Nash equilibrium of G, denoted by s. Let C^n be the event in Ω on which G is common p^n belief, and let $[u]$ be the event in which the game is actually G. Finally, let $\Omega_i^n = B_i^{q^n}(C^n)$ – the event on which i believes with probability at least q^n that there is common q^n belief that the game is G; and $\Omega_*^n = C^n \cap [u]$. Note that for ω in Ω_i^n, i is almost certain that the event Ω_*^n

occurred: $\omega \in \Omega_i^n \Rightarrow p_i^n(\Omega_N \mid \mathscr{F}_i(\omega)) \to 1$. (This is because for such ω, i is almost certain that the game is G and that G is almost common 1 belief, so he is almost certain that both these events obtain.) Consider the following strategies $\mathbf{s}^n$ in G^n. (α) Play s_i on Ω_i^n. (β) Find a Nash equilibrium where i is choosing strategies only for other states assuming everyone is playing as given in (α). (That is, consider the restricted game where every i is required to play s_i on Ω_i^n and free to choose anything at all other states. Of course strategies must still only depend on a player's information.) Since s is a strict Nash equilibrium, all players are happy with (α) when n is large. (This is because on Ω_i^n player i is almost certain that everyone else is playing s and that the game is G.) Since in (β) we constructed a Nash equilibrium, everyone is happy with (β) as well. Clearly the interim payoffs of the equilibria we have constructed on G^n converge to those of G; the *ex ante* payoffs converge as well if it is almost certain that G is almost common 1 belief.

If s were not a strict Nash equilibrium, but just a Nash equilibrium, then the construction would yield an interim ε Nash equilibrium in G^n, i.e., strategies that are ε optimal at every information set. Thus, the notion of interim ε Nash equilibrium is robust in that given any $\tilde{G}$, every interim ε Nash equilibrium of G is played in some interim ε Nash equilibrium of $\tilde{G}$ in those states where G is almost common 1 belief.

69 The distribution of actions generated by $\mathbf{s}$ on $[u]$ is formally defined as the element $\phi \mid_u$ of $\Delta(S)$ given by $\phi \mid_u(s) = \Sigma_{\{\omega \in [u]: s(\omega) = s\}} p_n(\omega)/p_n([u])$.

70 See also the discussion of the characterization of Nash equilibria in games of incomplete information and games with moves by Nature at the end of section 4.

71 Kajii and Morris (1995) is not, strictly speaking, a generalization of Carlsson and van Damme (1993), because the latter allow for continuous noise, and do not assume that players know their own payoffs, but it seems that the methods of Kajii and Morris (1995) could be used to generalize Carlsson and van Damme (1993).

72 Monderer and Samet consider perturbations of $\mathscr{F}_i$s, whereas Kajii and Morris focus on changing only the probabilities as we do in this section. Formally, Kajii and Morris (1994a) show that given a Nash equilibrium $\mathbf{s}$ for G^∞ there is an ε-Nash equilibrium $\mathbf{s}^n$ for a sequence of games G^n, where the expected payoffs of $\mathbf{s}^n$ in G^n converge to those of $\mathbf{s}$ in G^∞ if an only if p^n converges to p in the sense that it is almost certain that it is almost common 1 belief that the difference in the conditional probabilities are uniformly almost zero. More precisely, the only if result is that when the sequence fails to converge there exists games where the expected payoffs fail to converge.

73 Assumption [N], that you always know Ω, is also strong – it implicitly rules out some forms of unawareness, in that the complete list of states is always known. Similarly, [MC] is strong since it implies monotonicity, $A \subset B \Rightarrow K_i(A) \subset K_i(B)$, which in turn implies that at any state at which you know anything, you also know the complete state space Ω. For now [N] and [MC], which are necessary and sufficient for the translation between possibility correspondences and knowledge operators, are maintained.

