Correlated equilibria, good and bad: an experimental study

John Duffy
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260, USA
jduffy@pitt.edu

Nick Feltovich*
University of Aberdeen Business School
Edward Wright Building
Aberdeen AB24 3QY, UK
n.feltovich@abdn.ac.uk

October 23, 2008

Abstract

We report results from an experiment that explores the empirical validity of *correlated equilibrium*, an important generalization of the Nash equilibrium concept. Specifically, we examine the conditions under which subjects playing the game of Chicken will condition their behavior on private third–party recommendations drawn from publicly–announced distributions. We find that when recommendations are not given to subjects, aggregate behavior is characterized well by mixed–strategy Nash equilibrium play, though it does less well at lower levels of aggregation. When recommendations are given, behavior differs from both mixed–strategy Nash equilibrium and behavior without recommendations, with the nature of the differences varying according to the treatment. Our main finding is that subjects will follow third–party recommendations only if those recommendations derive from a correlated equilibrium, and further, if that correlated equilibrium is payoff–enhancing relative to the available Nash equilibria.

Journal of Economic Literature classifications: D83, C72, C73.

Keywords: correlated equilibrium, recommendation, coordination problem, chicken game, hawk–dove game, experiment, experimenter demand effects.

*Corresponding author. We thank Yasuyo Hamaguchi, Hans Hvide, Sobei H. Oda, Joe Swierzbinski, Peyton Young, three anonymous referees, and participants at several conferences and seminars for helpful suggestions and comments. Any remaining errors would have been less likely if we had followed third–party recommendations more often.

1 Introduction

A standard assumption in noncooperative game theory is that players' strategies—whether pure or mixed—are probabilistically independent. However, researchers at least as long ago as Aumann (1974, 1987) recognized that relaxing this assumption by allowing *correlation* in players' strategies could greatly enlarge a game's equilibrium possibilities beyond the set of Nash equilibria. The equilibria that result are known as *correlated equilibria*.[1] As an illustration, consider the two–player game of Chicken, shown in Figure 1; strategies are defect (D) and cooperate (C). This game has two asymmetric pure–strategy Nash equilibria—(D,C) and (C,D)—as well as a mixed–strategy Nash equilibrium in which each player chooses D with probability two–fifths.

		Player 2	
		D	C
Player	D	0,0	9,3
1	C	3,9	7,7

Figure 1: The basic Chicken game

The mixed–strategy equilibrium of this game has the attractive feature of symmetry—thus avoiding the "symmetry–breaking" question implicit in asymmetric equilibria (see Crawford (1998)). Evolutionary dynamics often favor such symmetry and indeed, the Nash equilibrium mixed strategy is the unique evolutionarily stable strategy of this game (see, for example, Hofbauer and Sigmund (1998)). However, as Skyrms (1996) and others have observed, this mixed–strategy equilibrium is inefficient: in the Chicken game of Figure 1, it yields expected payoffs of just 5.4 for each player. By contrast, if the players somehow agreed to condition their behavior on a fair coin toss, playing (for example) the strategy profile (D,C) after Heads and (C,D) after Tails, each could improve her ex ante expected payoff to 6. Moreover, since both recommended outcomes are strict Nash equilibria, both would strictly prefer to honor such an agreement as long as they believed that the other would, even after knowing which recommendation was received.[2]

Furthermore, as Aumann (1974) first pointed out, the players could actually do even better in this game by enlisting an "objective chance mechanism", that randomly chooses one of three signals called (for example) "X", "Y", and "Z", with equal probability. Player 1 learns only whether X was chosen or not while Player 2 learns only whether Z was chosen or not. Aumann then shows that if both players know the set and distribution of signals (possible states of the world), and if Player 1 plays strategy D if the state is "X" and C otherwise, while Player 2 plays strategy D if the state is "Z" and C otherwise, that these correlated strategies are mutual best responses, i.e., a *correlated equilibrium* that yields expected payoffs of $6\frac{1}{3}$ to each player in the Chicken game of Figure 1—an expected payoff that is higher than that obtained under the mixed–strategy equilibrium.[3]

[1] An early example of a correlated equilibrium is also found in Luce and Raiffa (1957, pp. 115-120).

[2] By contrast, see Young (2005, Chapter 3) for a discussion of a variant on this setup (due to Moulin and Vial (1978)) in which players choose—*before* receiving recommendations—whether to commit to following them or not. Young calls a distribution of recommendations under this setup a *coarse* correlated equilibrium if all players are willing to commit to following recommendations, given that the others also choose to commit.

[3] The objective chance mechanism induces the outcomes (D,C), (C,C) and (C,D) with equal (one–third) probability, so expected payoffs are $(3 + 7 + 9)/3$ for each player.

A correlated equilibrium is a probability distribution over outcomes—that is, a joint distribution over players' strategies—such that under the assumptions mentioned above, all players prefer to follow their state–contingent correlated strategy. Then, a Nash equilibrium is just a special case of correlated equilibrium, in which the joint distribution of strategies is the product of the corresponding marginals (that is, the resulting players' strategies are probabilistically independent of one another).

Aumann (1987) argued that correlated equilibria follow naturally from a "modern subjectivist, Bayesian view of the world" (p. 2)—that is, when all events can be assigned subjective probabilities and individuals are Bayesian–rational. Indeed, he shows that if players are Bayesian–rational and hold common priors concerning the probability distribution of observations from the randomization device, then the distribution of actions chosen by those players *must* be a correlated equilibrium distribution. As Aumann observes, while correlated equilibria and mixed equilibria both rely on observations from a randomization device, correlated strategies (and thus correlated equilibria) are *more general* as there is no need to assume that the observations from the randomization device are independent of one another, as is assumed under mixed strategies.

On the other hand, Gul (1988) has argued that Aumann's argument for the naturalness of correlated equilibria relies heavily on the assumption of common prior beliefs, which is not so easily justified. Gul argues instead that common priors should be explicitly modeled as having been achieved based on some prior stage of the game. One possibility is that players have learned over time to hold such common beliefs as in the work of Hart and Mas-Colell (2002) and others.[4] A second possibility is to adopt Myerson's (1991, p. 250) mechanism–design approach where, in a first stage, a neutral third-party "mediator" (which Myerson describes as "a person or machine that can help the players communicate and share information") draws outcomes for all players from a *commonly known distribution*, thus ensuring common prior beliefs. For instance, the mediator might announce to players that he will draw each of the three Chicken game outcomes (D,C), (C,C), and (C,D) with equal probability. The mediator then recommends to each player only the player's *own* strategy for the outcome chosen—*not* that of the other player (e.g., if the outcome randomly drawn is (C,D), the mediator privately recommends to Player 1 that she play C and privately recommends to Player 2 that he play D). In the second stage, players may choose actions conditional on the recommendation given to them by the mediator. This latter approach is perhaps the one that is best suited to the laboratory, as the experimenter can announce the distribution of outcomes used publicly thereby assuring common priors, and the experimenter can also play the role of the neutral, third–party mediator. This recommended–play approach has the added advantage of yielding a clearer mapping from realizations of the randomization device to each player's strategy space. This is the approach we take in this paper.[5]

The purpose of this paper is to examine the empirical validity of the correlated equilibrium concept with an external mediator. We study correlated equilibria in the controlled environment of the laboratory, as this enables us to clearly assess the role of well–defined, correlated signals as coordinating devices, providing the theory with its best chance of success. Specifically, we design and conduct an experiment in which human subjects play the game shown in Figure 1. Prior to making their choices, subjects receive

[4] See, e.g., Foster and Vohra (1997), Fudenberg and Levine (1998 Chapter 8, 1999), Vanderschraaf (2001), Vanderschraaf and Skyrms (2003), and Brandenburger and Friedenberg (2008).

[5] Sharma and Torres (2004) provide a model of how such a neutral, third-party mediator could play a welfare–improving role in implementing correlated equilibria in a team production model.

private signals ("recommendations") generated according to a known distribution of outcomes that serves as our main treatment variable. Three of the distributions we use are symmetric correlated equilibria. In one of our treatments, which we call our "Nash–recommendations" treatment, the correlated equilibrium we attempt to implement is simply a convex combination of Nash equilibria. In a second treatment—our "good–recommendations" treatment—the correlated equilibrium is the one described above, which yields payoffs that are Pareto superior to all symmetric payoff vectors in the convex hull of Nash equilibrium payoff vectors.

