American Economic Review 2008, 98:1, 72–86
http://www.aeaweb.org/articles.php?doi=10.1257/aer.98.1.72

Rational Expectations in Games

By Robert J. Aumann and Jacques H. Dreze*

A player i's actions in a game are determined by her beliefs about other players; these depend on the game's real-life context, not only its formal description. Define a game situation *as a game together with such beliefs; call the beliefs—and i's resulting expectation*—rational *if there is* common knowledge of rationality *and a* common prior. *In two-person zero-sum games, i's only rational expectation is the game's value. In an arbitrary game G, we characterize i's rational expectations in terms of the correlated equilibria of the* doubled *game 2G in which each of i's strategies in G appears twice.* (*JEL* C72, D83, D84)

Modern game theory was born in 1928, when John von Neumann published his Minimax Theorem. Inter alia, this theorem ascribes to all two-person zero-sum games a *value*—what rational players should expect to get.

Almost 80 years later, strategic game theory has not gotten beyond that initial point, insofar as the basic question of value is concerned. To be sure, we do have equilibrium theories: the initial concept of John F. Nash (1951) and its various refinements and coarsenings. But when the game is not two-person zero-sum, none of these theories actually tells the players what to expect. Even when there is just one Nash equilibrium, it is not at all clear that the players "should" expect its payoff.[1] Can one ascribe a value to each player—what she should expect—in an arbitrary *n*-person game?

As stated, the question has no answer; the problem is underspecified. Formally, a game is defined by its strategy sets and payoff functions. But in real life, many other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties each hold a third of the seats in parliament, or, say, 49 percent, 39 percent, and 12 percent, respectively. But the political situations are quite different. The difference lies in the attitudes of the players, in their expectations about each other, in custom, and in history, though the rules of the game do not distinguish between the two situations. Another example revolves around the ultimatum game (Werner Güth, Rolf Schmittberger, and Bernd Schwarze 1982), which, when played in different cultures, leads to systematically different outcomes (Alvin E. Roth et al. 1991).

Thus, if one is given only the abstract formulation of a game, one cannot reasonably hope for an expectation. Somehow, the real-life context in which the game is played must be taken into account. We are discussing not just a game, but a "game situation," i.e., a game played in a specific context; and we should be prepared to let a player's expectation depend upon the context—the "situation."

* Aumann: Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel (e-mail: raumann@math.huji.ac.il); Dreze: Center for Operations Research and Econometrics, Université Catholique de Louvain, 34 Voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium (e-mail: jacques.dreze@uclouvain.be). Aumann gratefully acknowledges research support from Israel Science Foundation grant 1343105.

[1] For example, see Example IIIB.

Call the player in whose expectation we are interested the *protagonist,* and designate her[2] *Player* 1 (P1). The essential element in the above notion of "context" is the protagonist's belief about the actions and beliefs of the other players. These are described by a *belief hierarchy:* her belief about what the others play, about what they believe she—and the others—play, about what they believe about *that,* and so on. Formally, therefore, we define a *game situation* Γ to consist of (a) a strategic game G as defined by its strategy sets and payoff functions, and (b) a belief hierarchy of P1 in G; we say that Γ is *based* on G and G *underlies* Γ.

To start with, each game situation has a well-defined *expectation:* the protagonist's expected payoff, as calculated from her strategy and her belief about the strategies of the others. But in itself, this doesn't get us very far: without further restrictions, even two-person zero-sum game situations may well have expectations that are very far from the value. Indeed, they may be anything at all between the protagonist's maximum and minimum payoffs.

To achieve a meaningful extension of von Neumann's value, we must take into account the interactive nature of games: that the players are rational, and reason about each other. This may be done by restricting attention to game situations with *common knowledge of rationality* (*CKR*) and *common priors* (*CP*).[3] The relevance of these two conditions to the value problem is brought out by

THEOREM A: *The expectation of any two-person zero-sum game situation with common knowledge of rationality and common priors is the value of the underlying game.*

Both *CKR* and *CP* are needed here; without either one, the result fails (Section IV, A and B). It is thus natural to define a *rational expectation*[4] in a game G as the expectation of a game situation based on G for which *CKR* and *CP* obtain. In Section IVD, we relate this definition to the classical notion of rational economic expectations, as originally promulgated by John F. Muth (1961).

What Theorem A says is that in two-person zero-sum games, rational expectations are *not* situation-specific: any such expectation must be the value of the game. But in general, rational expectations *are* situation-specific. Here, we investigate the set of rational expectations that can arise in situations based on a given game $G;$ i.e., the restrictions on such expectations that are implied by the formal definition of the game itself, whatever the specific situation may be.

On the face of it, the task of operationally calculating these expectations appears formidable; inter alia, because of the complexity of belief hierarchies. It is made approachable by our Theorem B, which characterizes the rational expectations in terms of the correlated equilibria (Aumann 1974) of the *doubled* game $2G$—which is the same as $G,$ except that each of the protagonist's strategies is listed twice.[5] To state it, recall that a *correlated equilibrium* of G is a probability distribution ρ on pure strategy profiles (a.k.a. n-tuples), such that if "chance" chooses a profile in accordance with ρ, and informs each player only of *his* strategy in the chosen profile, then it is optimal for him to play that strategy, assuming that the others play *their* strategies. The protagonist's expected payoff, after being told which strategy to play, is called a *conditional*

[2] The protagonist is female. The other players are of indeterminate gender; to distinguish them from the protagonist, we use masculine pronouns for them. Similarly, we use masculine pronouns for players in general, who may or may not include the protagonist.

[3] A player is *rational* if his strategy choice maximizes his expected payoff. *CKR* means that all players are rational, all know it, all know *that*, and so on. Roughly, *CP* means that differences in the players' probability assessments are due to differences in information *only*; for a precise definition, see Section I. If *CKR*—or *CP*—obtain, then all players know it, so it can be read off from the protagonist's belief hierarchy.