74 While some authors have briefly considered the effect of just dropping [T], without imposing [D], (Geanakoplos (1989), Samet (1990), Brandenburger, Dekel, and Geanakoplos (1992)), we feel that [D] is a basic property of belief and knowledge. Thus, since weakening [T] to [D] has already been analyzed and shown to correspond to weakening knowledge to belief, we will not focus on weakening [T] in this section. Nevertheless, we will present a result where both [D] and [T] are dropped – Brandenburger, Dekel, and Geanakoplos show that this is w.l.o.g. for some purposes, and the extra generality comes at no cost in the presentation. Naturally, one can ask many of the other questions that follows in the context of assuming [D] and not [T]. We leave such exercises to the interested reader.

75 Morris (1992, 1996) extends this in several directions. He provides a multi-stage dynamic decision-making context, and derives additional properties that the possibility correspondence must satisfy if it will meet some additional requirements concerning dynamic consistency and the value of information in these richer contexts; he allows for non-Bayesian updating of probabilities; and he considers non-expected utility preferences. The latter is an attempt to motivate non-partitional structures from an axiomatic perspective, and will be mentioned again below when we turn to the motivation for these structures.

76 As with partitions and common q belief, there exists a non-iterative characterization of common knowledge in this more general context as well: E is common knowledge at ω' if there is a self-evident F, (i.e., $\mathcal{F}_i(\omega) \subset F$ given any $\omega \in F$), that contains ω' and that is perceived to be a subset of E, i.e., given any ω in F, $\mathcal{F}_i(\omega) \subset E$. If $\mathcal{F}_i$ satisfies [P1] (equivalently, if K_i satisfies [T]), then this reduces to the following: E is common knowledge at ω if there exists a self-evident F with $\omega \in F \subset E$.

77 The hypothesis in proposition 19 may appear weaker than those in proposition 18, but this is not the case: the proposition 19 assumes that it is common knowledge that players say yes whereas in 18 the assumption is that we are considering a Nash equilibrium.

78 This clarifies why non-partitions with common priors leads to the same behavior as do non-partitions and partitions without common priors. The lack of a common prior corresponds to the disagreement about i's information structure.

79 This is similar to the syntactic construction of Ω; the difference is that the order of the elementary facts in X is now important.

80 See, e.g., Binmore and Brandenburger (1990).

81 A more common version of this story is Watson's deductions concerning the guilt of a person based on a dog not barking – he fails to use the lack of barking as an indication that the dog knew the individual, while he would (we presume) deduce from any barking that the person was not known to the dog.

82 What happens in the non-partition model of the scientist is instructive. The typical model has $\Omega = \{d, \neg d\}$: a state d which is characterized by the ozone layer disintegrating and γ rays appearing, and a state $\neg d$ which is characterized by no γ rays appearing and no disintegration. As before we specify

$\mathcal{F}(d) = \{d\}, \mathcal{F}(\neg d) = \{d, \neg d\}$. So at d, the scientist knows there are γ rays and the ozone layer is disintegrating, so is aware of everything including γ rays. At $\neg d$ she does not know d. Moreover, at $\neg d$, she does not know that she does not know d, since the set of states at which she does not know d is $\neg d$, and at $\neg d$ she does not know $\neg d$. So she is unaware of γ rays. So far so good. But consider the person's awareness of *no* γ rays. While it is true that at $\neg d$ she does not know there are no γ rays, she does know that she does not know it. This is because she knows Ω, and at both states she will not know that there are no γ rays, so she always knows that she does not know there are no γ rays. Thus, at state d she is aware of the sentence "there are no γ rays," so Modica and Rustichini argue she should be aware of the possibility of γ rays.

83 Dropping [N] and [MC] is related to dropping logical omniscience, the requirement that an individual can deduce all the logical implications of his knowledge. Dekel, Lipman, and Rustichini (1996) argue that, in fact, the state-space model is inappropriate for modeling unawareness as it imposes a form of logical omniscience by identifying events with all sentences that are true in that event, so that knowledge of any sentence implies knowledge of any other sentence that is logically equivalent. Various aspects of logical omniscience are weakened (in very different models) by Fagin and Halpern (1988), Lipman (1995a), and Modica and Rustichini (1993, 1994).