It is often forgotten, though, that there also exist correlated equilibria in which payoffs are *Pareto inferior* to all symmetric payoff vectors in the convex hull of Nash equilibrium payoff vectors. If correlated equilibrium is to be taken seriously as a descriptive device, and not just a theoretical curiosity, then it should be possible to induce these bad correlated equilibria as well as the good ones. To our knowledge, however, there has never been an experimental test of a bad correlated equilibrium. We remedy this, with what we call our "bad–recommendations" treatment. Despite the "bad" moniker, the distribution over outcomes we use in this treatment is every bit as much a correlated equilibrium as that in our good– and Nash–recommendations treatments. In particular, it is still optimal for a player to follow her recommendations, as long as she believes her opponent will follow the recommendations given him.

Finally, we attempt to distinguish between subjects' following recommendations as part of a correlated equilibrium and their following of recommendations for other reasons—for example, out of a desire to please the experimenter (an example of "experimenter demand effects")—with our "very–good–recommendations" treatment. In this treatment, the distribution of recommended outcomes is *not* a correlated equilibrium, but the temptation to follow recommendations may be great, because if both players follow recommendations, payoffs are Pareto superior to all symmetric correlated–equilibrium payoff vectors.

In the experiment, subjects play the game shown in Figure 1 repeatedly against changing opponents. In half of the rounds, they receive recommendations (always according to the same correlated strategy distribution), while in the remaining rounds, they do not receive any recommendations. The main results are as follows. When players do not receive recommendations, their behavior is described fairly well (though not perfectly) by the mixed–strategy Nash equilibrium. Giving subjects recommendations has an effect that depends on which underlying distribution of outcomes is used. The likelihood of following a recommendation is higher in the good– and Nash–recommendations treatments and lower in the bad– and very–good–recommendations treatments, and also varies somewhat with which of the available actions is recommended. In nearly all cases, subjects follow recommendations more often than chance would predict, but there is no treatment where subjects follow recommendations all the time.

2 Correlated equilibrium—theory and tests

The game we use is the Chicken game shown in Figure 1 above. We chose Chicken as it is perhaps the simplest game with the property that there exist correlated equilibrium payoff pairs that lie outside the convex hull of Nash equilibrium payoff pairs. Under the assumption that players are risk neutral with regard to monetary payoffs, the game has three Nash equilibria: (D,C), (C,D), and a mixed–strategy Nash equilibrium in which each player chooses D with probability $\frac{2}{5}$. Payoffs in these three equilibria are, respectively, (9,3), (3,9), and (5.4,5.4).

As mentioned in the introduction, one way to think about correlated equilibria is as involving a "mediator"—a non–strategic third party—in the game. The mediator chooses one of the four pure–strategy profiles according to a commonly–known probability distribution, and to each player "recommends" that player's component in the profile. (The mediator never recommends a mixed strategy.) The probability distribution is a correlated equilibrium of the original game if each player at least weakly prefers following her recommended action to choosing any other action. (Thus, a correlated equilibrium of the original game corresponds to a Nash equilibrium of this new game, in which players' strategies are mappings from recommended actions to chosen actions.[6])

Define p_{DD}, p_{DC}, p_{CD}, and p_{CC} to be the probabilities of the outcomes (D,D), (D,C), (C,D), and (C,C), according to the commonly–known distribution characterizing the mediator's behavior. Suppose Player 1 is given a recommendation of D. Then, the conditional probability that the chosen outcome was (D,D) is $\frac{p_{DD}}{p_{DD}+p_{DC}}$, and the probability that the chosen outcome was (D,C) is $\frac{p_{DC}}{p_{DD}+p_{DC}}$. If Player 1 believes that Player 2 will follow the recommendation given to him, then Player 1's conditional expected payoff from following her recommendation of D is

$$\frac{p_{DD}}{p_{DD}+p_{DC}} \cdot 0 + \frac{p_{DC}}{p_{DD}+p_{DC}} \cdot 9 = \frac{9p_{DC}}{p_{DD}+p_{DC}},$$

and her conditional expected payoff from choosing C instead is

$$\frac{p_{DD}}{p_{DD}+p_{DC}} \cdot 3 + \frac{p_{DC}}{p_{DD}+p_{DC}} \cdot 7 = \frac{3p_{DD}+7p_{DC}}{p_{DD}+p_{DC}},$$

so (if risk neutral) she prefers to follow the D recommendation if $\frac{9p_{DC}}{p_{DD}+p_{DC}} \geq \frac{3p_{DD}+7p_{DC}}{p_{DD}+p_{DC}}$—that is, if $2p_{DC} \geq 3p_{DD}$. Using similar reasoning for Player 1 following a C recommendation, Player 2 following an D recommendation, and Player 2 following a C recommendation gives us a total of four inequalities:

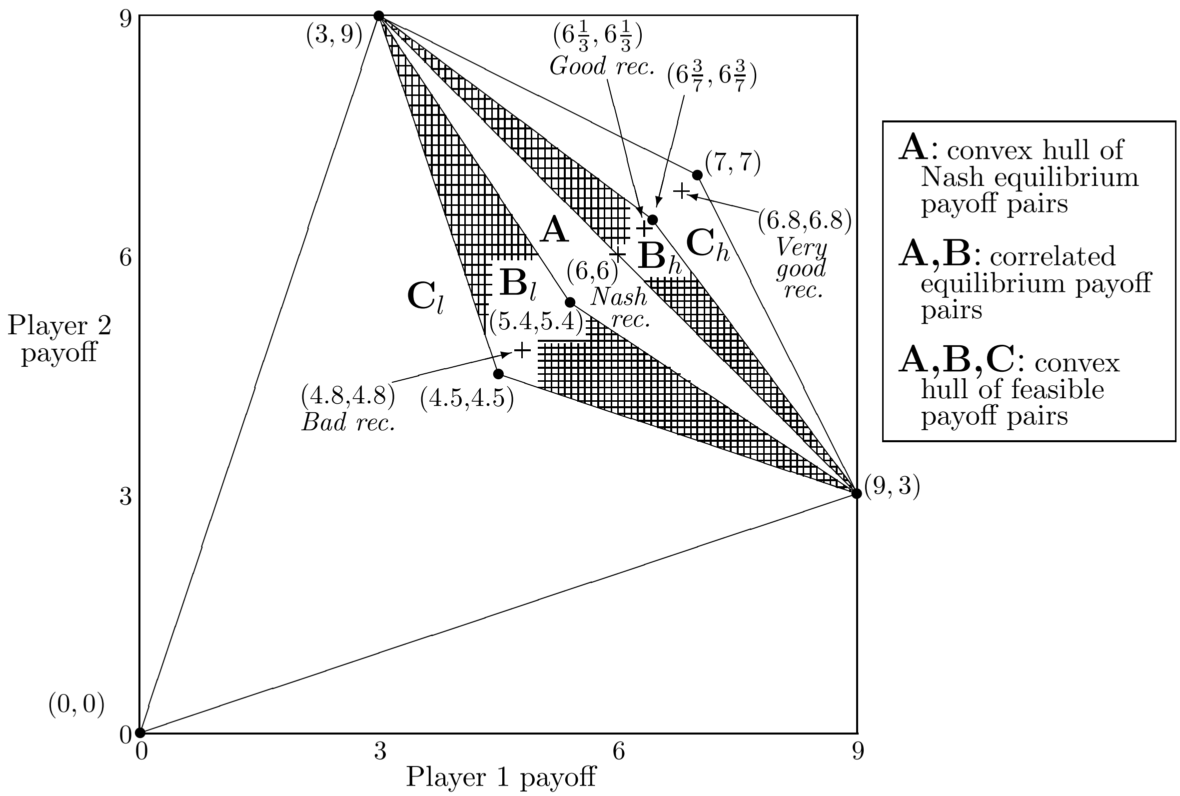
$$\begin{aligned} 2p_{DC} &\geq 3p_{DD} \\ 3p_{CD} &\geq 3p_{CC} \\ 2p_{CD} &\geq 3p_{DD} \\ 3p_{DC} &\geq 3p_{CC}. \end{aligned}$$

A correlated equilibrium is a quadruple $(p_{DD}, p_{DC}, p_{CD}, p_{CC})$ that satisfies these four inequalities, along with $p_{DD} + p_{DC} + p_{CD} + p_{CC} = 1$.