[4] This definition is not used in Enrico Minelli (1995).

[5] Doubling a game affects its correlated equilibria in a nontrivial way; they are *not* just doubled versions of the correlated equilibria of the original game.

payoff to ρ (because it is conditional on her information, namely the strategy). Correlated equilibria are described by a limited number of explicit linear inequalities;[6] so they, and the corresponding conditional payoffs, are explicitly describable in terms of G. We then have

THEOREM B: *The rational expectations in a game G are precisely the conditional payoffs to correlated equilibria in the doubled game* $2G$.

The above presentation contains the gist of the paper. The remainder is organized as follows: Section I discusses belief hierarchies, and characterizes them in terms of our main conceptual tool—that of *belief system*. Some auxiliary propositions are stated in Section II, and an alternative formulation of Theorem B appears in Section V. Section IV discusses two-person zero-sum games; in particular, it argues that the classical justifications for the value in such games (von Neumann and Oskar Morgenstern 1944) are less than convincing. Numerical examples are adduced in Sections III and IV. Section VI provides intuitions for Theorems A and B. Section VII discusses the relation of our concept to earlier work of Aumann (1987), which seems to give a different answer to the same question (VIIA), and to Nash equilibrium (VIIB). Section VIIC interprets rational expectations as "benchmark" outcomes of games; Section VIID relates them to Muthian rational expectations. Section VIII is devoted to the background and literature, and Section IX concludes. The Appendix provides formal proofs.

I. Belief Hierarchies and Belief Systems

Formally, a (strategic) n -person *game* G consists of n finite sets $S_1, S_2, \ldots, S_n$ (the *strategy sets* of the players) and n functions $h_1, h_2, \ldots, h_n$ from $S = S_1 \times S_2 \times \cdots \times S_n$ *to* $\mathbb{R}$ (the *payoff functions).* A belief hierarchy is most easily represented by means of a *belief system*[7] B for G consisting of

(i) For each player i a finite set T_i whose members t_i are called *types* of i*;* and

(ii) For each type t_i of each player i:

(a) A strategy[8] of i in G denoted $s_i(t_i)$; and

(b) A probability distribution on $(n - 1)$-tuples of types of the other players, called t_i's *theory.*

It may be seen that a player's type uniquely determines the whole hierarchy of his beliefs. Conversely, every belief hierarchy satisfying certain basic consistency conditions is the belief hierarchy of some type in some belief system (Jean-Francois Mertens and Shmuel Zamir 1985; see also Aumann and Heifetz 2002, sect. 8).

A *common prior* is a probability distribution π on $T_1 \times \cdots \times T_n$ that assigns positive probability to each type of each player, such that the theory of each type of each player is the conditional

[6] See Aumann (1987), Proposition 2.3. With two players having k and l strategies, respectively, there are $k^2 - k + l^2 - l$ inequalities in kl variables, plus the $kl + 2$ inequalities that say that the probabilities are nonnegative and sum to one (no more and no less).

[7] Originated by John C. Harsanyi (1967–1968).

[8] Member of S_i—what is usually called a *pure* strategy. Mixed strategies play no explicit role in this treatment. If a player wishes to use a mixed strategy, say by tossing a coin, he may certainly do so; but then the type describes the situation *after* the coin toss.

of π given that that player is of that type.[9] Less formally, such that each player's probability for a strategy profile is its probability under the common prior, conditioned on his information—i.e., on his being the type he is; examples are provided Section III. A type of a player is *rational* if the strategy it prescribes maximizes his expected payoff given its theory. Rationality is *commonly known* if this is so for all types of all players.[10] A belief system is *rational* if *CKR* and *CP* obtain. A *rational expectation* in *G* is an expected payoff of some type of the protagonist in some rational belief system for *G*.

II. Auxiliary Results

Our main results are Theorems A and B, stated in the introduction. The results in this section are used in analyzing the examples below, and in the proofs of Theorems A and B.

PROPOSITION C:

(i) Every conditional correlated equilibrium payoff—in particular, every Nash equilibrium payoff—is a rational expectation.

(ii) Rational expectations are unchanged by iterated deletion of strongly dominated strategies.

(iii) Every rational expectation is at least the protagonist's maxmin payoff.[11]

(iv) The rational expectations are covariant under multiplication of a player's payoffs by a positive constant, and under addition of a constant to a player's payoffs for fixed strategies of the other players.

Item (i) says that the notion of rational expectation is weaker than that of Nash or even correlated equilibrium. Nevertheless, it is strong enough to yield the value in two-person zero-sum games (Theorem A). Proposition C follows from Theorem B; for item (iv), we note that these transformations do not change the correlated equilibria.

To state the next proposition, recall that Roger B. Myerson (1997) called a game *elementary* if it has a correlated equilibrium that assigns positive probability to each strategy of each player, and all the inequalities associated with this equilibrium are strict.[12] We then have

PROPOSITION D: *Every elementary game has a maximum rational expectation, namely the (protagonist's) highest payoff at any strategy pair.*

Finally, we have

PROPOSITION E: *If all correlated equilibrium payoffs are the protagonist's minimax payoff v, then v is the only rational expectation.*

[9] In symbols, $\pi_i(t^{-i}; t_i) = \pi(t)/\pi(t_i)$ for each i and each t in $T_1 \times \cdots \times T_n$ where $\pi_i(\cdot\,; t_i)$ is t_i's theory and t^{-i} is the $(n-1)$-tuple of types assigned by t to players other than i.

[10] This is equivalent to the definition in terms of iterated knowledge given in the introduction.

[11] That is, it is her security level, what she can guarantee to herself when using mixed strategies.

