GAMES AND ECONOMIC BEHAVIOR 8, 123–163 (1995)

Subjective Games and Equilibria

EHUD KALAI AND EHUD LEHRER*

Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208

Received November 29, 1993

Applying the concepts of Nash, Bayesian, and correlated equilibria to the analysis of strategic interaction requires that players possess objective knowledge of the game and opponents' strategies. Such knowledge is often not available. The proposed notions of subjective games and of subjective Nash and correlated equilibria replace essential but unavailable objective knowledge by subjective assessments. When playing a subjective game repeatedly, subjective optimizers converge to a subjective equilibrium. We apply this approach to some well known examples including single- and multi-person, multi-arm bandit games and repeated Cournot oligopoly games. *Journal of Economic Literature* Classification Numbers: C73 and C83. © 1995 Academic Press, Inc.

1. INTRODUCTION AND SUMMARY

The concept of Nash (1950) equilibrium and its extensions to Bayesian equilibrium by Harsanyi (1967) and to correlated equilibrium by Aumann (1974) have become the main tools for modeling strategic interaction under uncertainty. In addition to their logical elegance these concepts give researchers the ability to make predictions in uncertain environments. How-

* The authors acknowledge valuable communications with Andreas Blume, Eddie Dekel-Tabak, Itzhak Gilboa, David Kreps, Sylvain Sorin, as well as participants in the 1993 Summer in Tel Aviv Workshop and in seminars at the University of California, San Diego; the California Institute of Technology; and the University of Chicago. The research was supported by NSF Economics Grants SES-9022305 and SBR-9223156 and by the Division of Humanities and Social Sciences of the California Institute of Technology. This paper is an extended version of "Bounded Learning Leads to Correlated Equilibrium" (see Kalai and Lehrer, 1991).

0899-8256/95 $6.00

ever, in applying these concepts researchers often make the following assumptions about the players in a game:

a. *A Complete Model.* Each player has complete detailed information about the identity of his opponents, their feasible sets, information structures, utilities, and so forth.

b. *Correct Common Priors.* When such information is missing, the player assigns correct prior probabilities to all possible values of the unknown parameters.

c. *A Closed Model.* Each player assumes that his opponents model the game exactly as he does and, moreover, that they too assign the same correct prior probabilities to all unknown parameters.

In many applications the above assumptions are unrealistic and therefore the prediction power of these models is suspect. For example, the complete model assumption seems to be too demanding even for highly rational players engaged in moderate size problems. The common prior assumption also seems highly incredible. Such problems are severe given the non-robustness of the equilibrium concepts, i.e., the fact that small changes in game specification can change drastically the predicted outcome, and this is especially so when combined with the closed model assumption.

The subjective approach proposed in this paper attempts to present a more realistic open model,[1] where each player takes only a subjective partial view of his individual decision problem. Since the approach makes weaker assumptions, its prediction power is, in general, weaker. For many applications, however, weaker and more reliable predictions seem preferable to sharp but less reliable ones. Moreover, despite the weaker assumptions, the theory still leads to a natural equilibrium concept and to meaningful results. In particular, it identifies parameters researchers must study in order to eventually obtain better, sharper, and more robust predictions.

Before discussing the general model, the results, and the relationship to earlier literature, we illustrate the approach and concepts through an n-person, infinitely repeated Cournot game with differentiated products. This game may be thought of as a multi-product extension of the Porter (1983) model, used later by Green and Porter (1984) and by Abreu *et al.* (1986).

At the beginning of every period, each of the n-producers chooses a non-negative quantity of his own good to produce for the coming period. Following these choices, a fixed probability distribution, which depends

[1] See Sorin (1992) for an earlier game-theoretic discussion of these notions.

on the chosen production levels, determines a random vector of n individual prices for the n-producers. Each producer is informed of his realized price, with which he computes his period's profit. The game continues in this manner where, prior to every period, each producer's strategy may depend on his own history of past production levels and realized prices. (There is a high level of imperfect monitoring here since a player sees only his own realized prices, but the general model presented in the sequel allows for a large variety of information systems, ranging from perfect, full information to only learning one's own payoffs.)

Let us consider the problem from a producer's viewpoint. In order to compute his optimal (dynamic) strategy using a Nash equilibrium approach, he must model the game. This requires detailed knowledge of his competitors, their production capabilities, information structures, utility, and so on. (Note that indirect competitors are also important since their actions affect the actions of immediate competitors). For example, a soft drink producer must have such information about other soft drink producers and about producers of related products—e.g., fruit juice, milk, and their related products. Assuming that he modeled the entire game correctly, as did all other players, the player will select the strategy which is his part of the one Nash equilibrium selected by everyone.

The assumption, that all the ingredients of the game are known, seems unrealistic here. To deal with the lack of knowledge, researchers often resort to Harsanyi's extension to Bayesian equilibrium. However, Bayesian equilibrium suffers even more from this difficulty. While it does not require the players to know the correct realization of a large vector of parameters, it requires them to know the prior probability distribution by which these parameters were drawn. No player is likely to know, or believe that his opponents know, the correct prior distribution. Just describing this distribution seems to be too large a task, no matter how rational players are.

The subjective approach, which we proceed to describe now, is significantly less demanding for the individual players, and in this sense it is more realistic (it is still highly unrealistic in requiring the player to solve an infinite horizon optimization problem).

Rather than modeling the parameters of all potential producers of related products, the subjective player restricts himself to assessing only his own *environment response function*. This function specifies, for every individual history of past production levels and realized prices, and for every contemplated next-period production level, the probability distribution over his next-period prices. In other words, it is his individual (dynamic) market demand function. Two facts are important to notice. First, the market and opponents' chosen strategies fully determine such a function. Second, whatever the real game and opponents' strategies are, finding

an optimal strategy reduces to finding an optimal strategy relative to the environment response function. In other words, this function compresses all the payoff relevant uncertainties of the game into a one-person decision problem, and in this sense it is the only part of the game of any relevance to him.

It follows that, from the point of view of an individual player, assessing the environment response function instead of the entire game and selected equilibrium strategies involves no loss of generality. In addition, the environment response function may be easier to assess since it is defined over a drastically smaller space. Moreover, many different games give rise to the same environment response function and they may be considered as one.

We therefore describe a subjective version of our Cournot game by the real repeated Cournot game, known possibly only to the modeler, together with an n-vector of subjective environment response functions assessed by the individual players. A vector of strategies in the above game is *subjectively rational* if each player strategy, i.e., dynamic production plan, is optimal against his subjective environment response function.

The question of how players assess environment response functions is important, difficult, and largely not addressed in the current paper (except for Example 4.4.1 which shows how to obtain them from a Bayesian equilibrium). Also not addressed are questions of completing and closing the model. These include questions like how players assess the assessments of others, and if their assessments can be completed in a consistent manner. Taking environment response functions as primitives of the model, the paper concentrates on studying the type of equilibrium that subjective optimizers may settle on.

A vector of subjectively optimal strategies may not be in equilibrium because of discrepancies between the subjective environment response functions and the actual ones. Under such discrepancies, as the game progresses (along a play path), a player's observations may contradict his beliefs. Such contradictions may be drastic, for example, if he experiences a price to which he had assigned probability zero, but they may also be just probabilistic, as occurs when we believe in one probability distribution but observe a sequence of events generated by a different distribution. These contradictions may cause the player to modify his beliefs and, as a result, his chosen strategy.

To rule out such contradictions, at a *subjective Nash equilibrium* we add for each player a condition of uncontradicted beliefs. This condition requires the coincidence of the correct individual forecast, i.e., the probability distribution on the actual individual price paths, with the individual's subjective forecast, i.e., the one computed using his subjective environment response function and his chosen strategy. Under such full coinci-

dence no contradictions are possible since the subjective probabilities of all events observable to a player are correct. Despite this full coincidence, however, the player may still be wrong in his assessments of conditional probabilities regarding events that are not on the play paths. And, for this reason, there are subjective Nash equilibria which are not Nash equilibria.

While the uncontradicted-beliefs condition just stated seems too demanding for interactions that have just started, it is natural for long ongoing interactions. Indeed, the paper presents general sufficient conditions under which subjective optimizers must converge with time to play a subjective equilibrium. Due, however, to the possibility of imperfect monitoring, the limit may be a *subjective correlated equilibrium* (see Aumann (1974, 1987) for the concepts of correlation and subjectivity in games; see Fudenberg and Tirole (1992) and Myerson (1991) for additional discussion) rather than a subjective Nash equilibrium. The past play that has lead them to equilibrium turns out to serve as a natural, unavoidable correlation device (see Lehrer (1991) for a study of this phenomenon). In our Cournot example, dependencies in the stochastic realizations of past market prices serve as a device correlating players' future beliefs and strategies.

The subjective approach proposed in this paper is a straightforward generalization of the above example. For a player in a general extensive form game an environment response function describes probabilities in the decision tree defined on his personal information sets. For any information set and a feasible action at the information set, the function specifies the probability distribution over his "next information sets." Restricting ourselves to perfect-recall games, "next information sets" is a well defined concept. Moreover, if we add to a player's set of information sets the terminal nodes with his individual payoffs, then knowing the correct environment response function is sufficient for the purpose of determining an optimal individual strategy. Thus, rather than assessing the whole game tree and opponents' strategies, it is sufficient for a player to assess his environment response function. Section 6 presents such an example.

The main body of this paper is devoted, however, to a special case of n-player infinitely repeated subjective games. In addition to being a useful modeling tool for applications, such games are convenient for the presentation of learning theories and convergence. The notion of a repeated game that we use may be thought of as a generalization of a one-player multi-arm bandit problem to n players. But it is significantly more general since it allows for all levels of monitoring, from perfect to a minimal level where each player only learns his own payoffs as the game progresses. We present a strong sufficient condition of compatibility of beliefs with the truth, as in Kalai and Lehrer (1993b), that guarantees convergence of subjectively optimal strategies to a subjective correlated equilibrium. The reader interested in weaker sufficient conditions, yielding weaker notions

of convergence, is referred to Lehrer and Smordinsky (1993). For extensions to long but finite games we refer to reader to Sandroni (1994).

While the notion of subjective equilibrium that the players learn to play is weaker than the objective counterpart, it still has interesting implications. In addition to some general properties, the paper illustrates implications to a one-player multi-arm bandit game and coincidence, under additional assumptions, with a Walrasian and Cournot equilibrium.

The need to distinguish subjective from objective knowledge in social interaction is not new or unique to this paper. The proposed notion of a subjective equilibrium has its roots already in Von Hayek (1937)[2]. He proposed that, at equilibrium, "the individual subjective sets of data correspond to the objective data, and ... in consequence the expectations in which plans were based are born out by the facts." Since Von Hayek, other economists have advocated and used such subjective notions, see for an example Hahn (1973). Aumann (1974) introduced and studied a model of a one-shot game with players assigning subjective probabilities to the outcomes of an exterior correlation device.

The newer literature on game theory contains an increasing number of concepts reducing the objective-knowledge assumed by the Nash approach, and moving in the direction of a subjective equilibrium. Rationalizable equilibria—see Bernheim (1984), Pearce (1984), and the more recent Rubinstein and Wolinsky (1994) rationalizable conjectural equilibria—are such examples. The notions most closely related to the ones proposed here are by Battigalli (1987) (see also Battigalli and Guaitoli (1988) and Battigalli *et al.* (1992)), the self-confirming equilibrium of Fudenberg and Levine (1993) (see also Fudenberg and Kreps (1988) for earlier motivation), and the earlier version of a subjective equilibrium proposed in Kalai and Lehrer (1993a). Our convergence result is closely related to earlier Bayesian learning papers, for example Jordan (1991) and Kalai and Lehrer (1993b).

To understand the connection of this paper to some of the earlier literature we need to consider special cases where the players know the game and uncertainty is restricted to opponents' choice of strategies. Such cases can be accommodated in the current model by having each player start with a subjective assessment of (distribution over) opponents' strategies which he uses, together with the known game, to compute a subjective environment response function. For games played with perfect monitoring, the notion of a subjective Nash equilibrium proposed in the current paper coincides with the subjective equilibrium introduced in Kalai and Lehrer (1993b). The proposed notion of a subjective correlated equilibrium

[2] See also Von Hayek (1974) and Matsuyama (1994) for a discussion of distortions introduced into economic policy analysis by the use of game theory.

generalizes the notion of a self-confirming equilibrium in Fudenberg and Levine (1993) (they allow for correlation in beliefs but not in strategies). Thus, we may think of the new notions of subjective equilibria as generalizations of the earlier notions to cases where the game itself, in addition to opponents' strategies, is not known and to cases where monitoring is not perfect.