Since the set of correlated equilibria can be characterized as an intersection of sets defined by linear equations and inequalities, it is a convex set, and because it contains the set of Nash equilibria, it must also contain the convex hull of Nash equilibria. The same is true in payoff space; that is, the set of correlated–equilibrium payoffs of a game always contains the convex hull of the set of Nash equilibrium payoff pairs. However, in most games—including ours—there also exist correlated equilibria that are *not* in the convex hull of Nash equilibrium payoff pairs. Figure 2 shows the regions corresponding to the sets of Nash equilibrium payoff pairs and correlated equilibrium payoff pairs. The Nash equilibrium payoff pairs of this game are (3,9) (corresponding to the equilibrium (C,D)), (9,3) (corresponding to (D,C)), and (5.4,5.4) (corresponding to the mixed–strategy equilibrium). Therefore, the convex hull of Nash equilibrium payoff

[6]This Nash equilibrium is not unique. There always exist three "babbling" equilibria corresponding to the three Nash equilibria of the original game, in which both players completely ignore the recommendations given them, and play Nash equilibrium strategies instead. There exist additional equilibria as well.

Figure 2: Characteristics of the Game



pairs is the triangle with these three points as vertices (region A in the figure); in particular, 6 is the highest symmetric payoff in this convex hull, and 5.4 the lowest. The set of correlated equilibrium payoff pairs is the quadrilateral with vertices (3,9), (4.5,4.5), (9,3), and $(6\frac{3}{7}, 6\frac{3}{7})$ (the union of regions A, B_l, and B_h in the figure), so that $6\frac{3}{7}$ is the highest symmetric correlated equilibrium payoff and 4.5 the lowest.

Relatively little experimental research has looked at correlated equilibria that are not convex combinations of Nash equilibria.[7] The earliest such study that we know of is that by Moreno and Wooders (1998), who examine the ability of several game–theoretic solution concepts (including Nash equilibrium and correlated equilibrium) to characterize subject behavior in a three–player version of a one–shot matching pennies game, in which two of the players have perfectly aligned interests; their game is shown on the left of Figure 3. Instead of giving players recommendations as we do, they allowed subjects to participate in a round of cheap talk prior to play of the game; subjects could send messages to either other player individually, or to both at once. Moreno and Wooders found that the choices of the players with aligned interests were highly correlated, so that mixed–strategy Nash equilibrium poorly described the distribution of outcomes. Rather, they concluded that the best–performing solution concept was coalition–proof correlated equilibrium (Einy and Peleg (1995), Moreno and Wooders (1996)).

More recently, Cason and Sharma (2007) attempted to induce a correlated equilibrium through the

[7] Experimental studies of correlated equilibria that are convex combinations of Nash equilibrium include Van Huyck, Gilette, and Battalio (1992), Brandts and McLeod (1995), and Seely, Van Huyck, and Battalio (2005). In a market setting, Duffy and Fisher (2005) examine whether subjects will coordinate on the closely related concept of a sunspot equilibrium involving a randomization over two certainty equilibria.

use of private recommendations to subjects, as we do. The game they use is a version of Chicken, shown on the right of Figure 3. The correlated equilibrium they attempt to induce has (Up, Right) and (Down,

Moreno and Wooders (1998) — Player 3: H

		Player 2: H	Player 2: T
Player 1	H	1,1,–2	–1,–1,2
	T	–1,–1,2	–1,–1,2

Moreno and Wooders (1998) — Player 3: T

		Player 2: H	Player 2: T
Player 1	H	–1,–1,2	–1,–1,2
	T	–1,–1,2	1,1,–2

Cason and Sharma (2007)

		Player 2: Left	Player 2: Right
Player 1	Up	3,3	48,9
	Down	9,48	39,39

Figure 3: Games used in previous correlated–equilibrium experiments

Left) occurring with probability 0.375 each, and (Down, Right) with probability 0.25, with (Up, Left) never occurring. This correlated equilibrium yields expected payoffs of 31.125 for each player: higher than the mixed–strategy Nash equilibrium expected payoffs of 20.4, and indeed, higher than any symmetric payoff pair in the convex hull of Nash equilibrium expected payoffs. In the experiment, subjects often did follow recommendations, doing so roughly 80% of the time in their baseline treatment, and earning payoffs well above the mixed–strategy Nash equilibrium prediction (though below the prediction of the correlated equilibrium) as a result.[8]

However, by only considering a correlated equilibrium that was *payoff–enhancing* relative to Nash equilibrium, Cason and Sharma's study risks confounding the coordinating role of third-party recommendations with a general interest by subjects in earning higher payoffs. Further, in Cason and Sharma's experimental instructions, they explicitly tell subjects that they ought to follow recommendations, as doing so will result in higher payoffs as long as the opposing player also follows recommendations.[9]

By contrast with Cason and Sharma's (2007) experiment, which considered a single type of correlated equilibrium, our experimental design considers three different correlated equilibria, each associated with a different probability distribution for recommended play. In our "Nash–recommendations" treatment, the recommendations (D, C) and (C, D) are each selected with probability one–half, and (C, C) and (D, D) are selected with probability zero. This distribution of recommended outcomes is a correlated equilibrium, and moreover, is a convex combination of Nash equilibria, with payoffs of 6 for each player. We also consider two correlated equilibria that are not convex combinations of Nash equilibria. In our "good–recommendations" treatment, the recommended outcomes (D, C), (C, D), and (C, C) are each selected with probability one–third, and (D, D) is selected with probability zero. These probabilities satisfy the conditions for a correlated equilibrium, and yield payoffs of $6\frac{1}{3}$ for each player—more than any point in the convex hull of Nash equilibrium payoff pairs. In addition to the good–recommendations treatment, however,

[8]Cason and Sharma—somewhat pessimistically, in our opinion—conclude from these results that "players frequently reject recommendations," and their experiment includes additional treatments designed to increase the likelihood that recommendations are followed, such as having human subjects play against a computer program that always follows recommendations. Recommendations are typically followed even more often in these variations.

[9]For example, their instructions state "[y]ou should follow the recommendation given by the computer, because as long as the person you are paired with also follows his or her recommendation then you earn more on average by following the recommendation" and "[t]o reiterate: you always earn more by following your recommendation as long as the participant you are paired with also follows his or her recommendation"; see http://www.krannert.purdue.edu/faculty/cason/papers/corr-inst.pdf.

we also consider a "bad–recommendations" treatment, in which the recommended outcomes (D, C) and (C, D) are each selected with probability 0.4, and (D, D) is selected with probability 0.2, so that (C, C) is selected with probability zero. These probabilities also satisfy the conditions for a correlated equilibrium, but result in payoffs of only 4.8 for each player—less than any point in the convex hull of Nash equilibrium payoff pairs. As far as we know, there are no existing experimental studies of correlated equilibria that are payoff–reducing relative to Nash equilibrium.

Finally, as an even stronger test of the correlated equilibrium concept, we consider one distribution of recommended outcomes that is *not* a correlated equilibrium. In our "very–good–recommendations" treatment, the recommended outcome (C, C) is selected with probability 0.8, (D, C) and (C, D) are each selected with probability 0.1, and (D, D) is selected with probability zero. Given these probabilities, a player receiving a D recommendation will prefer to follow it—assuming she believes her opponent will also follow recommendations—but a player receiving a C recommendation will not, instead preferring to choose D. If recommendations are followed, however, payoffs are 6.8 for each player—higher than in any of three correlated equilibria discussed above.

Some features of the four recommended outcome distributions we use, as well as the mixed–strategy Nash equilibrium, are shown in Table 1. The expected payoffs from following these distributions of recommended outcomes are also shown in Figure 2 (as plus signs).