84 In order to focus on the foundations of knowledge and rationality we will not present the extensions to more interesting economic environments, such as Milgrom and Stokey (1982).

85 Actually, there is a precursor in Aumann (1974). The result stated there, that in a zero-sum game allowing correlations without differing priors will not change the value of the game, can be shown to imply result 3 below.

References

Anderlini, L. (1990). "Some notes on church's thesis and the theory of games." *Theory and Decision*, 29: 19–52.

Armbruster, W. and Böge, W. (1978). "Bayesian game theory." In Moeschlin, O. and Pallaschke, D. (eds.), *Game Theory and Related Topics*. Amsterdam: North-Holland.

Aumann, R.J. (1974). "Subjectivity and correlation in randomized strategies." *Journal of Mathematical Economics*, 1: 67–96.

(1976). "Agreeing to disagree." *The Annals of Statistics*, 4(6): 1236–9.

(1987). "Correlated equilibrium as an expression of Bayesian rationality." *Econometrica*, 55: 1–18.

(1995a). "Backward induction and common knowledge of rationality." *Games and Economic Behavior*, 8: 6–19.

(1995b). "Interactive epistemology." Working Paper No. 67, The Hebrew University of Jerusalem.

Aumann, R.J. and Brandenburger, A. (1995). "Epistemic conditions for Nash equilibrium." *Econometrica*, 63: 1161–80.

Bacharach, M. (1985). "Some extensions of a claim of Aumann in an axiomatic model of knowledge." *Journal of Economic Theory*, 37: 167–90.

(1988). "When do we have information partitions?" Mimeo, Oxford University.

Bacharach, M. and Mongin, P. (1994). "Epistemic logic and the foundations of game theory." *Theory and Decision*, 37: 1–6.

Barwise, J. (1988). "Three views of common knowledge." In Vardi, M. Y. (ed.), *Theoretical Aspects of Reasoning about Knowledge*. Los Altos, CA: Morgan Kaufman.

Basu, K. (1990). "On the non-existence of a rationality definition for extensive games." *International Journal of Game Theory*, 19: 33–44.

(1995). "A paradox of knowledge and some related observations." Mimeo, Cornell University.

Battigalli, P. and Bonanno, G. (1995). "The logic of belief persistency." Mimeo, University of California, Davis.

Ben-Porath, E. (1994). "Rationality, Nash equilibrium and backwards induction in perfect information games." Mimeo, Tel Aviv University.

Bernheim, D. B. (1984). "Rationalizable strategic behavior." *Econometrica*, 52: 1007–28.

Bhattacharyya, S. and Lipman, B. L. (1995). "Ex ante versus interim rationality and the existence of bubbles." *Economic Theory*, 6: 469–94.

Bicchieri, C. (1988). "Strategic behavior and counterfactuals." *Synthese*, 76: 135–69.

(1989). "Self-refuting theories of strategic interaction: a paradox of common knowledge." *Erkenntnis*, 30: 69–85.

Binmore, K. G. (1984). "Equilibria in extensive games." *Economic Journal*, 95: 51–9.

(1987–8). "Modeling rational players I and II." *Economics and Philosophy*, 3 and 4: 179–214 and 9–55.

Binmore, K.G. and Brandenburger, A. (1990). "Common knowledge and game theory." In Binmore, K. G. (ed.), *Essays on the Foundations of Game Theory*. Cambridge, MA: Blackwell, chapter 4.

Blume, L., Brandenburger, A., and Dekel, E. (1986). "Lexicographic probabilities and choice under uncertainty." *Econometrica*, 59: 61–79.

(1986). "Lexicographic probabilities and equilibrium refinements." *Econometrica*, 59: 81–98.

Böge, W. and Eisele, Th. (1979). "On solutions of Bayesian games." *International Journal of Game Theory*, 8: 193–215.