Convergence to subjective correlated equilibria, as described in this paper, is also closely related to the convergence to subjective equilibria in Kalai and Lehrer (1993b). There, the repeated game is played with perfect monitoring and after every history a new repeated game, or a formal subgame, starts. The result is that the overall strategies induce an approximate subjective Nash equilibrium play in the subgames that start after long histories. (They also show that, for perfect monitoring games, approximate subjective Nash equilibrium approximately plays like an ε − Nash equilibrium.) In the current paper, due to imperfect monitoring, different players observe different histories and there may not exist formal subgames. Thus, we cannot discuss the new game that starts after a long history. Instead, we consider the correlated game starting after a long time T but across all histories of length $T - 1$, where the play and information revealed up to time $T - 1$ serve as a correlation device. The weaker result is that if T is large, the play starting from time T together with the past information up to T approximates a subjective correlated equilibrium. However, the probablistic methods used to obtain the needed merging of measures are exactly the ones used in Kalai and Lehrer (1993b).

The general literature on learning in strategic interaction has exploded over the last few years. It includes a very large number of models assuming bounded or myopic players, as well as a large number of rational learning papers. A sample of recent related rational-learning models includes Crawford and Haller (1990), Monderer and Samet (1990), Nyarko (1991b), Vives (1992), Koutsougeras and Yannelis (1993), Goyal and Janssen (1993), and Fujiwara–Greve (1993). Blume and Easley (1992), Jordan (1993), and Nachbar (1994) present excellent critical evaluations of this approach. Also, a growing literature on strategic rational learning concentrating on reputation and forgiveness aspects is emerging—see, for example, Cripps and Thomas (1991), Schmidt (1991), and Watson (1992). These directions are especially important since forgiving strategies invite experimentation, a phenomenon that sometimes creates a coincidence of subjective with objective equilibria.

The present paper is also a direct contribution to the literature on players who do not know their own utility functions, as in the case-based approach of Gilboa and Schmeidler (1992). We discuss this after we study the multi-arm bandit example.

Two interesting connections to explore are with subjective variants of

the Mertens and Zamir (1985) hierarchies of rationality model, as in Nyarko (1991a) and El-Gamal (1992), and with a new literature on endogenous uncertainty in economics, as in Chichilinsky (1992) and Kurz (1994). It seems that there should be close relationships between the proposed notion of subjective equilibria to the ideas presented in these papers.

2. Examples and Intuition

It is obvious that a subjective equilibrium may give rise to drastically different outcomes from those of an objective equilibrium. Even the well known repeated prisoners' dilemma game with myopic players may be "solved." For example, if each player believes that whenever he acts non-cooperatively he will be severely punished by an outside force, his best response is to cooperate repeatedly. Thus, the two players play the fully cooperative path as a response to their beliefs. Moreover, their beliefs are not contradicted, since neither ever acts non-cooperatively to find out that his fear of severe punishment is not founded.

Before we turn, however, to additional multi-person examples with less "dramatic" beliefs, we start with the well known one-person multi-arm bandit problem (see Whittle (1982) for the general problem, and see Rothschild (1974) and Banks and Sundaram (1993) for more recent references and economic applications). It turns out to be a special, stationary case of our general formulation. The need to distinguish between subjective and objective equilibria becomes very clear here.

Example 2.1 (A Two-Arm Bandit Game). The player in each period $t = 1, 2, \ldots$, has to engage in one of two possible activities, L and R. (A special case where these activities represent handles of two different slot machines motivates the name of this problem.) Each activity, L and R, has a stationary distribution, Π_L and Π_R, describing independent probabilities of realized payoffs when the corresponding activity is used. The player's goal is to maximize the expected present value of his total payoff, discounted by some fixed parameter. Clearly, the optimal objective solution is to repeatedly use the activity with the higher per-play expected value.

What makes the problem interesting is that the player may not know Π_L, Π_R, or both. Instead, as he plays he observes random payoffs generated by these distributions according to the actions that he uses. So every time he uses L he observes his realized payoff generated by Π_L, and the same for R. But in every period, before making his choice, he knows the full history of his past choices and resulting payoffs. In order to maximize his expected payoff, depending on his discount parameter and subjective

beliefs, it may pay him to experiment with both activities in order to learn something about their payoff distribution. Clearly, higher discount factors, representing more patient players in our conventions, lead to more experimentation, even if some immediate payoffs may seem to be sacrificed. But the problem is difficult and the question of how much and how to experiment depends in a fairly complex way on the subjective beliefs. These are described by prior probability distributions on sets of possible payoff distributions associated with each activity.

Suppose, for our example, that activity L generates payoffs of \$0 or \$2 with equal probability, i.e., $\Pi_L(0) = \Pi_L(2) = 0.5$. Let us also assume that the player knows that. On the other hand, he does not know Π_R and assigns positive probabilities λ^G and λ^B ($\lambda^G + \lambda^B = 1$) to it being one of two possible distributions Π^G and Π^B. The "good" distribution Π^G has $\Pi^G(2) = 0.6$ and $\Pi^G(0) = 0.4$, but the "bad" distribution has $\Pi^B(2) = 0.4$ and $\Pi^B(0) = 0.6$. The following scenarios give rise to equilibria, or the lack of such, of different types.

Scenario 1. $\Pi_R = \Pi^B$, λ^B is high, and the player chooses to play repeatedly activity L. This is an objectively optimal strategy, since he chooses the optimal strategy against the true payoff distributions of the two machines. It is also a subjective equilibrium since, for sufficiently high λ^B, the best response is not to experiment and to just use activity L. Moreover, his beliefs are not contradicted since the only uncertainty is regarding Π_R but he never uses R.

Scenario 2. As above, λ^B is high and the player uses repeatedly activity L, but now the real payoff distribution $\Pi_R = \Pi^G$. Now his strategy is still a subjective equilibrium, i.e., he is best responding to beliefs that are not contradicted, but it is not an objective equilibrium. If the player knew that $\Pi_R = \Pi^G$ he would not want to stay with the constant left strategy.

Scenario 3. λ^G is one, the player uses repeatedly activity R, but $\Pi_R = \Pi^B$. This is obviously not an objectively optimal solution. But also the subjective equilibrium fails. While the player maximizes against his beliefs, with increasingly high probability he will find out that his beliefs are wrong, i.e., he will observe persistence inconsistencies between his empirical payoffs and his beliefs. In other words, the condition of uncontradicted beliefs is violated here.

The previous example with the three scenarios illustrates the relationships of the different equilibria. Every objective equilibrium is a subjective one, when the subjective assessments happen to coincide with the true distributions. However, when the subjective assessments are not accurate, as in Scenario 2, we may have a discrepancy. Thus, the set of subjective

equilibria is really larger. But, as Scenario 3 illustrates, not all strategies and beliefs constitute subjective equilbria.

In Scenario 2 above, the discrepancy between subjective and objective equilibria is due to the fact that the player does not "know the game" he is playing. In this example, he does not know the payoff rules. When we move to multi-player situations, different types of information imperfections may cause such discrepancies. In the next example, even though both players fully know the game, imperfect monitoring (of each other's actions) alone gives rise to a subjective equilibrium which is not objective (Nash equilibrium). Moreover, from a modeler's viewpoint, this subjective equilibrium seems as appealing as the Nash equilibria.

EXAMPLE 2.2 (Acting in the Dark). This symmetric two-person game has two actions for each player: "r" for rest and "a" for act.

	r	a
r	0, 0	0, 1
a	1, 0	−1, −1

A player choosing r is paid 0 regardless of his opponent's choice. A player choosing a, on the other hand, is paid 1 if his opponent chooses r, but -1 if his opponent chooses a too.

We assume that the two players choose their actions repeatedly and simultaneously in the beginning of periods $t = 1, 2, \ldots$. However, in each period, after the choices are made, each player is only told his realized payoff and is not told his opponent's choice. We assume also, as is done in standard game theory models, that all the above is commonly known to both players. This implies that when a player chooses to rest he learns nothing about his opponent's choice. But when he chooses to act he learns his opponent's choice indirectly, through his payoff and knowledge of the game.

Let A be the constant strategy of acting in each period and R be the constant rest strategy. It is easy to see that (A, R) and (R, A) are objective Nash equilibria of the repeated game with imperfect monitoring. These are equilibria because the best reply to A is R and vice versa.

What about (R, R)? The first player may be playing R because he thinks that player two is playing A (as in the Nash equilibrium (R, A)). With the imperfect monitoring he never finds out that he is wrong and playing R against the conjecture that the other is playing A is as justified as playing R when the other player really plays A. Applying this reasoning to both players, it seems that (R, R) is as reasonable an equilibrium as the above

two Nash equilibria (we are ignoring at this time issues of robustness and trembles). It is easy to see, following the suggested reasoning, that (R, R) is a subjective Nash equilibrium, even though it is not a (objective) Nash equilibrium.

Our next example is also of a two-player game. It illustrates several important items. First, moving to correlation, it shows a subjective correlated equilibrium which is not an objective correlated equilibrium of the same game. Second, it illustrates the process of learning and converging, in one step here, to the subjective correlated equilibrium. Finally, it illustrates the generality of a class of games included in our model. In particular, our stage game can be viewed as a multi-person, multi-arm bandit problem.

EXAMPLE 2.3 (Winners and Losers Acting in the Dark). As in the previous example we consider a two player game with each one having to repeatedly choose between resting (r) or acting (a), and each being informed only about his resulting payoffs. Again, a player that rests receives a zero payoff and a player that acts, at a period where his opponent rests, receives a payoff of 1. The above information is known to both players. However, now when both players act a random pair of payoffs will be generated according to a fixed probability distribution $\Pi_{a,a}$.

	r	a
r	0, 0	0, 1
a	1, 0	$\Pi_{a,a}$

We consider two scenarios that may arise as we vary the beliefs and actual payoffs when both players choose to act.

We first restrict ourselves to the case where $\Pi_{a,a}$ can be only one of two possible distributions: $\Pi_{a,a} = \Pi^{W,L}$ or $\Pi_{a,a} = \Pi^{L,W}$. Under $\Pi^{W,L}$ player 1 is a "likely big winner" and player 2 is a "likely loser." Specifically, we define the probabilities for pairs of payoffs to be $\Pi^{W,L}(10, -1) = 0.99$ and $\Pi^{W,L}(-1, 10) = 0.01$. Under $\Pi^{W,L}$ the roles are reversed, and thus $\Pi^{L,W}(10, -1) = 0.01$ and $\Pi^{L,W}(-1, 10) = 0.99$.

Scenario 1: A standard common knowledge game. Nature moves first and chooses randomly with equal probability $\Pi_{a,a}$ to equal $\Pi^{W,L}$ or $\Pi^{L,W}$. The realized choice, which is to be fixed now for the duration of the infinite game, is not revealed to the players. However, following standard game theoretic models, we assume that all the information above is common knowledge.

If the players are sufficiently patient, then each would want to learn if he is a "likely big winner" or a "likely loser" in order to continue playing the game optimally. A reasonable Nash equilibrium of this Bayesian game has each player experimenting by acting at the first stage. If he loses he stops acting forever, but if he wins 1 or 10's he acts again. (After a while the computation of equilibrium becomes more complicated, since every time that he receives a 10 or a -1 he can update his prior as to the underlying $\Pi_{a,a}$ being $\Pi^{W,L}$ or $\Pi^{L,W}$. Like the one-arm bandit problem, this is a relatively simple analysis. Once a player decides to rest at some stage he receives no new information. Assuming therefore a "once-rest, rest-forever" strategy facilitates the computation of a relatively simple equilibrium. We choose not to complete this computation here, since it is tangential to the points we wish to make.)

Scenario 2: Both players are wrong and learn in one step to play a subjective correlated equilibrium which is not objective. Suppose the players believe everything as in Scenario 1 and, therefore, choose Nash equilibrium strategies of the type described there, i.e., they both act in the first period in order to discover who is the winner. But assume that they are both wrong in that the payoff distribution of both acting $\Pi_{a,a}$, is really the random distribution Π^R, defined as follows. $\Pi^R(10, -1) = \Pi^R(-1, 10) = \Pi^R(-1, -1) = 1/3$. In other words, in each period when they both act it is equally likely, independent of the past, that one will win, the other will win, or that they will both lose.