Table 1: Outcome frequencies imposed in the experiment

	Probability of (D,D) outcome	Probability of (D,C) outcome	Probability of (C,D) outcome	Probability of (C,C) outcome	Probability of C choice	Expected payoffs
Good recommendations	0.000	0.333	0.333	0.333	0.667	(6.333,6.333)
Bad recommendations	0.200	0.400	0.400	0.000	0.400	(4.8,4.8)
Nash recommendations	0.000	0.500	0.500	0.000	0.500	(6,6)
Very good recommendations	0.000	0.100	0.100	0.800	0.900	(6.8,6.8)
Mixed–strategy NE	0.160	0.240	0.240	0.360	0.600	(5.4,5.4)

3 Experimental procedures

Besides varying the type of recommendations that were given to subjects (that is, the probability distribution over outcomes), we varied whether recommendations were given at all. All experimental sessions lasted for 40 rounds: 20 rounds with recommendations and 20 rounds without recommendations. We also varied the order of these; in half of our sessions, the 20 rounds without recommendations came first, and in the other half, the 20 rounds with recommendations came first. (Thus, whether or not recommendations were given was varied within–subject, while the type of recommendations and the ordering between recommendations and no recommendations were varied between–subjects.) Each experimental session involved 12 subjects. Subjects were primarily undergraduate students from University of Pittsburgh, and were recruited by newspaper advertisements and email. No one took part in more than one session of this experiment.

At the beginning of a session, subjects were seated in a single room and given a set of written instructions for the first twenty rounds.[10] They were told at this time that there would be a second part to the session, but details of the second part were not announced until after the first part had ended. The instructions for the first part were read aloud to subjects, in an attempt to make the rules of the game common knowledge. After the instructions were read, subjects were given a short quiz to ensure that they understood the instructions. After subjects' quizzes were completed, they were graded anonymously. If any question was answered incorrectly, the experimenter went over the question and answer out loud for the benefit of all subjects (without identifying which subject had answered incorrectly). After any incorrect answers were discussed, the first round of play began. After the twentieth round of play was completed, each subject was given a copy of the instructions for the remaining twenty rounds. These were also read aloud, after which another (shorter) quiz was given out, before the final twenty rounds were played.

In the instructions, we strove to use neutral terminology. Instead of relatively loaded terms such as "opponent" or "partner", we used phrases such as "the player matched with you". Also, in our discussion of recommendations, we never went so far as to instruct subjects to follow recommendations, or even to point out that following recommendations might lead to higher payoffs; rather, we merely provided the outcome distribution from which the recommendations were generated (both in the written instructions and in our public reading of those instructions), and noted that a player's recommendation may or may not convey information about the recommendation given to the player matched with him.[11]

The experiment was run on networked computers, using the z–Tree experiment software package (Fischbacher (2007)). Subjects were asked not to communicate directly with one another, so the only interactions were via the computer program. Subjects were paired using a round–robin matching format, in an attempt to minimize incentives for reputation building and other potential supergame effects; for the same reason, subjects were not given identifying information about their opponents in any round.

A round of the game in which there were *no* recommendations (either rounds 1–20 or rounds 21–40, depending on the cell) began by prompting subjects to choose one of the two available actions. (In the instructions and during the session, the actions were named X and Y instead of D and C, respectively.) After the action choices were entered, each subject was shown the following information: own action, opponent action, own payoff, and opponent payoff. In a round of the game *with* recommendations, the sequence of play was similar except for the recommendations. Specifically, subjects would first be shown their "recommended action", which was randomly drawn from the appropriate outcome distribution. Then, they were prompted to choose an action. After action choices were entered, each subject was shown the following information: own recommendation, own action, opponent recommendation, opponent action, own payoff, and opponent payoff. In all treatments, subjects were not given information about the results of any other pairs of subjects, either individually or in aggregate. At the end of the round, subjects were

[10]The set of instructions given to subjects—as well as additional materials given to them (quizzes and record sheets)—from one of our cells can be found at http://www.abdn.ac.uk/~pec214/papers/corr_instructions.pdf. Materials used in the other cells and screenshots of the computer interface seen by subjects, as well as the raw data from the experiment, are available from the corresponding author upon request.

[11]One passage from our instructions states, "These recommendations are optional; it is up to you whether or not to follow them. Notice that your recommendation may give you information about the recommendation that was given to the person matched to you." To further emphasize this point, one of the questions in the quiz given to subjects after reading the instructions was, "You are required to follow the recommendations shown on your computer screen (circle one): TRUE FALSE"—to which the correct answer was FALSE. We acknowledge the possibility that our use of the term "recommendations" itself might have influenced subjects to follow them to some extent.

asked to observe their result, write the information from that round down onto a record sheet, and then click a button to continue to the next round.

At the end of round 40 of any treatment, the experimental session ended. One of the first twenty rounds and one of the last twenty rounds were randomly chosen, and each subject received his/her earnings from these two rounds, at an exchange rate of $1 per point. Additionally, all subjects received a $5 show–up fee. Total earnings for subjects participating in a session averaged about $15, and sessions typically lasted between 45 and 60 minutes.

4 Experimental results

A total of 16 sessions were conducted—four of each treatment—with 12 subjects per session, for a total of 192 subjects. Each subject played 20 rounds without recommendations and 20 with recommendations, giving us 7580 observations overall: 1920 of each treatment. Aggregate outcome frequencies and payoffs are shown in Table 2.

Table 2: Aggregate observed outcome frequencies

Recommendations	Ordering	Outcome				Average
		(D,D)	(D,C)	(C,D)	(C,C)	payoff
None	+	0.145	0.200	0.274	0.381	5.513
	−	0.174	0.226	0.295	0.305	5.261
	Combined	0.159	0.213	0.284	0.343	5.387
Good	+	0.154	0.258	0.363	0.225	5.300
	−	0.125	0.258	0.279	0.338	5.588
	Combined	0.140	0.258	0.321	0.281	5.444
Bad	+	0.213	0.267	0.246	0.275	5.000
	−	0.171	0.242	0.208	0.379	5.354
	Combined	0.192	0.254	0.227	0.327	5.177
Nash	+	0.121	0.271	0.300	0.308	5.583
	−	0.104	0.258	0.300	0.338	5.713
	Combined	0.113	0.265	0.300	0.323	5.648
Very good	+	0.138	0.175	0.254	0.433	5.608
	−	0.200	0.279	0.229	0.292	5.092
	Combined	0.169	0.227	0.242	0.363	5.350

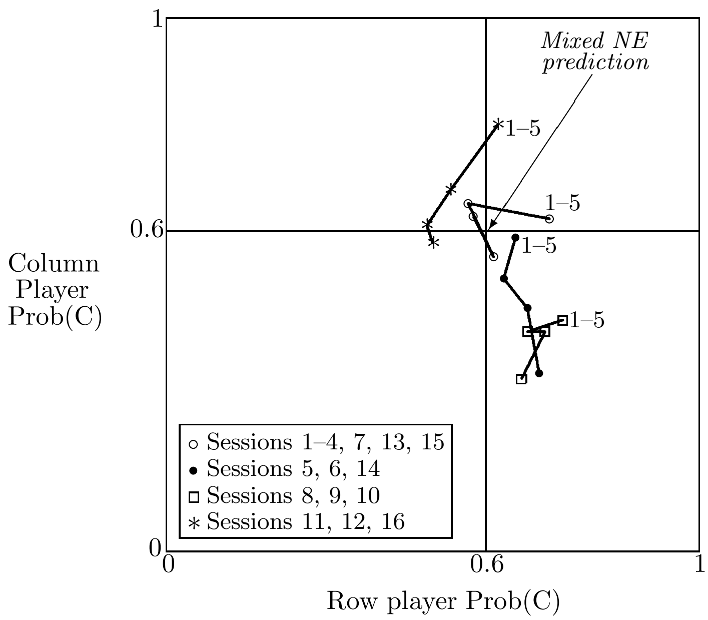
Note: "+": recommendations received in rounds 21–40; "–": recommendations received in rounds 1–20

4.1 Behavior without recommendations

We first examine subject behavior in rounds where subjects do not receive recommendations; this is shown in the top three rows of data in Table 2. There are minor, but insignificant, differences in aggregate choice frequencies according to whether the no–recommendation rounds were 1–20 or 21–40 (robust rank–order test, session–level data, $p > 0.10$ for (C,C) frequencies, $p > 0.20$ for the other three frequencies).[12] If we pool (C,D) and (D,C) outcomes, aggregate behavior comes very close to the mixed–strategy Nash equilibrium prediction of 16% (D,D) outcomes, 48% (C,D) and (D,C) outcomes, and 36% (C,C) outcomes. When these outcomes are treated separately, however, the substantially larger frequency of (C,D) than (D,C) outcomes means that mixed–strategy Nash equilibrium does less well. In fact, a chi–square test strongly rejects the null hypothesis that behavior in the no–recommendations rounds is generated by i.i.d. mixed–strategy equilibrium play ($p < 0.001$) when (C,D) and (D,C) outcomes are disaggregated, but not when they are pooled ($p > 0.20$). We note that this test assumes independence across subjects in a session and for each subject over time (so to the extent that these assumptions do not hold, the test will be excessively liberal).