Bonanno, G. (1991). "The logic of rational play in games of perfect information." *Economics and Philosophy*, 7: 37–65.

(1993). "Information partitions and the logic of knowledge and common knowledge." Mimeo, University of California, Davis.

(1995). "On the logic of common belief." Mimeo, University of California, Davis.

Bonnano, G. and Nehring, K. (1995). "Intersubjective consistency of beliefs and the logic of common belief." Mimeo, University of California, Davis.

(1994). "Weak dominance and approximate-common knowledge." *Journal of Economic Theory*, 4: 265–76.

Börges, T. and Samuelson, L. (1992). "Cautious utility maximization and iterated weak dominance." *International Journal of Game Theory*, 21: 13–25.

Brandenburger, A. (1992). "Knowledge and equilibrium in games." *Journal of Economic Perspectives*, 6: 83–101.

Brandenburger, A. and Dekel, E. (1987a). "Rationalizability and correlated equilibria." *Econometrica*, 55: 1391–402.

(1987b). "Common knowledge with probability 1." *Journal of Mathematical Economics*, 16: 237–45.

(1989). "The role of common knowledge assumptions in game theory." In. Hahn, R. (ed.), *The Economics of Missing Marketes, Information and Games*. Oxford: Oxford University Press.

(1993). "Hierarchies of beliefs and common knowledge." *Journal of Economic Theory*, 59: 189–98.

Brandenburger, A., Dekel, E., and Geanakoplos, J. (1992). "Correlated equilibrium with generalized information structures." *Games and Economic Behavior*, 4: 182–201.

Campbell, R. and Sowden, L. (1985). *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcombe's Problem*. Vancouver, British Columbia: University of British Columbia Press.

Canning, D. (1992). "Rationality, computability, and Nash equilibrium." *Econometrica*, 60: 877–88.

Carlsson, H. and van Damme, E. (1993). "Global games and equilibrium selections." *Econometrica*, 61: 989–1018.

Chellas, B. F. (1980). *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

Cotter, K. (1991). "Correlated equilibrium with type-dependent strategies." *Journal of Economic Theory*, 54: 48–68.

Cubitt, R. (1989). "Refinements of Nash equilibrium: a critique." *Theory and Decision*, 26: 107–31.

Dekel, E. and Fudenberg, D. (1990). "Rational behavior with payoff uncertainty." *Journal of Economic Theory*, 52: 243–67.

Dekel, E., Lipman, B., and Rustichini, A. (1996). "Possibility correspondences preclude unawareness." Mimeo, Northwestern University.

Engl, G. (1994). "Lower hemicontinuity of the Nash equilibrium correspondence." Mimeo, University of California, Irvine.

Fagin, R., Geanakoplos, J., Halpern, J., and Vardi, M. (1993). "The expressive power of the hierarchical approach to modeling knowledge and common knowledge." Mimeo, IBM Almaden Research Center.

Fagin, R. and Halpern, J. (1988). "Belief, awareness, and limited reasoning." *Artificial Intelligence*, 34: 39–76.

Fagin, R., Halpern, J. Y., Moses, Y., and Vardi, M. Y. (1995). *Reasoning About Knowledge*. Cambridge, MA: MIT Press.

Fagin, R., Halpern, J., and Vardi, M. (1991). "A model-theoretic analysis of knowledge." *Journal of the Association for Computing Machinery*, 38: 382–428.