[12] That is, if a player is informed that the chosen strategy profile calls for him to play a certain strategy, then it is strictly *better* for him to choose that strategy than any other strategy. Myerson showed that, in a certain sense, all games may be "reduced" to elementary games.

III. Examples

In the two-person games below, the row and column players are Rowena and Colin, respectively. Rowena is the protagonist.

	L	R
T	6,6	2,7
B	7,2	0,0

Figure 1A

The game G

	L	R
T	1/2	1/2
B	7/8	1/8

Figure 1B

Rowena's beliefs

	L	R
T	1/2	7/8
B	1/2	1/8

Figure 1C

Colin's beliefs

	L	R
T	7/22	7/22
B	7/22	1/22

Figure 1D

The common prior

FIGURES 1A THROUGH 1D

A. *Rational Expectations May Be Mutually Inconsistent*

The game G in Figure 1A ("Chicken") has three Nash equilibria: two pure, yielding (2, 7) and (7, 2), and one mixed, yielding ($4\frac{2}{3}$, $4\frac{2}{3}$). Consider now a belief system with four states, TL, TR, BL, and BR, with each player's probabilities *for* each state *in* each state as depicted in Figures 1B and 1C. For example, in BL as well as in BR, Rowena's probabilities for BL and BR are $\frac{7}{8}$ and $\frac{1}{8}$ respectively, while for TL and TR they are 0. Rowena has the two types T and B, Colin the two types L and R.

The expectation of Rowena's type B is $6\frac{1}{8}$. She attributes probability $\frac{1}{8}$ to Colin's type being R, in which case his expectation, too, will be $6\frac{1}{8}$. So in that case, the players will each expect $6\frac{1}{8}$. These expectations are mutually inconsistent; ($6\frac{1}{8}$, $6\frac{1}{8}$) is infeasible—it is outside the convex hull of the possible payoff vectors. And this in spite of common knowledge of rationality, which the reader may verify, and the existence of a common prior, depicted in Figure 1D.

BR is the conflict outcome in Chicken. We see here that conflict may occur even when the players reason perfectly rationally and attribute rationality to each other; both players know about the inconsistency, and indeed it is commonly known that it may occur. Contrary to common wisdom (or rather, foolishness), the conflict is not due to any irrationality, but simply to differing assessments, which may well ensue when players are provided with different information.

	L	R
T	1/2	1/2
B	1	0

Figure 1E

Rowena's beliefs

	L	R
T	1/2	1
B	1/2	0

Figure 1F

Colin's beliefs

	L	R
T	1/3	1/3
B	1/3	0

Figure 1G

The common prior

FIGURES 1E THROUGH 1G

To be sure, the mixed Nash equilibrium may also lead to conflict. But in that case, the players' assessments of the situation are not inconsistent. It is the inconsistency of the assessments that is noteworthy here.

The distribution in Figure 1D is a correlated equilibrium of G, and $6\frac{1}{8}$ is its conditional payoff given B.

Another belief system for G is depicted in Figures 1E–1G. Here, it is common knowledge that the conflict outcome BR is impossible. In particular, a type B Rowena expects 7 and knows that Colin is of type L, so expects 4. The payoff pair (7, 4) is, however, infeasible. Thus here again, the expectations of the players are mutually inconsistent, in spite of there being no irrationality in the system.

B. *Different Conditional Correlated Equilibrium Payoffs in 2G and G*

The game G of Figure 2A (Lloyd S. Shapley 1964) has a single Nash equilibrium, namely, (($\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{3}$), ($\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{3}$)), yielding the payoff (3, 3). Consider now a belief system with seven states, T_1R, T_2C, T_2R, ML, MR, BL, and BC, with each player's probabilities set forth in Figures 2B and 2C (as in Section IIIA); Rowena's strategy in rows T_1 and T_2 is T. If Rowena's type is T_1, the game situation has expectation 5. She knows that Colin's type is R,

	L	C	R
T	0,0	4,5	5,4
M	5,4	0,0	4,5
B	4,5	5,4	0,0

FIGURE 2A

The game G

so that his expectation is 4½. Thus Rowena *knows* that the players' expectations are the infeasible[13] pair (5, 4½)—in spite of common knowledge of rationality and a common prior (depicted in Figure 2D). Here, unlike in the previous example, 5 is not a conditional payoff to a correlated equilibrium of G, given any strategy of Rowena.[14]

But it *is* a conditional payoff to a correlated equilibrium of the doubled game $2G$, depicted in Figure 2E; the correlated equilibrium in question is depicted in Figure 2F. Note that if we eliminate the rows in which all the probabilities vanish, Figure 2F becomes Figure 2D. A similar relationship obtained in our first example, but with G instead of $2G$.

	L	C	R
T^1	0	0	1
T^2	0	2/3	1/3
M	1/2	0	1/2
B	1/2	1/2	0

Figure 2B

Rowena's beliefs

	L	C	R
T^1	0	0	1/4
T^2	0	1/2	1/4
M	1/2	0	1/2
B	1/2	1/2	0

Figure 2C

Colin's beliefs

	L	C	R
T^1	0	0	1/12
T^2	0	1/6	1/12
M	1/6	0	1/6
B	1/6	1/6	0

Figure 2D

The common prior

FIGURES 2B THROUGH 2D

0,0	4,5	5,4
0,0	4,5	5,4
5,4	0,0	4,5
5,4	0,0	4,5
4,5	5,4	0,0
4,5	5,4	0,0

Figure 2E

The doubled game $2G$

0	0	1/12
0	1/6	1/12
1/6	0	1/6
0	0	0
1/6	1/6	0
0	0	0

Figure 2F

A correlated equilibrium in $2G$

FIGURES 2E AND 2F

Alternatively, a correlated equilibrium of this game may be obtained by assigning probability 1/6 to the entries with payoffs (4, 5) or (5, 4). The associated inequalities are all strict, so the game is elementary. So by Proposition D, the maximum rational expectation is 5.