Since both players choose to act in the first period, they will be paid according to a random draw of Π^R. Therefore, there are three possible developments from period two on. With 1/3 probability, the first draw is a $(10, -1)$. Following this, and holding the beliefs described in Scenario 1, they will each assign high subjective probability to $\Pi_{a,a} = \Pi^{W,L}$ and will continue by playing the constant strategies (A, R). Similarly, with 1/3 probability they will be paid $(-1, 10)$, assign high probability to $\Pi^{L,W}$, and play (R, A). But also with 1/3 probability they will draw $(-1, -1)$ in the first period. This will lead each player to assign high probability to the distribution in which he is a loser and as a result the pair of constant strategies (R, R) will be played.

So if we consider the game, starting from period two on, we have a correlated strategy assigning probabilities of 1/3 each to (R, A), (A, R), and (R, R). Correspondingly, we have correlated beliefs where, with a probability of 1/3 each, the players respectively assign high likelihoods to the pairs of payoff distributions $(\Pi^{L,W}, \Pi^{L,W})$, $(\Pi^{W,L}, \Pi^{W,L})$ and $(\Pi^{L,W}, \Pi^{W,L})$. This is a subjective correlated equilibrium, since from period 2 on the correlated strategies are best response to the correlated beliefs and the induced subjective distributions on the future play of the game coincide with the objective one.

Consider, for example, the initial message to be the draw $(-1, -1)$. Now each player believes that he has encountered an acting opponent in the first period. Moreover, given that he lost (he does not, of course, even consider the possibility that his opponent lost too), his updated posterior beliefs are that he is a likely loser and he decides to stay our forever. Since both stay our forever, their incorrect beliefs are never contradicted.

Since the actual expected payoffs of the action vector (a, a) are (2.66, 2.66), one can check that the correlated strategies assigning probabilities 1/3 each to (A, R), (R, A), and (R, R) are not a (objective) correlated equilibrium of the repeated game.

3. Subjective Equilibrium of a Single Decision Maker

We consider a player with a nonempty countable (which can be finite) set of *actions* A, a countable set of *consequences* C, and a bounded von Neumann–Morgenstern utility function $u: A \times C \to \mathbb{R}$.

Dynamically, the player will choose actions $a^1, a^2, \ldots$ from A. In every period t, after he chooses the action a^t, an outcome $c^t \in C$ will be stochastically determined, reported to him, and he will collect the payoff $u(a^t, c^t)$. The player's objective is to maximize the present value of his expected utility discounted by a fixed parameter λ, $0 < \lambda < 1$.

The above formulation implicitly assumes that the player knows A, C, u, and λ. What he does not know is the stochastic rule by which consequences are generated.

Examples of such problems are numerous. We will analyze the multi-arm bandit problem, where A represents a set of possible "arms" or activities to use, $c \in C$ represents a stochastically generated payoff, and $u(a, c) = c$. The stochastic choice of the consequence c in this example will be stationary and its distribution will depend entirely on the chosen a.

A more complex economic example concerns a producer in an oligopoly whose action a^t in each period t describes a chosen production level. Here, a consequence c^t describes his resulting market price. The stochastic determination of c^t is a function of his production level a^t, the production choices of his competitors, and a stochastic demand function which depends on the joint production vector plus a random noise. Here we will not assume stationary determination of consequences (prices) since the competitors are likely to change their production levels as they too observe the behavior of the market.

In the general formulation, the determination of consequences is described by a stochastic *environment response function* denoted by e. For

every *history* of actions and consequences, $h^t = (a^1, c^1, \ldots, a^t, c^t)$, and for any $t + 1$ period action a^{t+1}, e defines a probability distribution over C. Formally, $e|_{h^t a^{t+1}}(c)$ denotes the probability that the consequence c will be drawn after the play consisting of the history h^t followed by the action a^{t+1}. (Thus the above values must be nonnegative and sum to 1 over the possible values of c for any fixed h^t and a^{t+1}.) The unique empty history h^0 is allowed and thus $e|_{h^0 a^1}$ describes the distribution of initial outcomes as a function of every chosen initial action a^1. (When it does not create confusion, to simplify notation we will omit some time superscripts, e.g., write $e|_{ha}(c)$).

If the player knows the environment response function e, his problem is to choose a (*behavior*) *strategy* f to maximize the present value of his expected payoff computed with the distribution generated by his strategy and e. Formally such a strategy f assigns a probability distribution over the action set A for every history of past actions and outcomes. Thus, $f|_{h^t}(a)$ represents the probability that action a will be chosen in period $t + 1$ if the player observed the history h^t. (Fixing $h^t, f|_{h^t}(a)$ must sum to one as we vary $a \in A$.)

We choose not to restrict our analysis to *pure strategies*, where each $f|_{h^t}$ assigns probability one to a single $a \in A$. Such a restriction, even if not significant for the one player case, would limit the scope of the analysis in the sections that follow.

To define the expected present value of utility resulting from a strategy f, we first describe the underlying probability space. It consists of a set Z of infinite play paths of the form $z = (a^1, c^1, a^2, c^2, \ldots)$. For a history h^t, as described above, we will abuse notation and let it also denote the *cylinder set* in Z, consisting of all infinite play paths z whose initial t − period segment coincides with h^t. As usual, the σ–albegra used for Z is the one generated by all cylinder sets h^t, and to specify a probability on Z it suffices to assign consistent probabilities to all h^t's.

We do this inductively in the usual way. Given a strategy f and an environment reaction function e, we define $\mu_{f,e}(h^0) = 1$. For h^{t+1} described by h^t followed by $a^{t+1}c^{t+1}$, we define $\mu_{f,e}(h^{t+1}) = \mu_{f,e}(h^t) f|_{h^t}(a^{t+1}) e|_{h^t a^{t+1}}(c^{t+1})$.

Now we can define utility functions for strategies. First, the utility assigned to a play path $z = (a^1, c^1, a^2, c^2, \ldots)$ is computed by $u(z) = \sum \lambda^{t-1} u(a^t, c^t)$. The utility of a strategy f and an environment response function e is computed to be $u(f, e) = \int u(z)\, d\mu_{f,e}(z)$.

As stated earlier, the player's objective is to choose f that maximizes $u(f, e)$. However, since we assume that the player does not know e, he cannot solve the above problem.

Taking a subjective approach, we assume that the player holds an exogenously given subjective belief about the environment response function,

denoted by $\tilde{e}$, and that he chooses f to maximize $u(f, \tilde{e})$. But we do not assume that $\tilde{e}$ coincides with e. When this is the case we say that f is *subjectively optimal* relative to $\tilde{e}$. If f is optimal relative to the "real" e we say that it is *objectively optimal* or just optimal.

Remark 3.1. (A) *Beliefs over a Set of Possible Environments.* While in the above formulation the player's subjective belief is restricted to be a single environment response function it is really more general. For example, if the player assigned prior probabilities $q_1, q_2, \ldots, q_n$ to n possible environment response functions $\tilde{e}_1, \ldots, \tilde{e}_n$, he could replace this belief system by an equivalent single belief function $\tilde{e}$. This is done using the usual Bayes updating construction as, for example, in Kuhn's (1953) theorem. After every history h^t one computes posterior probabilities $\tilde{q}_1, \ldots, \tilde{q}_n$ for the environments $\tilde{e}_1, \ldots, \tilde{e}_n$ and assigns probabilities to the next outcome according to the $\tilde{e}_i$'s weighted by the updated posterior probabilities. We do such an explicit construction in our example of a multi-arm bandit problem discussed later.

(B) *Imperfect Updating of Environmental Response.* Updating posterior beliefs, as described above, assumes a type of consistency and perfect rationality on the beliefs of the player. However, the abstract formulation described by a single $\tilde{e}$, which is a function that can be freely defined after every history, allows for more general and imperfect updating. For example, a player with Bayesian posterior probabilities, $\tilde{q}_1, \ldots, \tilde{q}_n$, may choose to eliminate (reduce to zero) $\tilde{q}_i$'s whose values fall below some critical level and increase the other $\tilde{q}_i$'s proportionately.

The discrepancy between the real environment response function, e, and the subjective one, $\tilde{e}$, may make the player alert to the fact that his assessment is wrong. Given his choice of strategy f, his assessment of the stochastic evolution of his future outcomes is given by $\mu_{f,\tilde{e}}$, while the real evolution follows the distribution $\mu_{f,e}$. If, however, $\mu_{f,\tilde{e}} = \mu_{f,e}$ then it is impossible for him to detect, even with sophisticated statistical tests, that he is wrong. This is despite the fact that off the play paths serious discrepancies may exist between e and $\tilde{e}$. With such discrepancies, even if f is subjectively optimal it may be objectively suboptimal, but the player could not determine that and will have no cause to change his assessment or his strategy. This gives rise to the following definition.

DEFINITION 3.1. The strategy f with the environment response function e is a *subjective equilibrium* relative to the belief $\tilde{e}$ if the following two conditions hold.

1. *Subjective optimization;* f maximizes $u(f, \tilde{e})$; and
2. *Uncontradicted beliefs*; $\mu_{f,\tilde{e}} = \mu_{f,e}$.

Remark 3.2 (Optimizing Implies Experimenting). Reflecting on the definition above, a subjective equilibrium can be suboptimal because, and only because, its assessment of outcome probabilities off the equilibrium play paths is wrong. An obvious remedy to such a deficiency is for the player to experiment, in order to learn to the greatest extent possible the off-path outcome probabilities. Computing the optimal level of experimentation, however, may require some knowledge of the real distributions, which the player does not possess. But, under the subjective approach, it is naturally incorporated into his subjective optimization problem.

Consider, for example, the two-arm bandit player with two competing activities, L and R, of Example 2.1, with stationary payoff distributions Π_L and Π_R. Assume for simplicity, as we did there, that the subjective beliefs are accurate on the left, $\tilde{\Pi}_L = \Pi_L$, generating an expected utility of 1 for every use of L. On the other hand, for R, the player believes that there are two distributions Π^B and Π^G, which were drawn initially with probabilities 0.90 and 0.10, respectively. So, most likely, the right activity is bad. Recall that the corresponding expected values are 0.8 and 1.2. By the law of large numbers, sufficiently long use of R will reveal to the player whether Π^G or Π^B was drawn. Depending on his discount parameter, his subjective optimization will determine the optimal experimentation strategy. If the future is important enough, the 10% chance of the eventual generation of a payoff stream with an expected value of 1.2 in each period will dictate an initial experimentation period. But if future payoffs are sufficiently unimportant, it would be subjectively suboptimal to experiment.

The optimal strategy in the definition of subjective equilibrium above already includes a subjectively optimal level of experimentation. The actual computation of such optimal strategies is done using the well-known Gittins index, see Whittle (1982).

We will see in the sequel that under a certain condition, relating the belief to the truth, a subjective optimizer must converge eventually to a subjective equilibrium. In any finite time, however, he must converge only to an ε-subjective equilibrium where the subjective distribution, $\mu_{f,\tilde{e}}$, is only close to the objective one, $\mu_{f,e}$. To make this precise we must digress and discuss notions of closeness of distributions.

DEFINITION 3.2. For a given $\varepsilon > 0$ and two probability distributions, μ and $\tilde{\mu}$, we say that $\tilde{\mu}$ is ε-*close* to μ if for any event A, $|\mu(A) - \tilde{\mu}(A)| \leq \varepsilon$.

Remark 3.3: Interpretations of Closeness of Distributions. We say that $\tilde{\mu}$ is ε-near to μ if there is an event Q, with $\mu(Q)$ and $\tilde{\mu}(Q) \geq 1 - \varepsilon$, satisfying $|1 - \mu(A)/\tilde{\mu}(A)| \leq \varepsilon$ for every event $A \subseteq Q$ (we assume in the above that $0/0 = 1$).

As was shown in Kalai and Lehrer (1994), the two notions ε-closeness and ε-nearness are asymptotically equivalent, i.e., by making the distributions sufficiently close in one sense we can force them to be as close as we wish in the other sense. Thus, limit results, where we obtain eventual arbitrary closeness of two measures, are the same in both senses.

While the notion of ε-closeness is easier to state, the notion of ε-nearness is more revealing. First note that ε-closeness says little on small probability events. For example, we can have $\tilde{\mu}(A) = 2\mu(A)$ and still have $\tilde{\mu}$ be ε-close to μ provided that $\mu(A) < \varepsilon/2$. On the other hand, ε-nearness shows that this can be the case but not on events $A \subseteq Q$. Within the large set Q the ratios of the probabilities must be close to 1. This has important implications for conditional probabilities, which take on special importance in models with infinite horizons.