The reason for the partial failure of mixed–strategy Nash equilibrium to characterize play in this treatment becomes clearer when we disaggregate the data further. Figure 4 shows the frequencies of C choices by both types of player in the experiment, disaggregated into groups of sessions that had qualitatively similar results, and also disaggregated into five–round blocks. The first such five-round block in a session is labeled in the figure (as "1–5"), and the path of play through the other three five–round blocks are shown via line segments. In a plurality (7 out of 16) of sessions (numbers 1–4, 7, 13 and 15), behavior

Figure 4: Dynamics of no–recommendations rounds (5–round blocks)



is close to the mixed–strategy prediction after the first block of five rounds. In the other nine sessions, behavior is not well described by mixed–strategy play, though the nature of the deviations varies. In three sessions (5, 6, and 14), behavior starts out near the mixed–strategy outcome but moves over time toward the pure–strategy outcome (C,D), while in three other sessions (8, 9, and 10), behavior starts and remains between the mixed–strategy outcome and (C,D). In the remaining three sessions (11, 12, 16) average play is also away from the mixed–strategy outcome, in the direction of the (D,C) pure–strategy outcome (though

[12]See Siegel and Castellan (1988) for descriptions of the nonparametric tests used in this paper. Critical values for the robust rank–order test are from Feltovich (2005).

it does not move in this direction over time).

While behavior in this no–recommendations treatment can be different from i.i.d. mixed–strategy equilibrium play, a weaker condition—statistical independence of row and column player choices—is broadly satisfied. To verify this, we calculated the phi coefficient of association (a measure of correlation for categorical data) for each five–round block of every session, giving us 64 of these coefficients. Of these 64, only 6 (that is, 9.375% of them) were significantly different from zero at the 10% level, and 4 (6.25% of the 64) were significant at the 5% level: roughly what would be expected by chance if row and column player choices are independent.

4.2 Effect of recommendations on population aggregates

Having examined how subjects behave without recommendations, we next look at whether recommendations have any effect. Table 2 above provides some strong evidence that they do.[13] When recommendations are generated from a correlated equilibrium (all but the very–good–recommendations treatment), outcome frequencies are significantly different from mixed–strategy probabilities—irrespective of whether we pool (C,D) and (D,C) outcomes—and the difference is in the direction predicted by correlated equilibrium (following recommendations). Both good recommendations and Nash recommendations increase the likelihood of a pure–strategy Nash equilibrium outcome, from 49.7% without recommendations to 56.5% with Nash recommendations and 57.9% with good recommendations, though this likelihood decreases slightly in the game with bad recommendations—to 48.1%—and with very good recommendations, to the lowest frequency of 46.9%. Also, the Pareto–dominated (D,D) outcome becomes more likely under bad or very good recommendations (19.2% and 16.9% of the time respectively, versus 15.9% when no recommendations are given) and less likely under good recommendations (14.0%) or Nash recommendations (11.3%). For some of the treatments, this last result might be expected in light of the outcome probabilities we attempted to impose: a 20% chance of (D,D) in the bad–recommendations treatment and 0% in the good– and Nash–recommendations treatment as compared with 16% in the mixed–strategy Nash equilibrium. However, this does not hold for the very–good–recommendations treatment, as the frequency of (D,D) recommended outcomes was 0% in this treatment as well.

Two–sample chi–square tests imply that the distributions of outcomes in the good–, bad–, and Nash–recommendations treatments are significantly different from the distribution without recommendations ($d.f. = 3$, $p < 0.02$ for each comparison), but there is not a significant difference between no recommendations and very–good recommendations ($\chi^2 = 3.511$, $d.f. = 3$, $p > 0.20$). The finding of no difference between no recommendations and very good recommendations is striking: it suggests that subjects will not blindly follow just *any* recommendations, but rather will follow them only if they are consistent with implementation of a correlated equilibrium. Indeed, one–sample chi–square tests find no significant difference between the distribution of outcomes in either the bad– or very–good–recommendations treatment and that implied by mixed–strategy equilibrium play ($\chi^2 = 5.188$, $d.f. = 3$, $p > 0.10$ and $\chi^2 = 0.577$, $d.f. = 3$, $p > 0.20$ respectively), while we do find significant differences from mixed–strategy equilibrium for the good– and Nash–recommendations treatments ($\chi^2 = 23.26$, $d.f. = 3$, $p < 0.001$ and $\chi^2 = 17.01$,

[13]In the following discussion, we will concentrate on the pooled data from sessions with recommendations first and sessions with recommendations last. We will see in Table 4 and the surrounding discussion that pooling the data in this way is justifiable.

$d.f. = 3$, $p < 0.001$ respectively).[14]

Furthermore, we note that in the cases where recommendations *were* consistent with implementation of a correlated equilibrium, aggregate outcome frequencies—while different from the point predictions of Table 1—typically move in the direction predicted by correlated equilibrium relative to the mixed–strategy Nash equilibrium prediction. For example, if subjects were always to follow recommendations in the good–, bad– and Nash–recommendations treatments, the resulting frequency of (C,C) outcomes would be lower than in the mixed–strategy Nash equilibrium. As Table 2 shows, the frequencies of (C,C) outcomes in these cases are indeed lower than in mixed–strategy Nash equilibrium. By contrast, in the very–good–recommendations case, if subjects followed recommendations, the predicted frequency of (D,D) outcomes would be lower than in the mixed Nash equilibrium (0 versus 0.16), but Table 2 shows that the observed frequency is actually higher. Finally, two–sample chi–square tests usually find significant differences in the distribution of outcomes between any two of the recommendations treatments ($p > 0.10$ for comparison of the good– and Nash–recommendations treatments and for comparison of the bad– and very–good–recommendations treatments, $p < 0.02$ for any other pair of treatments).

Another way of assessing whether recommendations have any effect involves testing for independence between row and column player choices. Recall from Section 4.1 that players' choices were found to be independent of each other when players did not receive recommendations. We now construct the phi coefficient of association for each session and five–round block when recommendations are received by subjects, giving us a total of 64 of these coefficients—16 for each treatment. In the good–recommendations treatment, 5 of the 16 five–round blocks have a significantly negative correlation at the 10% level, and 3 of these are significant at the 5% level, while in the Nash–recommendations treatment, 4 of 16 are negative and significant at the 5% level (with the other 12 not significant even at the 10% level). These proportions are higher than chance would predict, giving additional evidence that recommendations are having an effect on behavior. In the bad– and very–good–recommendations treatment, on the other hand, only 2 and 1 (respectively) of 16 are negative and significant at the 10% level or better, suggesting that recommendations have less effect in these treatments. We note, however, that the level of correlation implied even by perfect following of recommendations varies sharply across treatments, from perfect negative correlation in the Nash–recommendations treatment, to fairly high (in absolute value) correlations of $-\frac{2}{3}$ and $-\frac{1}{2}$ in the bad– and good–recommendations treatments respectively, to the nearly zero correlation of $-\frac{1}{9}$ in the very–good–recommendations treatment. As a result, direct comparisons across treatments should be made with caution.