C. *The Set of Rational Expectations Need Not Be Convex*[15]

In each of the previous two examples adduced up to now, the set of rational expectations is an interval. In the game G of Figure 3, this is not so; here, there are precisely two rational expectations: 1 and 0.

	L	R
T	1,1	0,0
B	0,0	0,0

FIGURE 3

The game G

IV. Two-Person Zero-Sum Games

Von Neumann and Morgenstern (1944) advance two arguments to support the minmax value of two-person zero-sum games. One is the *equilibrium* argument; the other is the *guaranteed value* argument, which says that the row player can guarantee getting the value v, and the column player can guarantee not paying more than v, "so" rational players must reach precisely the value.

The equilibrium argument is of a formal, mathematical nature; one *proves* that there is a unique equilibrium payoff, namely the value. But the guaranteed value argument is more tenuous. We purposely put the word "so" in quotation marks, because there is a bit of a non sequitur there. Fully to justify this kind of argument, one needs a formal framework. In fact, though this "argument" appears to depend only on the rationality of the players, it is not true that players who

[13] The payoffs sum to 9½, whereas the maximum sum in the matrix is 9.

[14] By G's symmetry, we may suppose that Rowena's component of the strategy pair is T. The correlated equilibrium cannot then assign positive probability to TC, as Rowena's conditional payoff would then be < 5. So everything is eliminated by a sequence of strict dominations: C by L, then B by M, then L by R, then M by T, and finally R by C, leaving nothing. In fact, the highest conditional payoff to a correlated equilibrium of G, given a strategy of Rowena, is $5 - (1/3126) \approx 4.99968$.

[15] By contrast, with an ex ante viewpoint, as in Section VIIA, one gets the set of all unconditional correlated equilibrium payoffs, which is always convex.

are merely rational must necessarily reach the value; one needs *common knowledge* of rationality, and one also needs common priors (see Section IV, A and B).

One may think of Theorem A as reflecting the guaranteed value argument, but in a rather subtle way. The players do *not* actually guarantee the value. In many two-person zero-sum games,[16] it is in fact impossible to do so in pure strategies; and here, we think of the players as using pure strategies only. Rather, the protagonist *expects* the value. Guarantees enter the argument in showing that what she expects cannot be less than the value, because she *could*—by using mixed strategies—attain at least the value in expectation. One needs further arguments, revolving around the common knowledge, the common prior, and the zero-sumness to show that she also cannot expect more than the value.

Indeed, the current perspective shows exactly where the "classical" argument breaks down. It is true that Players 1 and 2 can guarantee v and $-v$, respectively; so, since the sum of payoffs is 0, the only feasible "individually rational"[17] payoff pair is $(v, -v)$. It is also true that any rational expectation (of either player) must be individually rational, that is, Proposition C (iii). What is *not* generally true is that the players' expectations must constitute a feasible pair, i.e., be "consistent." Indeed, we saw in Section III that inconsistent expectations are the rule rather than the exception; in particular, we saw that it is possible for Rowena to *know* that Colin expects a payoff that is inconsistent with the payoff she *knows* she is getting, even though rationality is commonly known and there is a common prior. On the face of it, there is no reason to suppose that a similar situation could not arise also in two-person zero-sum games.

But in fact, it cannot. Theorem A says that there is something special about two-person zero-sum games that makes it impossible. So this theorem goes considerably beyond the classical "guaranteed value" argument for the value.

As for the equilibrium argument, this is addressed in Section VIIB, where we discuss Nash equilibrium in general n-person games.

A. *Failure of Theorem A without Common Priors*

The game G is "matching pennies." With the depicted belief system—which has no common prior—it is common knowledge that it is optimal for each player to play that strategy with which his type is designated. In particular, Rowena's type T plays T and expects 0.8, whereas the value of the game is 0.

Careful consideration of the example leads to some discomfort. It is commonly known that Rowena believes[18] that Colin believes that Rowena does the opposite of what she really does—and vice versa,[19] i.e., that each player ascribes grave errors to the other. This is typical of situations without common priors. Common knowledge of rationality does obtain in this example.

B. *Failure of Theorem A without CKR*

Figure 5 depicts a common prior for a belief system in the game "matching pennies" (see Figure 4A). Type L_3 of Colin is irrational, but all other types of both players are rational. So, type T_1 of Rowena is rational, knows that Colin is rational, knows that he knows that she is rational, knows that he knows that she knows that he is rational, and knows that he knows that she knows that he knows that she is rational; moreover, T_1's expectation is 1, whereas the value of the game

[16] Specifically, unless the game has a pure strategy saddle point.
[17] This means that each player gets at least what he can guarantee to himself.
[18] Short for "ascribes probability 0.9 to."
[19] That is, with Rowena and Colin interchanged.

is 0. The example can be extended to an arbitrarily high level of iterated mutual knowledge of rationality; but, by Theorem A, not to common knowledge.

	L	R
T	1, −1	−1, 1
B	−1, 1	1, −1

Figure 4A

The game G

	L	R
T	0.9	0.1
B	0.1	0.9

Figure 4B

Rowena's beliefs

	L	R
T	0.1	0.9
B	0.9	0.1

Figure 4C

Colin's beliefs

FIGURES 4A THROUGH 4C

V. An Alternative Formulation of Theorem B

Theorem B is stated in terms of the doubled game $2G$. It can also be stated in terms of a set of "augmented" games, in each of which a single strategy of the protagonist is "doubled." Thus, given a strategy r_1 of the protagonist in G, define the *augmented* game G_{2r_1} as the n-person game in which just r_1 is replaced by two copies, and the payoff does not depend on which copy is used. We then have

	L_1	L_2	R_1	R_2	L_3
T_1	0.1	0.1	0	0	0
B_1	0.1	0	0.1	0	0
B_2	0	0.1	0	0.1	0
T_2	0	0	0.1	0	0.1
T_3	0	0	0	0.1	0.1

FIGURE 5
The common prior

THEOREM B′: *Rational expectations in the game G coincide with conditional correlated equilibrium payoffs in the augmented games* G_{2s_1}, *where* s_1 *ranges over the protagonist's strategies.*

Thus, α is a rational expectation in G if and only if the protagonist has a strategy s_1 such that in the augmented game G_{2s_1}, there are a correlated equilibrium ρ and a strategy such that α is the conditional payoff to ρ given that strategy.