Recall that our discussion of closeness of the measures $\tilde{\mu}$ and μ is motivated to capture the idea that a player believing $\tilde{\mu}$ but observing events generated by μ is not likely to suspect that $\tilde{\mu}$ is wrong. The notion of ε-closeness captures this idea for large events. Our player, however, after a long play is likely to observe small probability events consisting of the intersection of many past events. His forecast of future events then will be obtained by assigning probabilities to future events conditional on having observed low probability events. Thus, if our notion of closeness of $\tilde{\mu}$ and μ is such that the conditional probabilities they generate remain close, then the player using $\tilde{\mu}$ is not likely to modify his beliefs even after playing for a long time.

ε-nearness, and thus its asymptotically equivalent notion of ε-closeness, fulfills this property to a large extent. Since

$$\frac{\mu(A|B)}{\tilde{\mu}(A|B)} = \frac{\mu(A \text{ and } B)}{\tilde{\mu}(A \text{ and } B)} \frac{\tilde{\mu}(B)}{\mu(B)},$$

we can deduce that if A and B are events in Q, no matter how small, then closeness to 1 of the two factors in the right side implies closeness of the conditional probabilities in the left side.

DEFINITION 3.3. Given $\varepsilon > 0$, a strategy f, and environments e and $\tilde{e}$, we say that f is an *ε-subjective equilibrium* in the environment e relative to $\tilde{e}$ if the following two conditions hold:

1. *Subjective optimization,* f maximizes $u(f, \tilde{e})$, and
2. *ε-uncontradicted beliefs,* $\mu_{f,\tilde{e}}$ is ε-close to $\mu_{f,e}$.

Convergence of a subjectively optimal strategy to a subjective equilibrium is not guaranteed in general but is true under sufficient conditions of compatibility of the beliefs with the truth. The relationships between

notions of compatibility, notions of convergence, and alternative notions of ε-subjective equilibrium involve detailed mathematical analysis. To proceed with the presentation of the subjective approach, we present one such notion of compatibility that works well with our notion of ε-closeness (or ε-nearness) as defined above. For more general conditions we refer the reader to Lehrer and Smorodinsky (1993).

DEFINITION 3.4. We say that the subjective evolution described by $(f, \tilde{e})$ is *compatible with* the one generated by (f, e) if the distribution $\mu_{f,e}$ is absolutely continuous with respect to $\mu_{f,\tilde{e}}$, $\mu_{f,\tilde{e}} \gg \mu_{f,e}$. This means that for every event A,

$$\mu_{f,e}(A) > 0 \Rightarrow \mu_{f,\tilde{e}}(A) > 0.$$

In other words, events considered impossible according to the subjective belief of the agent, i.e., having subjective probability zero, are really impossible, i.e., they have objective zero probability.

Our goal is to show that after a sufficiently long time T, a subjective optimizer will play essentially an ε-subjective equilibrium for arbitrarily small ε. To make this formal we need to describe the environment response functions and strategies induced on the "new" problem starting from time T on.

DEFINITION 3.5. Let e be an environment response function, f a strategy, and h a history of length t. Define the environment response function e_h and the strategy f_h induced by h by

$$e_h|_{\bar{h}a}(c) = e|_{h\bar{h}a}(c)$$

and

$$f_h|_{\bar{h}}(a) = f|_{h\bar{h}}(a).$$

The notation $h\bar{h}$ denotes the *concatenation* of the histories h and $\bar{h}$, i.e., the history whose length is the sum of the lengths of h and $\bar{h}$ obtained by starting with the elements of h and continuing with the elements of $\bar{h}$.

THEOREM 3.1. *Let f be a subjectively optimal strategy relative to $\tilde{e}$ in the environment e, and assume that $(f, \tilde{e})$ is compatible with (f, e). For every $\varepsilon > 0$ there is a time T such that with probability greater than $1 - \varepsilon$, f_h is an ε-subjective equilibrium in the environment e_h relative to the beliefs $\tilde{e}_h$ for every history h of length greater than T.*

The probability $1 - \varepsilon$ in the statement of the theorem is the objective one, computed by $\mu_{f,e}$.

Theorem 3.1, which we do not prove here, is a direct consequence of Blackwell and Dubins (1962). The following is an easy version of presenting merging of measures, sufficient for the proof of the theorem (see Kalai and Lehrer (1994) for elaborations).

Let $X = (X^1, X^2, \ldots)$ be a vector of discrete random variables having a joint probability of distribution μ. For every time t, let $X^{t-} = (X^1, X^2, \ldots, X^t)$, i.e., the past, and let $X^{t+} = (X^{t+1}, X^{t+2}, \ldots)$, i.e., the future. Let $\tilde{\mu}$ be another (possibly incorrect) distribution for X which is compatible in the absolute continuity sense, i.e., $\mu \ll \tilde{\mu}$. Then with μ-probability 1 there is random time T s.t. for all $t \geq T$, $\tilde{\mu}(X^{t+}|X^{t-})$ is ε-close (or ε-near) to $\mu(X^{t+}|X^{t-})$.

EXAMPLE 3.1: *The Multi-arm Bandit Problem.* Generalizing the two-arm example of the previous section, we think of A as any finite set of activities that can be used repeatedly in periods $t = 1, 2, \ldots$. A countable set of consequences C consists of real numbers representing possible payoffs. For each activity $a \in A$ there is a fixed probability distribution Π_a over C with $\Pi_a(c)$ describing the (past independent) probability of the outcome c being realized when the action a is taken. The player's goal is to choose a sequence of actions $a^1, a^2, \ldots$, with each $a^t \in A$, that will maximize the present value of his expected payoff. However, he does not know the distributions, Π_a's, and whenever he uses the action a^t at time t he is told his realized payoff, which was drawn according to Π_a. Naturally, he can use this and all previous information before making his next choice, a^{t+1}.

In our general formulation, this example is modeled with A and C being described as above, $u(a, c) = c$, and a stationary environment function $e|_{ha}(c) = \Pi_a(c)$. Our player, not knowing the functions Π but knowing the stationary structure of the model, assumes that for every a, the distribution Π_a was initially chosen from among m possible distributions $\Pi_a^1, \ldots, \Pi_a^m$ with positive prior probabilities $\lambda_a^1, \ldots, \lambda_a^m$. We assume that Π_a indeed equals Π_a^j for some j.

The subjective environment response function, $\tilde{e}$, is computed by the standard method of Bayesian updating. First we compute inductively posterior probabilities $\lambda_a^j|_h$, $j = 1, \ldots, m$, for every a and h. Initially, $\lambda_a^j|_{h^0} = \lambda_a^j$. And for a history of the form $\overline{h}$ obtained by concatenating a history h with an action-outcome pair $(\overline{a}, c)$, $\lambda_a^j|_{\overline{h}} = \lambda_a^j|_h$ if $\overline{a} \neq a$, and $\lambda_a^j|_{\overline{h}} = \lambda_a^j|_h \, \Pi_a^j(c)/[\sum_i \lambda_a^i|_h \, \Pi_a^i(c)]$ if $\overline{a} = a$. Then $\tilde{e}$ is defined by $\tilde{e}|_{h,a}(c) = \sum_i \lambda_a^i|_h \, \Pi_a^i(c)$.

Since we assumed above that Π_a is assigned positive prior probability, for every strategy f, $(f, \tilde{e})$ is compatible with (f, e). Thus by Theorem 3.1

for every $\varepsilon > 0$ we can find a large enough time T such that with probability of at least $1 - \varepsilon$ the strategy and beliefs of the player from time T on constitute an ε-subjective equilibrium.

COROLLARY 3.1. *Suppose that the activities are strictly ranked by expected value, i.e., distinct objective expected values are generated by distinct activities, then for every $\varepsilon > 0$ there is a (random) time t such that with probability greater than $1 - \varepsilon$ the subjectively optimizing player described above will use only one activity from time t on.*

Proof. We may assume without loss of generality that the subjectively optimal strategy, f, is pure. (If f is a subjectively optimal behavior strategy, we study any *pure* strategy in the support of f.) We show that with probability 1 there is a (random) time t from which time on the strategy f prescribes playing one arm only. This obviously implies the corollary.

Assume to the contrary that there exists an event R, with positive probability such that on every infinite history $h \in R$ there are infinitely many truncations of h, h^t, after which f uses at least two arms. We denote by f_{h^t}, the continuation of f after the finite history h^t.

From Theorem 3.1, we deduce that on almost every $h \in R$, f_{h^t} is a δ_t-subjective equilibrium, where $\delta_t \to 0$. We take one $h \in R$ and consider the sequence of times t such that f_{h^t} prescribes the arm a_1 as its first action and $f_{h^{t+1}}$ prescribes the arm a_2 ($a_1 \neq a_2$) as its first action. We proceed by the following lemmas to the contradiction.

LEMMA 1. *Let f_t be a δ_t-subjective equilibrium, where $\delta_t \to 0$; then any limit of any subsequence of f_t ($t \to \infty$) is a subjective equilibrium.*

Proof. Denote by σ_t the belief over the distribution of outcomes that justifies f_t. In other words, f_t is a best response to σ_t and, moreover, the distribution induced by σ_t over infinite strings of outcomes is δ_t-close to the real distribution.

Suppose that $\bar{f}$ is a limit of some subsequence of $\{f_t\}$. There is a subsequence of $\{\sigma_t\}$ which converges to, say, σ_∞. Thus, σ_∞ is a distribution over outcomes and it coincides with the real distribution on those arms that are played, according to $\bar{f}$, at least once with a positive probability. From compactness we deduce that $\bar{f}$ is optimal against σ_∞ and, furthermore, σ_∞ confirms the real distribution of the outcomes of all those arms played with a positive probability according to $\bar{f}$. Therefore, $\bar{f}$ is a subjective equilibrium. ■

LEMMA 2. *If $\bar{f}$ is a subjective equilibrium in the set-up of Corollary* 3.1, *it uses only one arm, with probability* 1.

Proof. Let A' be the set of those arms used with a positive probability under $\bar{f}$. Since $\bar{f}$ is subjectively optimal in the grand game (with the full

set of arms, A), it is also subjectively optimal in the reduced game (with the restricted set of arms A' only). As $\bar{f}$ is a subjective equilibrium with A it is an objective equilibrium with A' (simply because there is full knowledge about the expected payoffs of all the arms available, A').

However, as an objective equilibrium, with the restricted set of actions A', $\bar{f}$ should prescribe using only one arm, the best one. ■

Returning to the proof of the corollary, recall that f is pure, that f_{h^t} prescribes pushing the arm a_1 first, and that $f_{h^{t+1}}$ prescribes the arm a_2 on its first move. Denote by w^{t+1} the outcome that forms with the history h^t, the longer history, h^{t+1} (i.e., h^{t+1} is the concatenation of h^t and w^{t+1}). Since there exist only finitely many w^{t+1} and infinitely many t, we may assume that all the w^{t+1} are the same. As the probability to get w^{t+1} by using the arm a_1 is stationary, say, $p > 0$, we deduce that under f_{h^t} the probability of arm a_2 being played at the second stage is at least p. Therefore, any limit of $f_t, \bar{f}$, assigns to two arms (a_1 and a_2) positive probabilities (probability 1 to a_1 and probability of at least p to a_2). By Lemma 1, f is a subjective equilibrium. However, it assigns to two arms positive probability, which contradicts Lemma 2. ■

Remark 3.4. Players Who Do not Know Their Own Utilities. Learning one's own utility function is an important problem in decision theory—see, for example, Gilboa and Schmeidler (1992) for a new approach and recent references. It deals with a player that can choose repeatedly activities a from a set A but does not know his own utility function $w(a)$. Our formulation, through consequences, includes this problem as a special case despite the fact that we assume that the player knows $u(a, c)$ for every action a and consequence c. Simply consider the (degenerate) stochastic rule with $c = w(a)$ and $u(a, c) = c$. Now every time the player uses activity a he learns his consequence $c = w(a)$, i.e., his utility of a. Obviously this is a special case of the multi-arm bandit game which allows for a richer class of "noisy" observations about one's own utility.

4. Multi-person Subjective Equilibria

4.1. *The Repeated Stochastic-Outcome Game*

We now assume that there are n-players, $n \geq 1$, each having a countable set of actions A_i, a countable set of consequences C_i, a bounded utility function $u_i: A_i \times C_i \to \mathbb{R}$, and a discount parameter λ_i. Also, as before, each player knows his individual data above and would like to choose a sequence of actions, $a^1, a^2, \ldots$, to maximize the present value of his

expected utility. But, again, he does not know the rule of how his actions affect his consequences, i.e., his environment response function.