Summarizing, we have:

Result 1 *When no recommendations are given, aggregate outcome frequencies are broadly similar to mixed–strategy Nash equilibrium frequencies, though in some sessions there is a tendency toward one of the pure–strategy Nash equilibria. When recommendations are given, they lead to significant differences in aggregate outcome frequencies, compared with the no–recommendations case, if and only if the recommen-*

[14] As in the previous section, we note here that these chi–square statistics assume independence across subjects in a session and for each subject over time. Also, note that observed frequencies in all treatments are significantly different from any of the *correlated*–equilibrium predictions, as each of the latter predicts zero probability of at least one outcome that occurs with positive frequency in the experimental data. Finally, pooling (C,D) and (D,C) outcomes has little qualitative effect on these significance tests; the lone exception is that outcomes in the bad–recommendations treatment are significantly different from mixed–strategy equilibrium at the 10% level when these are pooled, but not when they are treated separately.

dations come from a correlated equilibrium. Also, there are significant differences in aggregate outcomes across the treatments with recommendations. When recommendations come from a correlated equilibrium, the effect on aggregate outcome frequencies is consistent with the directional predictions—though usually not the point predictions—of the corresponding correlated equilibrium.

Table 2 also shows the per–round average payoff earned by subjects in each treatment; these vary from a low of 4.902 under bad recommendations to a high of 5.648 under Nash recommendations. However, a nonparametric Kruskal–Wallis one–way analysis of variance fails to reject the null hypothesis that average payoffs are the same in all four recommendations treatments (session–level data, $p > 0.10$), and robust rank–order tests find no significant differences in pairwise comparisons between treatments (session–level data, $p > 0.10$ in all cases). This lack of significance in the payoff dimension is likely owing to the relatively small differences in predicted expected payoffs amongst the various correlated equilibria, combined with the inherent conservatism of nonparametric tests using session–level data.

4.3 Effects of recommendations on individual behavior

Having shown that aggregate play with recommendations is usually different from aggregate play without recommendations—and that this difference depends on which recommendations are given—we next consider how subjects treat the particular recommendations they receive. Table 3 shows the frequencies with which recommendations are followed in each treatment over all twenty rounds as well as for the last five rounds of each treatment (after subjects have had time to gain experience with the strategic environment). For the sake of comparison, the table also shows the corresponding predicted frequencies according to mixed–strategy Nash equilibrium. (Note that the predictions in the last two columns depend on the actual frequencies of C versus D recommendations given in the experiment, so these will vary across treatments and rounds.[15])

Here we see more differences across treatments. In the good– and Nash–recommendations treatments, subjects are substantially more likely to follow recommendations than would be predicted by the mixed–strategy Nash equilibrium. They follow D recommendations 73.5% of the time in the good–recommendations treatment and 56.7% of the time in the Nash–recommendations treatment, compared to a prediction of 40%, and they follow C recommendations 73.2% of the time in the good–recommendations treatment and 77.7% of the time in the Nash–recommendations treatment, compared to a prediction of 60%. As the table shows, these frequencies are even higher if we concentrate on the last five rounds of the treatment, and all of these differences are significant (one–tailed sign test, session–level data, $p = 0.0625$).

In the bad– and very–good–recommendations treatments, evidence that subjects follow recommendations is weaker. If we look at frequencies for the entire bad–recommendations treatment, subjects are

[15] Specifically, the predictions in the third and fourth columns are based on the predictions of 0.4 and 0.6 for the frequencies of following D and C recommendations (from the first and second columns). For the third column—overall frequency of followed recommendations—we take the mixed–strategy Nash equilibrium prediction in a treatment to be the weighted average of 0.4 and 0.6, with the weightings equal to the actual observed frequencies of D and C recommendations in that treatment. As an example, in the good–recommendations treatment, C recommendations were actually made to subjects 628 times and D recommendations 332 times. The predicted frequency of following recommendations overall is then $0.6\left(\frac{628}{960}\right) + 0.4\left(\frac{332}{960}\right) = 0.531$. In a similar way, the predictions for the fourth column (frequency of both paired players following recommendations) are weighted averages of 0.16, 0.24, 0.24, and 0.36 (predictions conditional on recommended outcomes of (D,D), (D,C), (C,D), and (C,C) respectively), weighted by the actual frequencies of recommendation pairs.

Table 3: Frequencies of followed recommendations—all subjects, all rounds

Treatment		Frequency of followed D recommendations	Frequency of followed C recommendations	Frequency of followed recommendations (overall)	Frequency of followed recommendations (pairs)
Good	Observed (all rounds)	0.735*	0.732*	0.733*	0.531*
	Observed (rnds 16–20)	0.750*	0.770*	0.762*	0.583*
	Mixed NE prediction	0.400	0.600	0.531	0.277
Bad	Observed (all rounds)	0.477	0.631	0.541	0.269*
	Observed (rnds 16–20)	0.529*	0.530	0.529*	0.300*
	Mixed NE prediction	0.400	0.600	0.483	0.226
Nash	Observed (all rounds)	0.567*	0.777*	0.672*	0.454*
	Observed (rnds 16–20)	0.608*	0.792*	0.700*	0.517*
	Mixed NE prediction	0.400	0.600	0.500	0.240
Very good	Observed (all rounds)	0.511*	0.608	0.599	0.381
	Observed (rnds 16–20)	0.227	0.537	0.508	0.308
	Mixed NE prediction	0.400	0.600	0.580	0.336

*: *Significantly above corresponding mixed–strategy prediction (one–tailed sign test, session–level data,* $p = 0.0625$*)*

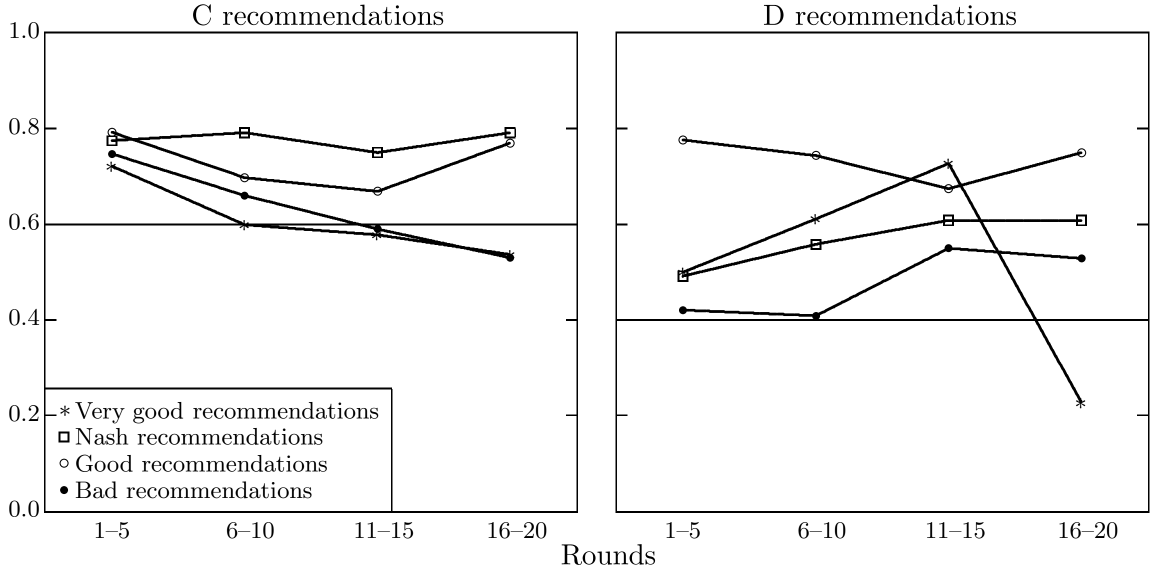
not significantly more likely to follow either type of recommendation than predicted by mixed–strategy equilibrium (one–tailed sign test, session–level data, $p > 0.10$), though the frequencies are slightly higher than predicted (47.7% for D recommendations and 63.1% for C recommendations, versus predictions of 40% and 60%). However, if we focus on the last five rounds, we find a higher frequency of following D recommendations, and this frequency is significantly higher than the mixed–strategy equilibrium prediction ($p = 0.0625$), though we also see that subjects actually become less likely to follow C recommendations in the last five rounds. Subjects in the very–good–recommendations treatment are not significantly more likely to follow C recommendations than mixed–strategy equilibrium predicts ($p > 0.10$), either over all rounds or in the last five. They are more likely to follow D recommendations over all rounds (51.1% versus a predicted 40%), and this difference is significant ($p = 0.0625$), but this frequency drops sharply in the last five rounds to below one–fourth, and in those rounds is not significantly different from the mixed–strategy equilibrium prediction ($p > 0.10$).