VI. Intuitions for Theorems B′, B, and A

Two intuitions underlie our results: (i) that the common prior probability of a rational belief system (henceforth RBS; see Section I) B in a game G is essentially the same thing as a correlated equilibrium (CE) of a game G_B closely related to G—that in which each strategy of each player appears as many times as there are types that play that strategy in B; and (ii) that the conditional expectation of a strategy in a CE does not change when *other* strategies that are "identical"—i.e., have the same payoffs no matter what the other players do—are amalgamated. *Amalgamation* of identical strategies in a CE of a game means going to the corresponding CE of a game in which the amalgamated strategies are replaced by a single strategy, inheriting as prior probabilities the sums of the prior probabilities before amalgamation.

By (i), every rational expectation in G is a conditional correlated equilibrium payoff in G_B. Define a game G_{2s_1} by amalgamating in G_B all "identical" strategies other than some specified strategy s_1 of the protagonist, retained twice. Then by (ii), a conditional CE payoff in G_B for strategy s_1 is also a conditional CE payoff in G_{2s_1} for s_1. This yields Theorem B′, and Theorem B follows.

Finally, let G be a two-person zero-sum game with value v. Call Player 1 Rowena, Player 2 Colin. Let t_1 be a type of Rowena in an RBS. Rowena has a mixed strategy μ that yields $\geq v$ against any pure strategy of Colin, and so also against her belief under t_1 of how he plays. So at least one of her pure strategies must yield an expectation $\geq v$ under that belief. So by *CKR,* the strategy prescribed by t_1 must also yield $\geq v$; thus, all types of Rowena expect $\geq v$. Similarly, all types of Colin expect $\geq -v$; so the average of all of Colin's payoffs under the common prior π is $\geq -v$. If some type of Rowena would expect $> v$, then the average of all of Rowena's payoffs under π would be $> v$, contradicting the zero-sumness. So all types of Rowena expect exactly v, as Theorem A asserts.

VII. Discussion

A. *Ex Ante and Interim Viewpoints*

Economists distinguish three stages in differential information environments. Ex ante, no one has any private information; in the *interim,* each agent has his private information only; ex post, all information is revealed to all. In our context, ex ante the protagonist knows only the belief system B—which is commonly known by all players; in the interim, each player knows his type, but not those of the others; ex post, each player knows the types of all players.

Some readers may be curious about the relationship of the current work to Aumann (1987), which appears to make the same assumptions—*CKR* and *CP*—but reaches distinctly different conclusions.[20] Namely, the 1987 paper arrives at *unconditional* expectations of correlated equilibria (CEs) of the *given* game G; here, we arrive at *conditional* expectations of CEs of the *doubled* game $2G$.

The puzzle is solved by noting that the 1987 paper concerns the possible ex ante expectations, under which the alternative interim situations are weighted by their probabilities under the common prior. In contrast, this paper characterizes all possible rational expectations at the interim stage.

In practice, the interim viewpoint is the more important. Almost invariably, real players have differential information. What they would have expected had they not had the information that they do have may be of theoretical interest, but has little practical significance. So with all due respect to the 1987 paper, we consider the current one more significant. Needless to say, the 1987 paper was an indispensable stepping stone.

B. *Nash Equilibrium*

A major argument for Nash equilibrium, advanced already by von Neumann and Morgenstern (1944), is that if game theory is to recommend strategies to the players in a game, then those recommendations must constitute a Nash equilibrium. When carefully examined, the argument breaks down. It assumes that the putative "recommendation of game theory" must be for each player to play some specified (mixed or pure) strategy, *known to all players.* But game theory need not make that kind of recommendation; its recommendation could be—indeed, *should* be—"respond optimally to your private information." The players are faced with a game *situation,* not just a game. Even when the *game* is commonly known, the game situation usually is not. It is, indeed, replete with private information, which the players ignore at their peril (see, e.g., Section IIIA).

Nash equilibrium is of central importance in studying *norms* of strategic behavior,[21] and we do not suggest abandoning it as a solution concept for strategic games. But for the one-shot game, rational expectations are more fundamental.

C. *Rational Expectations as a Benchmark*

Any reasonably intelligent child knows that tic-tac-toe "is" a draw—that that is the "right" outcome of the game. That is not to say that in real life she would always play for a draw; she might play for a win, and lose (or win!), even while knowing that the game "is" a draw.

[20] Indeed, many who have been present at preliminary presentations of this material have expressed puzzlement on this matter.

[21] In his thesis, Nash (1950, 21–22) suggested the related concept of "mass action" as an interpretation of his equilibrium.

The formal expression of this is that the value of tic-tac-toe is a draw. Similarly, in any two-person zero-sum game, the value is the "right" outcome. In *n*-person game situations, rational expectations capture this idea.

The building blocks of rational expectations—*CKR* and C*P*—are rarely literally true in real life. Nor are they in any sense normative: in a given situation, they either hold or they do not hold; it makes no sense to say that they "should" hold. Nevertheless, they constitute the appropriate benchmark—characterize the "right outcomes"—just like in tic-tac-toe.