Taking the approach of the previous section, we will assume that each individual starts with a subjective belief about his environment, described by $\tilde{e}_i$, and chooses an optimal strategy f_i relative to $\tilde{e}_i$, and conclude that eventually each player will play a subjective equilibrium. However, now we are also interested in the *interactive* aspects of the resulting equilibria, and for that purpose we must first be explicit about how the actions of one player enter the environment function of another.

We describe these cross effects by a collection of probability distributions, Π_a, defined for every *action vector* $a \in A \equiv \times_i A_i$. More precisely, $\Pi_a(c)$ denotes the probability that the *consequence vector* $c \in C \equiv \times_i C_i$ be realized if the vector of actions a is taken. (Thus, for a fixed a, the above quantities must sum to 1 as we vary c.) Note that the distributions Π_a, together with the action sets A_i and the utility functions u_i, fully determine an n-person stage game, G. In this game, for every action vector a, player i's (expected) utility is computed to be $u_i(a) = \sum_c u_i(a_i, c_i)\Pi_a(c)$. We refer to such a game as a game with stochastic consequences.

The above game will be played repeatedly as follows. In every period $t = 1, 2, \ldots$ each player, being informed of his past actions and realized individual consequences, will choose an action $a_i^t \in A_i$. Then, based on the vector of choices, a^t, nature will choose a vector of consequences $c^t \in C$ according to the distribution Π_a. Player i will be informed of his own outcome, c_i^t, will collect the payoff $u_i(a_i^t, c_i^t)$, and will proceed to choose a_i^{t+1}, and so on. Overall individual payoffs will be the present value of the total expected utility discounted by the individual discount parameters, λ_i. We denote this infinitely repeated game with stochastic consequences by G^∞.

EXAMPLE 4.1.1: A Cournot Game with Differentiated Products. We assume that each of the n players is a producer of a certain good, with A_i denoting the set of his possible period production levels. Now C_i describes a set of period market prices producer i may realize. Thus, for every vector of production levels $a \in A$, $\Pi_a(p)$ describes the probability of the vector of individual prices $p = (p_1, \ldots, p_n)$ being realized. The utility of player i is defined as usual by his resulting revenue minus cost, $a_i p_i - g_i(a_i)$, with g_i denoting his producting cost function. Thus, in each period the player knows his previous production levels and prices, and based on this knowledge he chooses his next production level.

When all producers produce a *homogeneous product,* and face the same market price, we model the situation by restricting the support of each Π_a to p's with $p_1 = p_1 = \cdots = p_n$.

Remark 4.1.1: Imperfect versus Perfect Monitoring. While the general

formulation, with each player being informed only of his own realized actions and consequences, describes imperfect monitoring and other types of information imperfection, it includes as special cases games with more monitoring and common information. For example, in the Cournot game above, $c_i = (a_i, p_i)$ is the minimum amount of information the player needs to compute his period payoffs. But partial monitoring can be modeled by letting each player's reported consequence be $c_i = (a_1, \ldots, a_n, p_i)$. So the outcome reported to player i includes all the production levels but only his realized price. Perfect monitoring and full common knowledge of histories can be modeled by letting individually reported consequences include all production levels and all realized prices, i.e., $c_i = (a_1, \ldots, a_n, p_1, \ldots p_n)$ (but his utility is still determined only by his components, $a_i p_i - g_i(a_i)$).

Regardless of how the c_i's are Π are defined, however, under the convention that a player knows all his previous realized actions and consequences before choosing his next action, our games always have perfect recall in Kuhn's (1953) sense.

The general subjective approach assumes that a real "objective game," G^∞, as defined above, is to be played. We will depart, however, from the assumption that the players know the game. Instead, we will assume that they hold beliefs about their individual decision problems and that these beliefs are represented by subjective environment response functions as defined in the previous section. It will ease the exposition, however, if we first review and establish notations for the objective notions of Nash and correlated equilibria as well as introduce the concept of an objective environment response function.

Formally, we define a *history* of length t, h^t, to consist of a vector $(a^1, c^1, \ldots, a^t, c^t)$ where each $a^j \in A$ and $c^j \in C$. An individual *player history* $h_i^t = (a_i^1, c_i^1, \ldots, a_i^t, c_i^t)$ with each $a_i^j \in A_i$ and $c_i^j \in C_i$. A play path $z = (a^1, c^1, a^2, c^2, \ldots)$ induces finite histories h^t and finite individual histories h_i^t by taking projections to the first t elements and then taking projections to player i's components.

A strategy of player i is a function f_i describing the probability that he takes a specified action after a specific history. Formally, $f_i|_{h_i}(a_i)$ denotes the probability that he would choose action a_i after observing his individual history h_i.

Below we define the utility function $u_i(f)$ for every *strategy vector* $f = (f_1, \ldots, f_n)$. As usual, a *Nash equilibrium* is a vector f^* with each f_i^* maximizing $u_i(f_{-i}^*, f_i)$. (Here and elsewhere, f_{-i} denotes a vector of strategies of all players but i where (f_{-i}^*, f_i) denotes the vector where all players but i play their star strategy but i plays f_i). To define the (expected) utility functions one needs to first establish the probability space describing the possible plays of the game.

We let Z denote the set of (infinite) play paths, and as before we let h^t denote a history of a finite length t but also the cylinder set defined by it. Given a strategy vector f we define the probability distribution it induces on finite histories, μ_f, inductively. For the empty history $\mu_f(h^0) = 1$, and assuming that μ_f was defined for all histories of length t, we define it for histories h of length $t + 1$ by

$$\mu_f(h, a, c) = \mu_f(h) \times_i f_i|_{h_i}(a_i)\Pi_a(c).$$

Since the above construction defines consistent probabilities for all cylinder sets it defines the distribution μ_f on the set of play paths.

Now for every play path $z = (a^1, c^1, a^2, c^2, \ldots)$ we define $u_i(z) = \Sigma_t \lambda_i^{t-1} u_i(a_i^t, c_i^t)$ and for a vector of strategies f we define $u_i(f) = \int u_i(z)\, d\mu_f(z)$.

Often the strategies of the players in the repeated game are correlated since their choice depends on correlated past individual messages. Formally, such a *correlation device* is described by two components. First is a nonempty countable set of *message vectors*, $M = \times_i M_i$, with each M_i denoting the set of *player i's messages*. The second component is a probability distribution p defined on M.

A *vector of correlated strategies*, $f = (f_1, \ldots, f_n)$ for the game G^∞, is defined by appending a correlation device to the beginning of the game. This is done by replacing the unique empty history by all possible elements $m \in M$ and allowing a player's strategy to depend on his reported initial message m_i. Formally a history of "length zero" is now any element of M, a history of length t is a vector of the form $(m, a^1, c^1, \ldots, a^t, c^t)$ and a play path $z = (m, a^1, c^1, a^2, c^2, \ldots)$. Individual histories are described as before by projecting to the player's component. So an individual history of player i is a vector of the form $(m_i, a_i^1, c_i^1, \ldots, a_i^t, c_i^t)$. Now a *vector of correlated strategies* $f = (f_1, \ldots, f_n)$ has each f_i describe a distribution over player i's actions for every individual history with an initial individual message. In other words, it is a vector of standard behavior strategies for the game with the initial correlation device, the *correlated game*, (M, p, G^∞).

The utility of player i is computed as before to be his expected present value where the probability distribution on the expanded Z includes the initial distribution p. Thus we only need to modify the distribution over length zero histories by defining $\mu_f(m) = p(m)$. The probabilities of longer histories are defined inductively as before.

A vector of correlated strategies, f, is a *correlated equilibrium* of G^∞ if it is a Nash equilibrium of the correlated game (M, p, G^∞) as defined above, for some correlation device (M, p).

As in the previous section, we will be interested in the play of the repeated game starting after a long time T. In the "new" game correlation cannot be ruled out since each player strategy from time T on may depend on his consequences up to time T. And, in general, these consequences are correlated.

Formally, given a vector of strategies for G^∞, f, and a positive integer T, we define the *induced vector of correlated strategies* $f^T = (f_1^T, \ldots, f_n^T)$ (for the game starting at period $T+1$) as follows. M is the set of length T histories and p is the distribution μ_f restricted to M. Following a history consisting of an initial message m_i followed by h_i, f_i^T randomizes over A_i with the same distribution that f_i does in the original game after the history obtained by concatenating m_i with h_i.

Remark 4.1.2: Nash Equilibrium Induces Correlated Equilibrium in Later Games. It is easy to see that if we start with a Nash equilibrium f, then f^T as defined above is only a correlated equilibrium of the repeated game starting at time $T+1$, see Lehrer (1991). Thus, in general games with imperfect monitoring, Nash equilibrium "deteriorates" to become a correlated equilibrium after time. This observation has important implications for learning theories. It suggests that, in general, we can at most hope to converge to correlated equilibrium.

It is easy to check that in the construction above we could have started with a vector of correlated equilibrium for G^∞, to conclude that it induces a correlated equilibrium after any time T.

4.2. *The Individual Environment Response Functions of the Repeated Game*

As already discussed, to compute a best response strategy, a player need not know the entire game or his co-players' strategies. It suffices to know his own personal decision problem, determined by the game and their strategies. This decision problem can be fully described by an environment response function as discussed in Section 3. However, since we have n players now, we let e_i denote the environment response function of player i. Thus, $e_i|_{h_i a_i}(c_i)$ denotes the probability of player i's next consequence being c_i given that he observed the history h_i and took the action a_i.

Given the opponents' strategy vector, f_{-i}, the computation of the environment function of player i, e_i, is straightforward. For every history of length t h_i, action a_i, and outcome c_i, we choose a strategy f_i for player i under which the individual history h_i followed by a_i has positive probability (or simply let player i play the actions of h_i up to time t, then a_i, and anything afterward) and let μ_f be the induced distribution on play

paths. Then define $e_i|_{h_i a_i}(c_i)$ to be the μ_f conditional probability of c_i being player i's outcome at time $t + 1$, given the individually observable play $h_i a_i$. If under the opponents' strategies, $h_i a_i$ is impossible, no matter what strategy player i chooses, then $e_i|_{h_i e_i}$ can be defined arbitrarily (since this situation will not arise).

Following the earlier discussion, it is straightforward to conclude the following equivalence.

PROPOSITION 4.2.1. *A vector of strategies f is a Nash equilibrium iff each player's strategy, f_i, is optimal relative to his environment response function, e_i (induced by f_{-i}).*

The above discussion and definitions are also applicable to correlated versions of the repeated game, with an initial correlated device (M, p). In this case, as before, each zero-length history consists of a message vector m and all other histories, individual or not, start with an initial message. Again, a correlated strategy vector f is a correlated equilibrium if and only if each f_i is a best response to the individual environment response functions induced by f_{-i} (this is now in the game with initial correlation).

Equivalently, one can discuss these notions on the strategies and beliefs induced by initial messages. For m_i, a positive probability message for player i, let[3] f_{m_i} and e_{m_i} be his induced strategy and environment response function after receiving the message m_i ($f_{m_i}|_{h_i} = f_i|_{m_i h_i}$ and $e_{m_i}|_{h_i a_i} = e_i|_{m_i h_i a_i}$). The following proposition is obvious.

PROPOSITION 4.2.2. *A vector of correlated strategies in the game (M, p, G^∞) is a correlated equilibrium if and ony if for every player i and every positive probability message m_i, f_{m_i} is optimal relative to e_{m_i}.*

4.3. *The Subjective Game and Equilibrium*

In this section we assume that a game, G^∞ as defined before, is played, but that the players do not fully know the game. We assume that each player knows his own components, i.e., feasible actions, possible consequences, and utility functions.

We model the situation by assuming that each player assesses his environment response function e_i by an environment response function $\tilde{e}_i$. The player will choose a strategy f_i to be optimal relative to the *subjective environment function* $\tilde{e}_i$. These choices, made by all n-players, result in a vector of strategies $f = (f_1, \ldots, f_n)$, which, in turn, induce the real objective environments $e_1, \ldots, e_n$. As we already discussed in the one-

[3] The more accurate notations, f_{i,m_i} and e_{i,m_i}, are abused here to reduce the number of subscripts.

person case, there is no reason to assume that $e_i = \tilde{e}_i$ and that if significant differences do exist, the player will observe that his assessments are wrong, change his subjective beliefs, and modify his strategy.

However, an equilibrium situation arises even if $\tilde{e}_i \neq e_i$, provided that the disagreements of the two functions are restricted to be after histories that are not observable, i.e., have zero marginal probabilities. When this is the case for each player, we are in a multi-person subjective equilibrium of the game.