Given how these results vary with the treatment, it should not be surprising that we find significant differences across treatments in how often recommendations are followed. A Kruskal–Wallis one–way analysis of variance rejects the null hypothesis that the likelihood of following a C recommendation is the same across the four treatments ($p < 0.05$), and similarly for the case of a D recommendation ($p < 0.05$). The null of equal frequencies across treatments of following recommendations overall is also rejected ($p < 0.05$), but we should note that mixed–strategy Nash equilibrium does not imply equal frequencies in this case.

We next examine how subjects' willingness to follow recommendations changes over time. Figure 5 shows the frequency with which recommendations are followed in each five–round block, disaggregated according to which correlated equilibrium was being implemented, and which action was recommended.

For C recommendations, there are no obvious time trends in the good– and Nash–recommendations treat-

Figure 5: Frequency of followed recommendations



ments, while subjects in the bad– and very–good–recommendations treatments become less likely over time to follow these (falling from about 75% to just over 50% in both). The frequency of following D recommendations stays roughly constant over time in the good–recommendations treatment and rises slightly in the bad– and Nash–recommendations treatment. In the very–good–recommendations treatment, sample sizes for D recommendations are small (since only one–tenth of recommendations is for a D choice), but their frequency of being followed rises somewhat from the first to the third five–round block, before plummeting in the last five–round block.

Further evidence of the effects of recommendations on individual subject choices can be found in Table 4, which reports the results of several probit regressions with the subject's choice of action as the dependent variable. (To be precise, the dependent variable is an indicator for a C choice.) Our main independent variables are two indicators for recommendations given to subjects—one for a C recommendation (viz., taking on the value of one if a C recommendation was made, and zero otherwise) and one for a D recommendation. (To avoid perfect collinearity, we do not include an indicator for no recommendation.) We also include variables for the products of these indicators with the round number, to capture any time–varying effect of recommendations that exists. Additionally, we include a variable for the round number itself, as well as an indicator variable that takes the value one in sessions in which recommendations were given in the first twenty rounds rather than the last twenty (to capture any order effects).

The regressions were performed using Stata (version 10) and incorporate individual–subject random effects; we estimate coefficients separately for the four treatments. The results are shown in Table 4, which shows the coefficient and standard error for each variable in our four model specifications. (We additionally estimated specifications with individual–session fixed effects, but the results were nearly identical to those reported here.) Also shown is the absolute value of the log–likelihood, as well as a pseudo–R^2, for each model specification.

We do not find evidence of substantial order effects between treatments (that is, our results are robust to whether the rounds with recommendations came before or after the rounds without recommendations),

Table 4: Results of probit regressions with random effects (standard errors in parentheses)

Dependent variable: cooperative action chosen in round t	Good–recommendations treatment ($N = 1920$)	Bad–recommendations treatment ($N = 1920$)	Nash–recommendations treatment ($N = 1920$)	Very–good–recommendations treatment ($N = 1920$)
constant	0.372	0.464	0.727***	0.956***
	(0.237)	(0.295)	(0.250)	(0.241)
Order (indicator for order effects)	0.046	0.232	–0.097	–0.440
	(0.312)	(0.392)	(0.327)	(0.314)
t (round number)	0.001	−0.022**	−0.027***	−0.048***
	(0.008)	(0.009)	(0.008)	(0.008)
Drec (D recom–mendation given)	−1.098***	–0.075	−0.463***	−1.313***
	(0.207)	(0.181)	(0.172)	(0.299)
Drec $\cdot\, t$	–0.001	−0.022***	–0.009	0.088***
	(0.017)	(0.014)	(0.015)	(0.025)
Crec (C recom–mendation given)	0.593***	0.338*	0.472**	0.086
	(0.162)	(0.200)	(0.184)	(0.150)
Crec $\cdot\, t$	–0.014	–0.021	0.021	0.002
	(0.013)	(0.016)	(0.015)	(0.012)
–ln(L)	970.789	898.429	950.705	971.544
pseudo–R^2	0.106	0.027	0.086	0.037

* (**,***): *Coefficient significantly different from zero at the 10% (5%, 1%) level.*

as the "Order" variable is never significant. On the other hand, there is some nonstationarity in the data, as shown by the negative and significant coefficient on the round number t in three of the four treatments (the lone exception being the good–recommendations treatment). The significance of the recommendation variables varies substantially across treatments. In the good and Nash–recommendations treatments, the C–recommendations and D–recommendations indicators are both significant, but their products with the round number are not, and each has the sign associated with subjects' following recommendations: positive for C and negative for D. In the bad–recommendations treatment, the D–recommendations indicator is insignificant, but its product with the round number is significant; the coefficient of the C–recommendations indicator is barely significantly different from zero, while that of its product with the round number is insignificant. In the very–good–recommendations treatment, both the C–recommendations indicator and its product with the round number are significant, with the former negative and the latter positive, but neither of the D–recommendations variables are significant.

Based on these results, we conclude:

Result 2 *There are significant differences across treatments in how subjects use their recommendations. Subjects are most likely to follow recommendations in the good– and Nash–recommendations treatments. In the bad–recommendations treatment, recommendations have less effect on behavior. In the very–good–recommendations treatment, subjects either don't follow recommendations at all, or learn over time not to follow them.*

We next consider whether the likelihood of a subject following recommendations depends on the subject's past history. One possibility is that subjects will be more willing to follow recommendations if following them has been successful in the past. As a first step, we look at how a subject's propensity to follow recommendations is affected by the success of following recommendations in the previous round. Specifically, we consider how consistent behavior in our experiment is with Selten and Stoecker's (1986) "direction learning theory". When a game has a one–dimensional strategy space, direction learning theory predicts that when a player changes strategy from one round to the next, the change will be in the direction of the (myopic) best response. In our setup, this implies that when a subject chose successfully (i.e., chose a best response to the opponent's action) in the previous round, she will continue following recommendations in the current round if she had done so in the previous round, or will not follow recommendations in the current round if she did not follow them in the previous round. On the other hand, if the subject chose unsuccessfully in the previous round, she will do the opposite of what she did in that round—follow recommendations if she had not done so, or not follow recommendations if she had. By and large, our data are only weakly consistent with direction learning. After a successful choice, subjects stay with the same strategy (follow or not follow recommendations) 64.6% of the time, but they stay with the same strategy almost as often (60.6% of the time) following an unsuccessful choice.

An alternative possibility is that subjects' choices depend not just on the previous round outcome, but instead on the entire history of play. To examine this possibility, we first construct measures for the success rate of following recommendations and that of not following recommendations. The success rate of following recommendations is set to one–half if a subject has never followed a recommendation; otherwise, it is equal to the proportion of times (in all rounds up through the previous round) in which following recommendations led to a best response to the opponent action. (For example, if a player has thus far followed recommendations 5 times, and 3 of these turned out to be best responses, then the success rate would be 0.6.) The success rate of not following recommendations is calculated in an analogous way.

We then use the difference between these success rates (the rate for following recommendations minus the one for not following recommendations) as an explanatory variable, where the dependent variable is an indicator for following recommendations in the current round. The results can be seen in Table 5. This table shows four alternative probit specifications, differing in which other explanatory variables are

Table 5: Results of probit regressions with random effects, rounds 2–20 of treatments with recommendations (standard errors in parentheses)

Dependent variable: follow recommendations in round t ($N = 3648$)	Model specification #1	Model specification #2	Model specification #3	Model specification #4
constant	0.381***	0.445***	0.294***	0.359***
	(0.049)	(0.066)	(0.093)	(0.105)
t (round number)		–0.006		–0.006
		(0.004)		(0.004)
Success–rate difference	0.196**	0.208***	0.150*	0.161**
	(0.079)	(0.079)	(0.081)	(0.081)
Good recommendations			0.357***	0.352***
			(0.134)	(0.134)
Bad recommendations			–0.194	–0.198
			(0.131)	(0.130)
Nash recommendations			0.200	0.194
			(0.135)	(0.135)
–ln(L)	2226.897	2225.886	2217.141	2216.220
pseudo–R^2	0.001	0.002	0.006	0.006

* (**,***): *Coefficient significantly different from zero at the 10% (5%, 1%) level.*

included: the round number, indicators for the treatments, or both. For all of these specifications, the coefficient for the success–rate difference is positive and significant, suggesting that subjects are indeed more likely to follow recommendations, the more successful following them has been in the past—relative to not following them. (On the other hand, the low pseudo–R^2 values suggest that past success is only a minor factor in explaining whether recommendations are followed.)