D. *Relation with Muthian Rational Expectations*

In his classic 1961 paper, John Muth characterized expectations as "rational" if they are "essentially the same as the predictions of the relevant economic theory" (316). In games, the "relevant theory," whatever it is, certainly "predicts" simple rationality—utility maximization—on the part of the players. Thus with rational expectations, each player expects all players to be rational. So the "relevant theory" predicts not only simple rationality on the part of all players (*R*), but also *mutual knowledge of rationality* (*KR*), i.e., that all players know[22] all players to be rational. But then all players "expect"—or know—*KR*; so the "relevant theory" predicts not only *KR*, but also K^2R. Continuing in this way, we get K^mR for all *m*, which amounts to *CKR*. So in games, Muth's rational expectations comprise *CKR*.

What about *CP*? One might think of the "relevant theory" as yielding a probability distribution π on the strategies and beliefs of the players—i.e., on their types. If an analyst is then informed of player *i*'s type, his "prediction" will be the conditional of π given that type. With rational expectations, *i* himself, who is indeed informed of his type, will have the same expectations, or beliefs, as the prediction of the informed analyst: the conditional of π given his type. But his beliefs are his theory (in the technical sense of "theory"; see Section I, (ii)(b)). So his theory is the same as the conditional—which is precisely *CP*.

VIII. Background and Literature

The theory presented in this paper differs from much of the strategic game literature in that it does not deal with equilibria; rather, it recommends that a player in a game, like in a one-person decision problem, should simply maximize against her subjective probabilities for the other players' actions. R. Duncan Luce and Howard Raiffa (1957, 306) were among the earliest to consider this approach. They wrote that in a game, "one *modus operandi* for the decision maker is to generate an *a priori* probability distribution over the … pure strategies … of his adversary by taking into account both the strategic aspects of the game and … 'psychological' information … about his adversary, and to choose an act which is best against this … distribution."

Joseph B. Kadane and Patrick D. Larkey (1982) expressed a similar view, but unlike Luce and Raiffa (1957), they eschewed the "strategic aspects." This ignores *the* fundamental insight of game theory, an insight that is captured by the idea of rational expectations introduced here: that a rational player must take into account that the players reason about each other in deciding how to play.

While the theory presented here is new, to a large extent it flows naturally from previous developments in game theory. First was von Neumann's (1928) minimax for two-person zero-sum games; this led to Nash's (1951) strategic equilibrium; this, in turn, to correlated equilibrium; and this, to the theory of rational expectations presented here. The crucial idea here is to analyze

[22] To be sure, this glosses over the difference between "expecting" and "knowing." In any case, the discussion here is meant to be suggestive, not mathematically precise.

game *situations* rather than games—to view games from "inside," without common knowledge of the situation—and it is Harsanyi's (1967-1968) theory of types that enables this.

IX. Conclusion

The fundamental object of study in game theory should be the game *situation* Γ rather than its underlying game G. In game situations, the recommendation of game theory is precisely that of one-person decision theory: choose your strategy to maximize your expected payoff, given your information. But unlike in one-person decisions, rational players should take account of the interactive nature of games—that all the players are rational, and reason about each other. Here, this notion is captured by restricting attention to game situations with *common knowledge of rationality* and *common priors.* We characterize, directly in terms of the underlying game G, the expectations that may result under this restriction; and show that in two-person zero-sum games, the only such expectation is von Neumann's value of the underlying game.

Appendix: Proofs

PROOF OF THEOREM B′:

Let B be an RBS for G with π the common prior, t_i^1 and t_i^2 two types of player i who play the same strategy, and B' the belief system obtained from B by *amalgamating* t_i^1 and t_i^2 into a single type u_i^0. Specifically, in B', the type space of each player j other than i is T_j; the type space U_i of i is obtained from T_i by removing t_i^1 and t_i^2 and replacing them by u_i^0, where

$$s_i(u_i^0) := s_i(t_i^1) = s_i(t_i^2); \tag{1}$$

and the common prior π' in B' is defined by

$$\pi'(u_i, t^{-i}) := \pi(u_i, t^{-i}) \text{ if } u_i \neq u_i^0, \tag{2}$$

and

$$\pi'(u_i^0, t^{-i}) := \pi(t_i^1, t^{-i}) + \pi(t_i^2, t^{-i}). \tag{3}$$

LEMMA 4: *CKR obtains in B′.*

PROOF:

We must show that in B', the strategy of each type maximizes that type's expectation. For types of players j other than i, this is immediate, since their (conditional) expectations are the same in B' as in B, whether or not they play the strategies prescribed by their types. The same holds for types u_i of i other than u_i^0. For i's type u_i^0, one must show that i's conditional expectation

$$\frac{\sum_{t^{-i} \in T^{-i}} \pi'(u_i^0, t^{-i}) h_i(s_i(u_i^0), t^{-i})}{\sum_{t^{-i} \in T^{-i}} \pi'(u_i^0, t^{-i})}$$

if he plays the strategy $s_i(u_i^0)$ prescribed by u_i^0 is at least as great as his conditional expectation

$$\frac{\sum_{t^{-i} \in T^{-i}} \pi'(u_i^0, t^{-i}) h_i(r_i, t^{-i})}{\sum_{t^{-i} \in T^{-i}} \pi'(u_i^0, t^{-i})}$$

if he plays some other strategy r_i, i.e., since the denominators are the same in the two expressions, that

$$(5)\qquad \sum_{t^{-i}\in T^{-i}} \pi'(u_i^0, t^{-i}) h_i(s_i,(u_i^0), t^{-i}) \geqq \sum_{t^{-i}\in T^{-i}} \pi'(u_i^0, t^{-i}) h_i(r_i, t^{-i}).$$

But by the same token, the optimality (in B) of $s_i(t_i^1)$ for t_i^1 and of $s_i(t_i^2)$ for t_i^2 yield

$$(6)\qquad \sum_{t^{-i}\in T^{-i}} \pi(t_i^1, t^{-i}) h_i(s_i,(t_i^1), t^{-i}) \geqq \sum_{t^{-i}\in T^{-i}} \pi(t_i^1, t^{-i}) h_i(r_i, t^{-i}),$$

and

$$(7)\qquad \sum_{t^{-i}\in T^{-i}} \pi(t_i^2, t^{-i}) h_i(s_i(t_i^2), t^{-i}) \geqq \sum_{t^{-i}\in T^{-i}} \pi(t_i^2, t^{-i}) h_i(r_i, t^{-i});$$

and, by (1) and (3), adding (6) and (7) yields (5). This establishes Lemma 4.