Forges, F. (1992). "Five legitimate definitions of correlated equilibrium in games

with incomplete information." Working Paper No. 383, Ecole Polytechnique.
Fudenberg, D., Kreps, D. M., and Levine, D. (1988). "On the robustness of equilibrium refinements." *Journal of Economic Theory*, 44: 354–80.
Fudenberg, D. and Tirole, J. (1991). *Game Theory*. Cambridge, MA: MIT Press.
Geanakoplos, J. (1989). "Game theory without partitions, and applications to speculation and consensus." Mimeo, Cowles Foundation for Research in Economics at Yale University.
(1993). "Common knowledge." *Journal of Economic Perspectives*, 6: 53–82.
Gilboa, I. (1988). "Information and meta information." In Halpern, J. Y. (ed.), pp. 227–43.
Gul, F. (1995a). "A comment on Aumann's Bayesian view." Forthcoming *Econometrica*.
(1995b). "Rationality and coherent theories of strategic behavior." Forthcoming *Journal of Economic Theory*.
Halpern, J. Y. (ed.) (1988). *Theoretical Aspects of Reasoning about Knowledge*. Los Altos, CA: Morgan Kaufman Publishers.
Halpern, J. Y. and Moses, Y. M. (1990). "Knowledge and common knowledge in a distributed environment." *Journal of the Association of Computing Machinery*, 51: 549–87.
(1992). "A guide to completeness and complexity for modal logics of knowledge and belief." *Artificial Intelligence*, 54: 319–79.
Harsanyi, J. (1967). "Games with incomplete information played by 'Bayesian' players, parts I–III." *Management Science*, 14: 159–82, 320–34, 486–502.
Hart, S., Heifetz, A. and Samet, D. (1995). "'Knowing whether', 'knowing that', and the cardinality of state spaces." Forthcoming *Journal of Economic Theory*.
Heifetz, A. (1993). "The Bayesian formulations of incomplete information – the non-compact case." *International Journal of Game Theory*, 21: 329–38.
(1995a). "Non-well-founded type spaces." Mimeo, Tel Aviv University.
(1995b). "Infinitary S5-epistemic logic." Mimeo, Tel Aviv University.
(1995c). "How canonical is the canonical model? A comment on Aumann's interactive epistemology." Discussion Paper No. 9528, CORE.
Heifetz, A. and Samet, D. (1995). "Universal partition spaces." Mimeo, Tel Aviv University.
Kajii, A. and Morris, S. (1994a). "Payoff continuity in incomplete information games." CARESS Working Paper No. 94-17, University of Pennsylvania.
(1994b). "Common p-belief: the general case." CARESS Working Paper No. 94-15.
(1995). "The robustness of equilibria to incomplete information." CARESS Working Paper No. 95-18.
Kaneko, M. (1987). "Structural common knowledge and factual common knowledge." RUEE Working Paper No. 87-27.
Kohlberg, E. and Mertens, J.-F. (1985). "On the strategic stability of equilibria." *Econometrica*, 54: 1003–38.
Lamarre, P. and Shoham, Y. (1994). "Knowledge, certainty, belief and conditionalisation." Mimeo, IRIN, Université de Nantes.