To make this precise let μ_{f_i, e_i} and $\mu_{f_i, \tilde{e}_i}$ be, respectively, the objective and subjective distributions induced on player i's play paths.

DEFINITION 4.3.1. Let $f = (f_1, \ldots, f_n)$ be a vector of strategies, and $e = (e_1, \ldots, e_n)$ be the induced environment functions. Let $\tilde{e} = (\tilde{e}_1, \ldots, \tilde{e}_n)$ be a vector of subjective environment response functions. The pair $(f, \tilde{e})$ is a *subjective Nash equilibrium of the game* G^∞ if for each player i the following two conditions hold:

1. *Subjective optimization,* f_i is optimal with respect to $\tilde{e}_i$;
2. *Uncontradicted beliefs,* $\mu_{f_i, e_i} = \mu_{f_i, \tilde{e}_i}$.

The beliefs a player holds at the beginning of the game, as described by his subjective environment function $\tilde{e}_i$, may depend on past stochastic observations. This dependency was ignored in the single player model of Section 3. However, in the multi-person case, if past observations of different players are correlated then it is useful to describe explicitly how they create correlation in the individual belief functions.

To permit the description of correlation, we replace the repeated game, G^∞, by one with a correlation device (M, p, G^∞). As described earlier, an individual strategy f_i can be thought of as a vector $(f_{m_i})_{m_i \in M_i}$, and for a vector of opponents' strategies, f_{-i}, we have the conditional environment response functions $(e_{m_i})_{m_i \in M_i}$. Recall that f is a correlated equilibrium iff every f_{m_i} is optimal relative to e_{m_i}.

A subjectively optimizing player can hold beliefs and choose strategies that depend on his individual message. Thus, we define a subjective environment response function in the game with correlation, $\tilde{e}_i$, to be a vector $\tilde{e}_i = (e_{m_i})_{m_i \in M_i}$. Note that under this definition the subjective player does not explicitly assess, nor even model, the correlation device. But his past messages can still affect his beliefs (this is in contrast to Aumann (1974) where the player's assessment focused on the correlation device).

DEFINITION 4.3.2. A *subjective correlated equilibrium* for G^∞ consists of a correlation device (M, p) as above, with a vector of strategies $f = (f_1, \ldots, f_n)$ of (M, p, G^∞) and a vector of (subjective) environment

response functions $\tilde{e} = (\tilde{e}_1, \ldots, \tilde{e}_n)$ satisfying for each player i and each positive probability message m_i the following two conditions:

1. *Subjective optimization,* f_{m_i} is optimal with respect to $\tilde{e}_{m_i}$; and
2. *Correlated uncontradicted beliefs,* $\mu_{f_{m_i}, \tilde{e}_{m_i}} = \mu_{f_{m_i}, e_{m_i}}$.

As already suggested, it is clear that every Nash equilibrium is a subjective Nash equilibrium, with $e_i = \tilde{e}_i$ for all i and, similarly, every correlated equilibrium is a subjective correlated equilibrium. However, the fact that the $\tilde{e}_i$'s may disagree with the e_i's off of the play path makes the set of subjective equilibria significantly larger than the corresponding objective notions. Subjective Nash equilibria, which are not Nash equilibria, could be of economic interest of their own, as can be seen in the following familiar example.

EXAMPLE 4.2: Competitive Equilibrium Is a Subjective Cournot Equilibrium with Finitely Many Producers. Consider a homogeneous-product repeated Cournot oligopoly game with n-identical producers. Each producer i has a constant marginal production cost of $\$g$/unit, with which he can produce any quantity a_i at every one of the discrete times $t = 1, 2, \ldots$. The market price in each period is deterministic and linear, i.e., $p = b - d \sum_i a_i$ for some positive b and d with $b > g$.

Consider a vector of production levels $a^* = (a_1^*, \ldots, a_n^*)$ resulting in a competitive market price $p = g$, i.e., $\sum a_i^* = (b - g)/d$. Suppose each player plays a constant strategy f_i^* which prescribes the constant production level a_i^* after every history. The vector of strategies $f^* = (f_1^*, \ldots, f_n^*)$ is not a Nash equilibrium of the repeated game since each firm i is making a zero profit which could be increased by reducing production.

Nevertheless, the above production levels are supported by a subjective equilibrium of the repeated game, if each of the finitely many players assumes that he cannot affect the prices. For example, assume that the outcome reported to each player at the end of each period consists of his own production level and realized market price. Let each player hold stationary beliefs described by the subjective environment response function $\tilde{e}_i|_{h_i a_i}(g) = 1$. That is, he assumes that with probability one the market price will be g regardless of past history of prices and regardless of his production level. Clearly, producing a_i^* is a best response to such $\tilde{e}_i$. Moreover, the price sequence $(g, g, \ldots)$ is assigned probability one by him and, indeed, it has probability one under f^*. So f^* does not contradict the beliefs $\tilde{e}$. Thus, we are in a subjective equilibrium.

It is easy to see in the above model that the only subjective equilibrium which is stationary in actions and beliefs is the competitive one. Thus, the only (doubly) stationary subjective equilibrium in the Cournot game

is the competitive one. This example illustrates that, while subjective equilibrium by itself may allow many outcomes in a game, in the presence of additional assumptions on beliefs it may lead to interesting conclusions.

Note that, in the above discussion, the stationarity of beliefs could be significantly weakened provided that we keep each player believing that his actions do not alter the price distribution. An interesting case of this type is when each player believes that tomorrow's price will be what today's price was.

4.4. *Convergence to Subjective Correlated Equilibrium*

In the previous section we justified the notions of subjective Nash and subjective correlated equilibria by arguing that players, finding themselves in such a situation, will have no reason to alter their beliefs or strategies. The condition of uncontradicted beliefs used was strong since it requires full coincidence of subjective and objective probabilities of all observable events. In this section we present sufficient conditions under which utility-maximizing players must converge to play such a subjective correlated equilibrium.

Since the individual strategies, however, may not in general converge to a stationary limit strategy, we will follow the same course as we did in the one person case. In other words, we will show that after a sufficiently long finite time they must play a subjective correlated ε-equilibrium for arbitrarily small ε.

DEFINITION 4.4.1. Let (M, p, G^∞) be a correlated game, f a vector of correlated strategies, $\bar{e}$ a vector of correlated subjective environment functions, and $\varepsilon > 0$. We say that $(f, \bar{e})$ is a subjective correlated ε-equilibrium if the following conditions hold.

1. *Subjective optimization,* for every player i and message m_i, f_{m_i} is optimal with respect to $\bar{e}_{m_i}$.

2. *Correlated ε-uncontradicted beliefs,* with p probability greater than $1 - \varepsilon$, a message vector m will be chosen with $\mu_{f_{m_i}, \bar{e}_{m_i}}$ being ε-close to $\mu_{f_{m_i}, e_{m_i}}$.

Before stating the convergence result we recall the terminology of Section 4.1. Let f be a vector of strategies of G^∞, e be the induced vector of environment response functions, $\bar{e}$ be a vector of (subjective) environment response functions, and t a positive integer. The correlated game induced from time t on is a correlated game (H^t, μ^t, G^∞), with H^t denoting all the possible histories of length t, and μ^t is μ_f restricted to the events in H^t.

f^t, e^t, and $\bar{e}^t$ are the concepts induced on the correlated game by the original game in the natural way, as already discussed.

We say that the players play a *subjective correlated ε-equilibrium from time t on* (correlated on the past) if $(f^t, \bar{e}^t)$ is a subjective correlated ε-equilibrium in the game (H^t, μ^t, G^∞).

Recalling the definition in Section 3, we say that $(f_i, \bar{e}_i)$ is compatible with (f_i, e_i) if μ_{f_i, e_i} is absolutely continuous with respect to $\mu_{f_i, \bar{e}_i}$. The following result is, mathematically, an immediate consequence of the convergence result for the one player case.

THEOREM 4.4.1. *Let f be a vector of strategies and $\bar{e}$ be a vector of subjective environment response functions. Suppose f and $\bar{e}$ satisfy the following two conditions for every player i:*

1. *Subjective optimization,* f_i is optimal relative to $\bar{e}_i$

2. *Beliefs compatible with the truth,* $(f_i, \bar{e}_i)$ is compatible with (f_i, e_i).

Then for every $\varepsilon > 0$ there is a time T such that for all times t, with $t \geq T$, from time t on, the players play a subjective correlated ε-equilibrium.

Remark 4.4.1: Starting with a Correlated Game. It is easy to see that Theorem 4.4.1 can be extended to the case that the original strategies were correlated. That is, instead of playing G^∞ directly, the players start at time zero with the observation of some correlation device and choose their strategies to be optimal relative to the message dependent subjective beliefs. The conclusion, that they will eventually play a subjective correlated equilibrium, will be identical to the one in the conclusion of the current Theorem 4.4.1.

EXAMPLE 4.4.1. Applications to Bayesian Equilibria. Bayesian equilibrium represents a special type of subjective optimization as we describe below. Therefore, the convergence result of this paper has important, if not yet fully understood, implications here.

Following the standard construction of such an equilibrium we let $T = \times_i T_i$ denote the set of *type vectors* with each T_i, *player i's type set,* assumed to be nonempty and at most countable. A prior probability distribution p is used to draw a type vector $\bar{t} \in T$ and each player i is then informed only of his own realized type $\bar{t}_i$.

There is a collection of possible n-person stage games $(G_t)_{t \in T}$, all sharing the same action sets, $A = \times_i A_i$. However, for every vector of types t, there are different stochastic rules generating vectors of consequences. In notations, $\Pi_{t,a}(c)$ denotes the probability of the consequence vector c being realized when the players, of the types specified by the vector t, take the actions specified by the vector a. After the type vector $\bar{t}$ is drawn

and each player is informed of $\bar{t}_i$, the infinitely repeated game $G_{\bar{t}}^{\infty}$ will be played.

The above formulation may seem unfamiliar due to its generality. The more common formulation is the special case with each consequence c_i constituting the player's payoff and $u_i(a_i, c_i) = c_i$. Also, the determination of payoffs (consequences) for a given vector of types is usually deterministic, i.e., $\Pi_{t,a}$ assigns probability one to a single vector of payoffs.

A vector of strategies $g = (g_i)_{i \in N}$, with each $g_i = (g_{t_i})_{t_i \in T_i}$, is a Bayesian Nash equilibrium if each g_{t_i} satisfies the usual best response property to the t_i-conditional distribution on opponents types assuming that they follow their equilibrium strategies. All of the above information is assumed to be common knowledge and we let e_{t_i} be the computed environment response function induced by (T, p, G, g) on player i's t_i-type. (Thus, alternatively, g is a Bayesian–Nash equilibrium if and only if each g_{t_i} is optimal relative to e_{t_i}).

To interpret the Bayesian equilibrium under the subjective approach we let the real repeated game G^{∞} be the realized game $G_{\bar{t}}^{\infty}$. For each player i we let the subjective environment response function $\bar{e}_i$ be $e_{\bar{t}_i}$, and the individual strategies f_i be $g_{\bar{t}_i}$. The real objective environment response function, e_i, is computed using the realized game $G_{\bar{t}}^{\infty}$ and realized opponents' strategies $(g_{\bar{t}_j})_{j \neq i}$ (e_i is different from $\bar{e}_i$ since the latter is computed not by the realized game and realized opponents' strategies, but by the posterior distribution over them given only player i's realized type). With this interpretation the two conditions of Theorem 4.4.1 are met. First subjective optimization is obviously satisfied directly under the definition of a Bayesian equilibrium. Beliefs compatible with the truth also hold. Since the beliefs that player type $\bar{t}_i$ holds about his environment are generated by a probability distribution over opponent strategies and realized game, assigning a positive probability to the chosen vector $(g_{\bar{t}_i})$ and $G_{\bar{t}}^{\infty}$, his forecast actually contains a grain of truth. That is, $\mu_{f_i, \bar{e}_i} = \lambda \mu_{f_i, e_i} + (1 - \lambda)\hat{\mu}$ with $\lambda > 0$. Since having a grain of truth implies absolute continuity, Theorem 4.4.1 is applicable and we can conclude that, for arbitrarily small positive ε, after a sufficiently long time the vector of strategies $g_{\bar{t}}$ will approximate a subjective correlated equilibrium of the realized game $G_{\bar{t}}^{\infty}$.