COROLLARY 8: *Amalgamation does not affect the expectation of any type of any player, except for the types that have been amalgamated.*

Let α be a rational expectation in G; we must prove that it is a conditional correlated equilibrium payoff in one of the augmented games. By definition of "rational expectation," there is for G an RBS B, and a type u_1 of the protagonist whose expectation is α. Let $r_1 := s_1(u_1)$ be the strategy played by type u_1. Without loss of generality, there is another type—different from u_1 —who plays r_1. For if not, we may split u_1 into two identical types, with the same strategy and theory as u_1. The new common prior (after the split) is then obtained from the original one by halving the probabilities of all states affected by the split.

Repeatedly amalgamating types and using Lemma 4 and its corollary, we arrive at an RBS B'' with a common prior π'', such that (i) for each strategy of each player—other than r_1 —at most one type plays that strategy; (ii) the protagonist has a type u_1^1 that is "like" u_1, in that it plays the same strategy r_1, and has the same expectation α; and (iii) the protagonist has exactly one other type, u_1^2, that plays[23] r_1.

The game G_{2r_1} is exactly like G, except that the strategy r_1 is "doubled": call the duplicates r_1^1 and r_1^2. Each type in B'' corresponds to a strategy in G_{2r_1}; specifically, u_1^1 and u_1^2 correspond to r_1^1 and r_1^2. Hence π'' induces a probability distribution ρ on the strategy profiles in G_{2r_1}, it being understood that strategy profiles without a counterpart in B'' are assigned probability 0.

LEMMA 9: *ρ is a correlated equilibrium in G_{2r_1}.*

PROOF:

CKR in B'' tells us that the expectation of any type is at least as great if it plays the strategy prescribed for that type, as if it plays some other strategy. Rephrasing, any strategy in G_{2r_1} with positive ρ-probability yields to its player a conditional expected payoff under ρ at least as great as any other strategy.

By Lemma 9 and Corollary 8, α is a conditional correlated equilibrium payoff in G_{2r_1}. This completes the proof of Theorem B′ in one direction.

[23] Obtained by amalgamating all the types of the protagonist who play r_1 other than u_1^1.

In the other direction, let β be a conditional correlated equilibrium payoff in an augmented game G_{2r_1}, specifically, the conditional payoff to the correlated equilibrium ρ, given some strategy s_1 in G_{2r_1}. Define a belief system B for G as follows: the types of i in B are in one-one correspondence with those of his strategies in G_{2r_1} to which ρ assigns positive probability; the theory of a type is the conditional of ρ given the strategy corresponding to that type. Then ρ is a common prior for B, and that CKR holds in B is the same as saying that ρ is a correlated equilibrium in G_{2r_1}. This completes the proof of Theorem B′.

PROOF OF THEOREM B:

If α is a rational expectation in G, then by Theorem B′, it is a conditional payoff to a correlated equilibrium ρ' in some augmented game G_{2r_1}, given a strategy x' of the protagonist in that game. The strategy profiles in G_{2r_1} are in one-one correspondence with those strategy profiles in $2G$ whose first component either is *either one* of the two copies of r_1 in $2G$, or is the *first* copy of a strategy other than r_1. Assigning to any such profile the ρ'-probability of the corresponding profile in G_{2r_1}, and 0 to all other profiles, yields a correlated equilibrium ρ in $2G$. The strategy x' in G_{2r_1} corresponds to some strategy x in $2G$, and then α is the conditional payoff to ρ in $2G$ given x. This completes the proof of Theorem B in one direction.

In the other direction, let β be a conditional payoff to a correlated equilibrium ρ in $2G$, given a strategy x of the protagonist; let r_1 be the strategy in G of which x is a copy in $2G$. "Amalgamating" the two copies in $2G$ of all strategies other than r_1 yields a correlated equilibrium ρ' in G_{2r_1}, and turns x into a strategy x' in G_{2r_1}. Then β is a conditional payoff to ρ' in G_{2r_1}, given x'. So by Theorem B′, β is a rational expectation in G. This completes the proof of Theorem B.

PROOF OF PROPOSITION C:

(i) The first part follows from Theorem B, as every correlated equilibrium in G can also be viewed as a correlated equilibrium in $2G$, since one can simply assign probability 0 to one of the two duplicates of each of the protagonist's strategies. The part about Nash equilibrium follows from the fact that at a Nash equilibrium ν, all "active" strategies of the protagonist (indeed, of any player)—i.e., those with positive probability at ν —must get the same payoff, so that the conditional expected payoffs coincide with the total expected payoff.

(ii) Follows from the fact that a strictly dominated strategy can never appear with positive probability in a correlated equilibrium, as it is always worthwhile to switch to the dominating strategy.