Lipman, B. L. (1994). "A note on the implications of common knowledge of rationality." *Games and Economic Behavior*, 6: 114–29.
(1995a). "Decision theory with logical omniscience: toward an axiomatic framework for bounded rationality." Mimeo, University of Western Ontario.
(1995b). "Information processing and bounded rationality: a survey." *Canadian Journal of Economics*, 28: 42–67.
Lismont, L. and Mongin, P. (1993). "Belief closure: a semantics for modal propositional logic." CORE Discussion Paper No. 9339.
(1994a). "On the logic of common belief and common knowledge." *Theory and Decision*, 37: 75–106.
(1994b). "A non-minimal but very weak axiomatization of common belief." *Artificial Intelligence*, 70: 363–74.
Maruta, T. (1994). "Information structures on maximal consistent sets." Discussion Paper No. 1090, Northwestern University.
Mertens, J.-F. and Zamir, S. (1985). "Formulation of Bayesian analysis for games with incomplete information." *International Journal of Game Theory*, 14: 1–29.
Milgrom, P. (1981). "An axiomatic characterization of common knowledge." *Econometrica*, 49: 219–22.
Milgrom, P. and Stokey, N. (1982). "Information trade and common knowledge." *Journal of Economic Theory*, 26: 17–27.
Modica, S. and Rustichini, A. (1993). "Unawareness: a formal theory of unforeseen contingencies. Part II." Discussion Paper No. 9404, CORE.
(1994). "Awareness and partitional information structures." *Theory and Decision*, 37: 107–24.
Monderer, D. and Samet, D. (1989). "Approximating common knowledge with common beliefs." *Games and Economic Behavior*, 1: 170–90.
(1990). "Proximity of information in games with incomplete information." Forthcoming *Mathematics of Operations Research*.
Morris, S. (1992). "Dynamic consistency and the value of information." Mimeo, University of Pennsylvania.
(1995). "The common prior assumption in economic theory." *Economics and Philosophy*, 11: 1–27.
(1996). "The logic of belief change: a decision theoretic approach." *Journal of Economic Theory*, 60: 1–23.
Myerson, R. B. (1991). *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.
Nielson, L. T. (1984). "Common knowledge, communication and convergence of beliefs." *Mathematical Social Sciences*, 8: 1–14.
Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. Cambridge, MA: MIT Press.
Pearce, D. G. (1982). "*Ex ante* equilibrium; strategic behavior and the problem of perfection." Working Paper, Princeton University.
(1984). "Rationalizable strategic behavior and the problem of perfection." *Econometrica*, 52: 1029–50.
Piccione, M. and Rubinstein, A. (1995). "On the interpretation of decision problems

with imperfect recall." Forthcoming *Games and Economic Behavior*.
Reny, P. (1985). "Rationality, common knowledge, and the theory of games." Mimeo, Princeton University.
(1992). "Rationality in extensive-form games." *Journal of Economic Perspectives*, 6: 103–18.
(1993). "Rationality in extensive-form games." *Journal of Economic Theory*, 59: 627–49.
Rosenthal, R. W. (1981). "Games of perfect information, predatory pricing and the chain-store paradox." *Journal of Economic Theory*, 25: 92–100.
Rubinstein, A. (1989). "The electronic mail game: strategic behavior under almost common knowledge." *American Economic Review*, 79: 385–91.
Rubinstein, A. and Wolinsky, A. (1989). "On the logic of 'agreeing to disagree' type results." *Journal of Economic Theory*, 51: 184–93.
Samet, D. (1990). "Ignoring ignorance and agreeing to disagree." *Journal of Economic Theory*, 52: 190–207.
(1993). "Hypothetical knowledge and games with perfect information." Mimeo, Tel Aviv University.
Samuelson, L. (1992). "Dominated strategies and common knowledge." *Games and Economic Behavior*, 4: 284–313.
Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.
Sebenius, J. and Geanakoplos, J. (1983). "Don't bet on it: contingent agreements with asymmetric information." *Journal of the American Statistical Association*, 18: 424–6.
Selten, R. (1975). "Re-examination of the perfectness concept for equilibrium points in extensive games." *International Journal of Game Theory*, 4: 25–55.
Shin, H. S. (1989). "Non-partitional information on dynamic state spaces and the possibility of speculation." Mimeo, University of Michigan.
(1993). "Logical structure of common knowledge." *Journal of Economic Theory*, 60: 1–13.
Stalnaker, R. (1994). "On the evaluation of solution concepts." *Theory and Decision*, 37: 49–73.
Stinchcombe, M. (1988). "Approximate common knowledge." Mimeo, University of California, San Diego.
Stuart, H. (1995). "Common belief of rationality in the finitely repeated prisoners dilemma." Mimeo, Harvard Business School.
Tan, T. C. C. and Werlang, S. R. C. (1984). "On Aumann's notion of common knowledge: an alternative approach." Mimeo, Princeton University.
(1988). "The Bayesian Foundations of solution concepts of games." *Journal of Economic Theory*, 45: 370–91.
Werlang, S. R. C. (1989). "Common knowledge." In Eatwell, J., Murrey, M., and Newman, P. (eds.), *The New Palgrave: A Dictionary of Economics*. New York, NY: W. W. Norton.