COROLLARY 4.4.1. *Let (T, p, G, g) be a Bayesian Nash equilibrium. For every vector of realized types $\bar{t}$ and for every $\varepsilon > 0$ there is a time T such that for every period $t \geq T$, from time t on, the induced strategies $(g_{\bar{t}_i}^{t})$ constitute a subjective correlated ε-equilibrium of the realized game $G_{\bar{t}}$.*

So even if the players started by playing an equilibrium of a Bayesian

game they must eventually play close to a subjective correlated equilibrium of the actually realized game. Moreover, this is a subjective correlated equilibrium with additional structure. Specifically, it is one in which the $\tilde{e}_i$'s can be justified by some actual opponents' strategies of the real game. In other words, an individual player uncertainty is only with regards to opponents' strategies and as if he knew the chosen payoff matrix. Characterizing the subset of such a subjective correlated equilibrium is an interesting open question.

5. Coincidence of Subjective and Objective Equilibria

The convergence theorem of the previous section illustrates conditions that must lead the players to a subjective correlated equilibrium. The subjective notion of equilibrium is, in most cases, more plausible than the objective counterpart, but as already discussed it entails a reduced prediction power. Since any player may hold his own individual hypothesis that justifies his actions, an outside analyst who wants to predict future outcomes must collect information about players' subjective beliefs. The potential contribution of subjective equilibrium to prediction power depends on the game and on players' beliefs. The preliminary examples in this section illustrate situations, with general conditions on beliefs, involving no loss of prediction power when compared to objective equilibrium. That is, subjective and objective equilibria predict the same behavior possibilities.

5.1. *Optimistic and Pessimistic Conjectures*

In the multi-arm bandit example discussed earlier, the player does not know the real distributions that determine his outcomes. A suboptimal arm will be repeatedly used whenever the payoffs of other arms are sufficiently underestimated. In other words, a subjective equilibrium in this case is not an objective one, since pessimistic conjectures regarding unused arms are held.

The same logic extends to multi-player games, as demonstrated in Example 2.2. It is natural to expect that, if we rule out pessimistic beliefs, a subjective equilibrium must be an objective one.

Let f be a vector of strategies of the infinite game, with or without correlation, and let e_i be the induced environment response function of player i. We say that $\tilde{e}_i$ has *optimistic conjectures relative to* f if, for every strategy g_i, $u_i(g_i, e_i) \leq u_i(g_i, \tilde{e}_i)$. In other words, under any possible strategy the player believes that he will do better (in a weak sense) than he actually would, given the game and opponents' strategies.

PROPOSITION 5.1. *Let* $(f, \tilde{e})$ *be a subjective correlated (resp. Nash) equilibrium with each* $\tilde{e}_i$ *holding optimistic conjectures relative to* f. *Then* f *is a correlated (resp. Nash) equilibrium.*

Proof. It suffices to show that for any player i, if his $\tilde{e}_i$ is optimistic and has the uncontradicted-beliefs property, then his strategy f_i must be objectively optimal. Suppose to the contrary that f_i is not optimal against the real e_i. Therefore, there exists a strategy g_i of player i satisfying $u_i(g_i, e_i) > u_i(f_i, e_i)$. However, by the optimistic conjecture assumption, $u_i(g_i, \tilde{e}_i) \geq u(g_i, e_i)$. Moreover, by the uncontradicted-beliefs property, $u_i(f_i, e_i) = u_i(f_i, \tilde{e}_i)$. As we combine the first two inequalities with the last equality we get $u_i(g_i, \tilde{e}_i) > u_i(f_i, \tilde{e}_i)$, which contradicts the optimality of f_i against $\tilde{e}_i$. ∎

The next example, using a familiar economic model, illustrates that subjective and objective equilibria may generate the same behavior pattern even if many important parameters of the game are not known. For this purpose, we first discuss equivalence of behavior.

Two strategy vectors f and g of G^∞ (or a correlated version of it (M, p, G^∞)) *play like* each other if the distributions they induce on the space of play paths, Z^∞, coincide, i.e., $\mu_f = \mu_g$. Note that when this is the case, players' payoffs, and even distributions over payoffs, under f and g coincide. Moreover, an outside observer with the ability to perfectly monitor all players' actions could not distinguish between f and g. Disagreements between f and g can only occur off the play paths, thus, with probability zero.

5.2. *Subjective Cournot Equilibrium Plays Like Cournot Equilibrium*

We consider n-producers (players) of an identical product in a market with a deterministic commonly known downward sloping demand function, D.

To fit our model, and to simplify the exposition, we make the following assumptions. The set of consequences, prices in this case, consists of all non-negative rational numbers. Thus, C_i is a countable set for $i = 1, \ldots, n$. Similarly, we let all players have the same set of actions, feasible production levels, $A_i = \{0, 1, 2, \ldots\}$. We assume that in each period the market price is established deterministically according to the vector of production levels, $a = (a_1, \ldots, a_n)$, by $c = D(\sum a_i)$. We also assume for simplicity that they each have a constant and positive marginal production cost, K. So if in a given period a player produces at a level a_i and the realized market price (determined by all production levels) is c, his period net profit is $a_i c - a_i K$.

We let (M, p, G^∞) be the above game with some initial correlation

device (M, p) describing the distribution of information available to the players prior to the start of the game. We assume that $(f, \tilde{e})$ is a subjective correlated equilibrium. Thus, each f_{m_i} is a best response to $\tilde{e}_{m_i}$ and $\mu_{f_{m_i}, \tilde{e}_{m_i}} = \mu_{f_{m_i}, e_{m_i}}$, where e_{m_i} is the real environment response function determined by the game and the other player's strategies, given m_i, while $\tilde{e}_{m_i}$ is the one induced by the subjective conjecture of i.

We assume that each player knows the demand function. Formally, we do so by assuming that for every Δ the distribution $\tilde{e}_{m_i}|_{h_i(a_i+\Delta)}$ coincides with the distribution $D(D^{-1}(\tilde{e}_{m_i}|_{h_i a_i}) + \Delta)$. Note that this rules out the price taking assumption we used earlier to obtain the competitive prices at a subjective equilibrium.

Our goal now is to show that f plays like some g which is a correlated equilibrium, or equivalently a Nash equilibrium of (M, p, G^{∞}).

We construct $g = (g_1, \ldots, g_n)$ as follows. For each player i, after every history h_i which has a μ_f-positive probability we define g_i to coincide with f_i, i.e., $g_i|_{h_i} = f_i|_{h_i}$. Note that this implies that g plays like f. For μ_f zero probability histories, h_i, define $g_i|_{h_i}$ to choose a large production level L with probability one. The level L is chosen in such a way that the amount $(n-1)L$, produced by $n-1$ producers, lowers market price below the marginal cost, K.

By the definition of g for every player i, f and g play alike. That is, player i cannot tell the difference between f and g because they induce the same distribution over the signals observed by player i.

In order to show that g is an equilibrium, we show that for every possible deviation, g_i', of player i, $u_i(g_i', e_i) \leq u_i(g_i, e_i)$, where e_i is the environment response function induced by g_{-i}.

We will show that g_i' is not a profitable deviation by showing that the outcome generated by (g_i', e_i) could be generated by some f_i' and $\tilde{e}_i$. Since f_i is individually optimal (against $\tilde{e}_i$), $u_i(f_i, \tilde{e}_i) \geq u_i(f_i', \tilde{e}_i) = u_i(g_i', e_i)$. As $u_i(f_i, \tilde{e}_i) = u_i(g_i, e_i)$ because f and g play alike, we conclude that $u_i(g, e_i) \geq u_i(g_i', e_i)$.

Recall that the demand function, D, is commonly known and that it has a negative slope. Suppose that after the history h_i the strategy f_i prescribes player i the action a_i with positive probability. Since f is a subjective equilibrium player i knows to predict the distribution over prices given his own a_i. As D is one-to-one, player i is able to forecast after h_i the distribution over the quantity produced by all his competitors. Therefore, he knows to predict the distribution over prices not only given a_i, but also given any other quantity player i may produce. We now deduce that after every history with positive probability (w.r.t. f) player i knows the distribution over prices induced by (g_i', e_i). In other words, the distribution over prices induced by (g_i', e_i) is the one induced by $(g_i', \tilde{e}_i)$ after every

history with positive probability. By iterating the same argument we infer that the probability assigned to history h_i by $(g_i', \bar{e}_i)$ and by (g_i', e_i) is the same, provided that h_i has a $(g_i, \bar{e}_i)$-positive probability.

Define $f_i'(h_i)$ to be identical to $g_i(h_i)$ for every history h_i which is positive w.r.t. $(g_i, \bar{e}_i)$. Otherwise, $f_i'(h_i)$ is zero. We will show that $u_i(f_i', \bar{e}_i) \geq u_i(g_i', e_i)$. Fix a time t and let h_i be a history of length t. Conditioned on h_i being $(g_i, \bar{e}_i)$-positive we get that $u_i(f_i', \bar{e}_i)$, which by definition equals $u_i(g_i', \bar{e}_i)$, is equal to $u_i(g_i', e_i)$. As for every other history h_i, since $f_i(h_i) = 0$ the return for player i is zero. On the other hand, the payoff $u_i(g_i, e_i)$ is at most zero (after h_i) because the total amount produced by all players drops market price below the cost per unit. Therefore, conditioned on h_i being a history with probability zero (w.r.t. $(g_i, \bar{e}_i)$), $0 = u_i(f_i', \bar{e}_i) \geq u_i(g_i, e_i)$.

We may conclude that any period t, $u_i(f_i', \bar{e}_i) \geq u_i(g_i, e_i)$ and therefore this is the case for the whole repeated game. This completes the proof, showing that g is an equilibrium.

6. Subjective Extensive Form Games

In the previous sections we restricted the subjective approach to repeated stochastic-consequence games. The extension to general extensive form games is straightforward.

We need only to generalize the definition of an environment response function. Recall that $e_{i|h_i a_i}(c_i)$ represented the probability that consequence c_i will be realized by player i after the plays described of the individual history h_i followed by the action a_i. The role of h_i's in the above must be replaced by the player's information sets. For every information set h_i, the a_i's following it must be restricted to actions feasible at this information set. The consequences, c_i's in the above, must be replaced by two types of objects. First, a consequence c_i can be a terminal node with its associated payoff to player i, provided that the action a_i taken at h_i can lead to such termination. But it can also describe a new individual information set, if following $h_i a_i$ the other players could lead player i back to a "next information set." Thus, $e_{i|h_i a_i}(c_i)$ is the probability of the play entering the information set c_i (allowing terminal nodes) conditional on being in the information set h_i and taking the action a_i.

We use the following example to illustrate the general approach.

Consider a three-stage alternating-offer bargaining between a seller, S, and a buyer, B. At stage one the seller can ask the buyer for two prices, VH (very high) or H (high). In the second stage, the buyer can accept the asked price, X, with $X = VH$ or $X = H$, yielding the respective payoffs

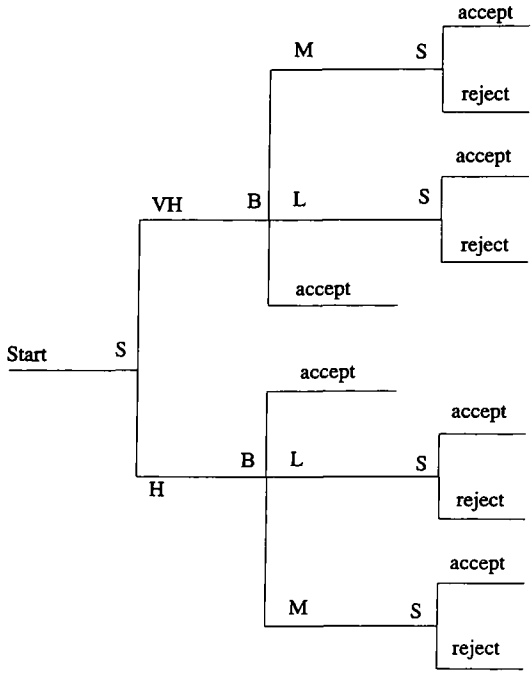

FIGURE 1

$X - R_s$, $R_b - X$, where R_s and R_b represent the respective reservation values. Or the buyer can counter propose two prices, L (low) or M (moderate), which the seller then accepts or rejects. If Y is accepted, $Y = L$ or $Y = M$, again the respective payoffs are $Y - R_s$, $R_b - Y$. If it is declined, the resulting payoffs are 0 and 0.

The extensive form game is described in Fig. 1. The decision tree of the seller, with O denoting nature's nodes, is described in Fig. 2. His subjective environment response function will specify six probabilities corresponding to the six arcs marked "accepted," L and M. For example, $e_s|_H$ (accepted) represents the probability that a seller's initial H offer will be accepted, and $e_s|_{VH}(M)$ represents the probability that an initial VH offer will be responded to with a counteroffer M.