(iii) Let α be a rational expectation, t_1^* a type of the protagonist with expectation α in an RBS, $s_1^* := s_1^*(t_1^*)$ the strategy played by type t_1^*, and p the probability distribution over $S_{-1} := S_2 \times \cdots \times S_n$ that t_1^*'s theory induces. Thus $\alpha = \sum_{s_{-1} \in S_{-1}} p_{s_{-1}} h_1(s_1^*, s_{-1})$. By CKR, s_1^* maximizes t_1^*'s expectation given its theory, so by the minimax theorem,

$$\alpha = \max_{s_1 \in S_1} \sum_{s_{-1} \in S_{-1}} p_{s_{-1}} h_1(s_1, s_{-1}) \geq \min_{q} \max_{s_1 \in S_1} \sum_{s_{-1} \in S_{-1}} q_{s_{-1}} h_1(s_1, s_{-1}) = \max_{r} \min_{s_{-1} \in S_{-1}} \sum_{s_1 \in S_1} r_{s_1} h_1(s_1, s_{-1}),$$

where q and r range over all probability distributions over S_{-1} and S_1, respectively.

(iv) Follows from Theorem B, since correlated equilibria are covariant in the required manner.

PROOF OF PROPOSITION E:

Proposition C(iii) says that every rational expectation is $\geq v$. Suppose α is a rational expectation that is $> v$. From Theorem B it follows that α is a conditional payoff of a correlated

equilibrium ρ in the doubled game $2G$. By Proposition C(iii) and Theorem B, all other conditional payoffs to ρ in $2G$ are $\geq v$. Since the unconditional payoff to ρ is the expectation of the conditional payoffs, and α appears in this expectation with positive probability, it follows that the unconditional payoff to ρ is $> v$. But an unconditional payoff to the correlated equilibrium ρ in $2G$ is also an unconditional payoff to a correlated equilibrium in G, obtained by amalgamating duplicated strategies. So we get a correlated equilibrium payoff in G that is $> v$, contrary to hypothesis.

PROOF OF THEOREM A:

Follows from Proposition E, since in two-person zero-sum games, the (unconditional) expected payoff to every correlated equilibrium is the value.[24]

PROOF OF PROPOSITION D:

By definition, the given game G has a correlated equilibrium μ that assigns positive probability to each strategy of each player, and in which the associated inequalities are strict. Let S be the set of strategy profiles in G. If ι assigns equal probabilities to all strategy profiles, and $\varepsilon > 0$ is sufficiently small, then $\lambda := (1 - \varepsilon)\mu + \varepsilon\iota$ assigns positive probability to each strategy profile, and the associated inequalities are still strict. Let w be a strategy profile in G that yields the protagonist her highest payoff in G. For each strategy profile s in G, let s^1 and s^2 be the two copies of s in $2G$. Let $0 < \delta < \lambda_w$. Define a probability distribution ρ on the set $2S$ of strategy profiles in $2G$ by $\rho_{s^1} := \lambda_s$, $\rho_{s^2} := 0$ for $s \neq w$, and $\rho_{w^1} := \lambda_w - \delta$, $\rho_{w^2} := \delta$. We will show that ρ is a correlated equilibrium of $2G$ when δ is sufficiently small.

Indeed, the inequalities associated with ρ in $2G$ are the same as those associated with λ in G, except for those that correspond to w_1^1 and w_1^2 in $2G$. Since δ is small, and the inequalities corresponding to w_1 in G are strict, those corresponding to w_1^1 in $2G$ still hold. As for w_1^2, if the protagonist is informed of w_1^2, she knows for sure that she will get the highest possible payoff in the whole game if she indeed plays w_1^2, so it certainly is not worthwhile for her to switch. Therefore, ρ is indeed a correlated equilibrium of $2G$.

It then follows from Theorem B that the conditional payoff corresponding to w_1^2 is a rational expectation. This conditional payoff is the protagonist's payoff in G at w, which is her highest payoff at any strategy pair.

REFERENCES

Aumann, Robert J. 1974. "Subjectivity and Correlation in Randomized Strategies." *Journal of Mathematical Economics*, 1(1): 67–96.

Aumann, Robert J. 1987. "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica*, 55(1): 1–18.

Aumann, Robert J., and Aviad Heifetz. 2002. "Incomplete Information." In *Handbook of Game Theory with economic applications, Volume 3,* ed. Robert J. Aumann and Sergiu Hart, 1665–86. Amsterdam: Elsevier.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, 3(4): 367–88.

Harsanyi, John C. 1967–1968. "Games of Incomplete Information Played by Bayesian Players I, II, III." *Management Science*, 14(3): 159–82, 320–34, 486–502.

Kadane, Joseph B., and Patrick D. Larkey. 1982. "Subjective Probability and the Theory of Games." *Management Science*, 28(2): 113–20.

Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions.* New York: John Wiley.

[24] Aumann (1974), last paragraph of Section 2.

Mertens, Jean-Francois, and Shmuel Zamir. 1985. "Formulation of Bayesian Analysis for Games with Incomplete Information." *International Journal of Game Theory*, 14(1): 1–29.

Minelli, Enrico. 1995. *Rational Expectations in Games.* Unpublished.

Muth, John F. 1961. "Rational Expectations and the Theory of Price Movements." *Econometrica*, 29(6): 315–35.

Myerson, Roger B. 1997. "Dual Reduction and Elementary Games." *Games and Economic Behavior*, 21(1-2): 183–202.

Nash, John F. 1950. "Non-cooperative Games." PhD diss. Princeton University. Repr. in *The Essential John Nash*, ed. Harold W. Kuhn and Sylvia Nasar, 53–84. Princeton: Princeton University Press, 2002.

Nash, John F. 1951. "Non-cooperative Games." *Annals of Mathematics*, 54(2): 286–95.

von Neumann, John. 1928. "Zur Theorie der Gesellschaftsspiele." *Mathematische Annalen*, 100: 295–320.

von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior.* Princeton: Princeton University Press.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. 1991. "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *American Economic Review*, 81(5): 1068–95.

Shapley, Lloyd S. 1964. "Some Topics in Two-Person Games." In *Advances in Game Theory*, ed. Melvin Dresher, Lloyd S. Shapley, and Albert W. Tucker, 1–28. Princeton: Princeton University Press.