However, drastically different games give rise to the same decision tree of the seller, as illustrated by the following two scenarios.

Scenario 1. The buyer consists of two players, b_1 and b_2, with hierarchical decision making. Upon hearing the asked price X, player b_1 can accept, counterpropose L, counterpropose M, or pass the decision to his

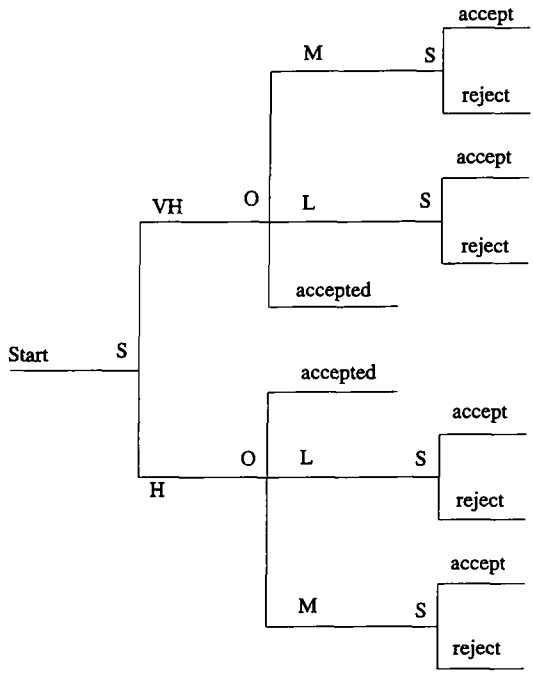

FIGURE 2

partner b_2. If b_1 passes, then b_2 decides whether to accept or counter with L or M. Now there are three reservation values, and if the item is sold at price P, the respective payoffs are $(P - R_s, R_{b_1} - P, R_{b_2} - P)$. Moreover, the extensive form game of Fig. 1 must be modified.

Note however, that the seller's decision tree in Fig. 2 is the same. We could construct a large number of scenarios, like the one above, all of which constitute different game trees with different groups of buyers, but all yielding the same individual decision tree for the seller. And for deciding his optimal strategy, the seller needs to assess only the response probabilities in this decision tree without considering all the possible game trees behind it.

Scenario 2. A Bayesian game with unknown buyer's reservation value. Suppose the single buyer has two possible reservation values chosen according to some prior probabilities. The buyer knows his realized value, but the seller does not. Now the extensive game has two versions of the original tree, with nature moving first and choosing which of the two trees to enter. The buyer knows which tree nature chose, but the

seller does not. Every pair of corresponding nodes in the two trees are put together for the seller in a single information set.

But, again, the seller's individual decision tree is unchanged and all he cares to know are the response probabilities to his various offers.

When we combine variations, as in Scenarios 1 and 2, we see that there is a large number of games, all yielding the seller the same individual decision tree. No matter what the game is, in order to choose an optimal strategy he only needs to know, or assess, the probability of various responses to his offers, i.e., his environment response function.

The buyer has a similar task. The real game defines for him a known decision tree. To decide on his optimal strategy, he needs to assess probabilities of how the seller (or sellers or sellers–agents or their types) will respond to his counteroffers.

The subjective extensive bargaining game is described by the real game together with the two decision trees and individually assessed environment response functions. If the situation modeled is a single isolated bargaining episode, then the reasonable solution concept is of subjectively optimal strategies, i.e., a pair of strategies with each being a best response to the subjective environment response function.

However, to repetitive bargaining situations that allow learning we may apply the concept of subjective equilibrium. It consists of a pair of subjectively optimal strategies that assign the correct probabilities to events on the play paths.

For example, suppose the seller assesses probability 1 to his proposed H price being accepted but 0.5 and 0.5 probabilities to his VH proposal being accepted or countered with M. Based on this assessment he proposes H and decides also to accept any counter proposal. The buyer chooses an optimal strategy which indeed counters H with an acceptance because he believes that rejecting H will lead the seller to reject any of his counterproposals. We are at a subjective equilibrium. Every player's expectations on the play path are met, even though their conjectures regarding off path responses may be wrong.

The above subjective equilibrium is closely related to earlier examples in Fudenberg and Kreps (1988) and to a self-confirming equilibrium in the Fudenberg–Levin (1993) sense.

A major difference, however, is that the current model does not assume that the players know the game. For example, the real game could be as in Scenario 1 above, yet with the seller behaving as if he is in the original game facing a single buyer.

Another interesting discrepancy between the players' model of the game and the actual game may be regarding the continuation of the game. The buyer may think that the game may continue with additional offers and

counteroffers, yet the seller may think that he must make a final response to the buyer's counteroffer. Since at the equilibrium described above the game ends after the first proposal, whoever is wrong regarding the possibility or impossibility of continuation never finds out.

REFERENCES

ABREU, D., PEARCE, D., AND STACCHETTI, E. (1986). "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica* **58,** 1041–1064.

AUMANN, R. J. (1974). "Subjectivity and Correlation in Randomized Strategies," *J. Math. Econ.* **1,** 67–96.

AUMANN, R. J. (1987). "Correlated Equilibrium as an Expression of Bounded Rationality," *Econometrica* **55,** 1–19.

BANKS, J. S., AND SUNDARAN, R. K. (1993). "Switching Costs and the Gittins Index," preprint, University of Rochester.

BATTIGALLI, P. (1987). *Comportamento Razionale ed Equilibrio nei Giochi e nelle Situazioni Sociali,* unpublished dissertation, Bocconi University, Milano.

BATTIGALLI, P., AND GUAITOLI, D. (1988). "Conjectured Equilibria and Rationalizability in a Macroeconomic Game with Incomplete Information," preprint, Bocconi University.

BATTIGALLI, P., GILLI, M., AND MOLINARI, M. C. (1992). "Learning Convergence to Equilibrium in Repeated Strategic Interactions: An Introductory Survey," *Ricerche Econ.,* in press.

BERNHEIM, D. (1984). "Rationalizable Strategic Behavior," *Econometrica* **52,** 1007–1028.

BLACKWELL, D., AND DUBINS, L., (1962). "Merging of Opinions with Increasing Information," *Ann. of Math. Statist.* **38,** 882–886.

BLUME, L., AND EASLEY, D. (1992). "Rational Expectations and Rational Learning," preprint, Cornell University.

CHICHILNSKY, G. (1992). "Existence and Optimality of a General Equilibrium with Endogenous Uncertainty," preprint, Columbia University.

CRAWFORD, V., AND HALLER, H. (1990). "Learning How to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica* **58,** 571–596.

CRIPPS, M., AND THOMAS, J. (1991). "Learning and Reputation in Repeated Games of Incomplete Information," preprint, University of Warwick.

EL-GAMAL, M. (1992). "The Rational Expectations of ε-Equilibrium," preprint, California Institute of Technology.

FUDENBERG, D., AND KREPS, D. (1988). "A Theory of Learning, Experimentation, and Equilibrium in Games," preprint, Stanford University.

FUDENBERG, D., AND LEVINE, D. (1993). "Self-Confirming Equilibrium," *Econometrica* **61,** 523–545.

FUDENBERG, D., AND TIROLE, J. (1992). *Game Theory.* MIT Press, Cambridge, MA.

FUJIWARA-GREVE, T. (1993). "A Note on Kalai–Lehrer Learning Model with a Generalized Semi-standard Information," preprint, Stanford University.

GILBOA, I., AND SCHMEIDLER, D. (1992). "Case-Based Decision Theory," preprint, Northwestern University.

GOYAL, S., AND JANSSEN, M. (1993). "Can We Rationally Learn to Coordinate?," preprint, Department of Economics, Erasmus University.

GREEN, E. J., AND PORTER, R. H. (1984). "Noncooperative Collusion under Imperfect Price Information," *Econometrica* **52,** 87–100.

HAHN, F. (1973). *On the Notion of Equilibrium in Economics: An Inaugural Lecture.* Cambridge: Cambridge Univ. Press.

HARSANYI, J. C. (1967). "Games of Incomplete Information Played by Bayesian Players, Part I," *Manage. Sci.* **14,** 159–182.

VON HAYEK, F. A. (1937). "Economics of Knowledge," *Economica* **4,** 33–54.

VON HAYEK, F. A. (1974)."The Pretence of Knowledge," Nobel Memorial Lecture.

JORDAN, J. S. (1991). "Bayesian Learning in Normal Form Games," *Games Econ. Behav.* **3,** 60–81.

JORDAN, J. (1993). "Three Problems in Learning Mixed-Strategy Nash Equilibria," *Games Econ. Behav.* **5**(3), 368–386.

KALAI, E., AND LEHRER, E. (1991). "Bounded Learning Leads to Correlated Equilibrium," preprint, Northwestern University.

KALAI, E., AND LEHRER, E. (1993a). "Subjective Equilibrium in Repeated Games," *Econometrica* **61,** 1231–1240.

KALAI, E., AND LEHRER, E. (1993b). "Rational Learning Leads to Nash Equilibrium," *Econometrica* **61,** 1019–1045.

KALAI, E., AND LEHRER, E. (1994). "Weak and Strong Merging of Opinions," *J. Math. Econ.* **23,** 73–86.

KOUTSOUGERAS, L. C., AND YANNELIS, N. C. (1993). "Convergence and Approximate Results for Non-Cooperative Bayesian Games: Learning Theorems," *Econ. Theory,* forthcoming.

KUHN, H. W. (1953). "Extensive Games and the Problem of Information," in *Contributions to the Theory of Games, 2* (H. W. Kuhn and A. W. Tucker, Eds.), Annals of Mathematics Studies, Vol. 28, pp. 193–216. Princeton, NJ: Princeton Univ. Press.

KURZ, M. (1994). "General Equilibrium with Endogenous Uncertainty," in *On the Formulation of Economic Theory* (G. Chichilnisky, Ed.), Cambridge: Cambridge Univ. Press, to appear.

LEHRER, E. (1991). "Internal Correlation in Repeated Games," *Int. J. Game Theory* **19,** 431–456.

LEHRER, E., AND SMORODINSKY, R. (1993). "Compatible Measures and Merging," preprint, Tel Aviv University.

MATSUYAMA, K. (1994). "Economic Development as a Coordination Problem," preprint, Northwestern University.

MERTENS, J. F., AND ZAMIR, S. (1985). "Formalization of Bayesian Analysis for Games with Incomplete Information," *Int. J. Game Theory* **14,** 1–29.

MONDERER, D., AND SAMET, D. (1990). "Stochastic Common Learning," *Games Econ. Behav.,* forthcoming.

MYERSON, R. (1991). *Game Theory, Analysis of Conflict.* Cambridge, MA: Harvard Univ. Press.

NACHBAR, J. (1994). "On Learning and Optimization in Supergames," preprint, Washington University.

NASH, J. F. (1950). "Equilibrium Points in *n*-Person Games," *Proc. Nat. Acad. Sci. USA* **36,** 48–49.

NYARKO, Y. (1991a). "The Convergence of Bayesian Belif Hierarchies," C. V. Starr Center Working Paper 91-50. New York: New York University.

NYARKO, Y. (1991b). "Bayesian Learning without Common Priors and Convergence to Nash Equilibrium," preprint, New York University.

PEARCE, D. (1984). "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica* **52,** 1029–1050.

PORTER, R. H. (1983). "Optimal Cartel Trigger-Price Strategies," *J. Econ. Theory* **29,** 313–338.

ROTHSCHILD, M. (1974). "A Two-Armed Bandit Theory of Market Pricing," *J. Econ. Theory* **9,** 195–202.

RUBINSTEIN, A., AND WOLINSKY, A. (1994). "Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability," *Games Econ. Behav.* **6,** 299–311.

SANDRONI, A. (1994). "Does Rational Learning Lead to Nash Equilibrium in Finitely Repeated Games?," preprint, University of Pennsylvania.

SCHMIDT, K. M. (1991). "Reputation and Equilibrium Characterization in Repeated Games of Conflicting Interest," preprint, Bonn University.

SORIN, S. (1992). "Information and Rationality: Some Comments," *Ann. Econ. Statist.* **25/26,** 315–325.

VIVES, X. (1992). "How Fast Do Rational Agents Learn?," *Rev. Econ. Studies*, in press.

WATSON, J. (1992). "Reputation and Outcome Selection in Perturbed Supergames: An Intuitive, Behavioral Approach," preprint, Stanford University.

WHITTLE, P. (1982). *Optimization Over Time*, Vol. 1. New York: Wiley.