We thus have

Result 3 *Subjects' following of recommendations is affected by history. The more successful following recommendations has been in the past, or the less successful not following recommendations has been, the more likely a subject is to follow recommendations in the current round.*

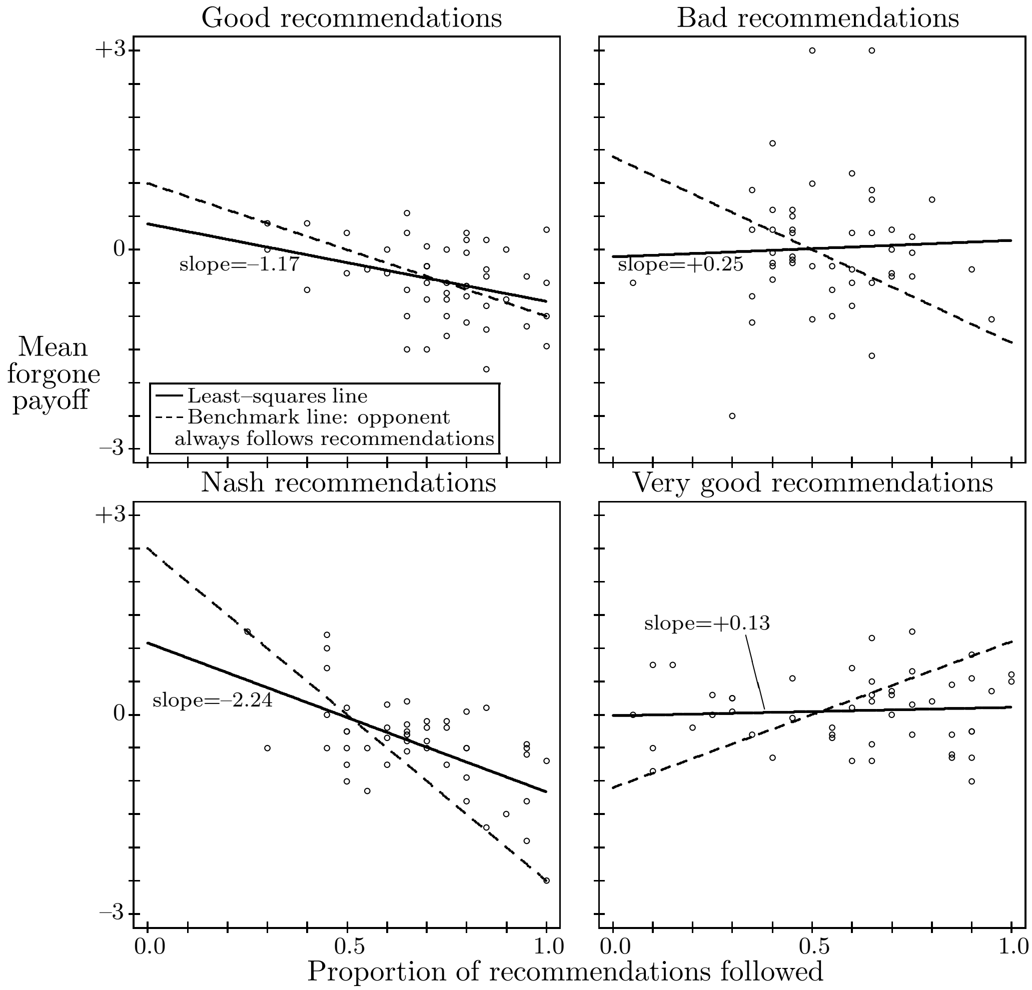
We next look at whether subjects who fail to follow recommendations suffer (monetarily) as a result. To examine this question, we consider subjects' *forgone payoffs*: the payoff a subject would have gotten from choosing the other action, minus the payoff the subject actually got. (Thus, a negative forgone payoff means the subject chose a best response.)

Overall, in rounds with no recommendations, forgone payoffs averaged –0.079 points per round, which is not significantly different from zero (two–tailed Wilcoxon signed–ranks test, session–level data, $p \approx 0.40$), meaning that on average, subjects earned approximately the same payoffs with the actions they chose than they would have earned by choosing the opposite action—as would be implied by mixed–strategy equilibrium play. Forgone payoffs averaged –0.467 points per round in the good–recommendations treatment and –0.428 points per round in the Nash–recommendations treatment, both of which are significantly less than zero (one–tailed Wilcoxon test, session–level data, $p = 0.0625$ for both), while forgone payoffs averaged +0.026 points per round in rounds with the bad–recommendations treatment and +0.059 points per round in the very–good–recommendations treatment, neither of which is significantly less than zero ($p = 0.3125$ and $p = 0.6825$ respectively), suggesting that subjects in the good– and Nash–recommendations treatments by and large made correct choices, while subjects in the other two treatments did not. Since the good– and Nash–recommendations treatments were also the ones where subjects were most likely to follow recommendations, the implication is that following recommendations is indeed positively associated with better outcomes for the individual subject, at least on average.

However, we are interested less in these treatment–wide aggregates than in how forgone payoffs are associated with how often subjects followed the recommendations they were given. In Figure 6, we present scatterplots showing, for each individual subject, the proportion of recommendations that were followed (on the horizontal axis) and the subject's mean forgone payoff (on the vertical axis). Also shown are two lines for each scatterplot: one is a least–squares line showing the actual average relationship between following recommendations and foregone payoffs (with the slope shown as well), and the other is a benchmark line showing what this relationship would be under the assumption that the opposing player always follows recommendations.[16] The benchmark line is negatively sloped for the three treatments in which recommendations form a correlated equilibrium, reflecting the incentives in these treatments for players to follow recommendations as long as their opponents are expected to do so. In contrast, if opponents always follow recommendations in the very–good–recommendations treatment, the incentives are against following recommendations, as evidenced by the positive slope of that benchmark line.

[16]This benchmark makes the additional assumptions that (1) the player receives C and D recommendations in proportion equal to the underlying probabilities of C and D recommendations in that treatment, and (2) the player's likelihoods of following C and D recommendations are equal.

Figure 6: Relationship between followed recommendations and forgone payoffs *(Individual subjects, all rounds)*



For subjects in the good– and Nash–recommendations treatments, the least–squares line shows a visible negative correlation between following recommendations and forgone payoffs, suggesting that following recommendations more often was associated with better payoffs for individual subjects in these treatments, as was true for the benchmark line. The least–squares lines are less steep in these treatments than the benchmark lines, due to not all opponents following recommendations in reality (while we've assumed they do for the benchmark). On the other hand, the least–squares line shows no apparent correlation for subjects in the bad– and very–good–recommendations treatments, by contrast with the benchmark lines. The implication is that in these two treatments, enough other subjects do not follow recommendations to remove nearly all incentives to follow recommendations in the bad–recommendations treatment, or not to follow recommendations in the very–good–recommendations treatment.

Spearman rank–order correlation tests provide further, quantitative, evidence of these results. The Spearman correlation coefficient between frequency of followed recommendations and mean forgone payoff is approximately –0.312 in the good–recommendations treatment and –0.469 in the Nash–recommendations treatment, both of which are significantly different from zero ($p \approx 0.02$ for the former and $p < 0.001$ for the latter), suggesting that following recommendations more often was associated with better payoffs for individual subjects. In the bad– and very–good–recommendations treatments, on the other hand, the

Spearman coefficients are approximately +0.003 and +0.092 respectively, neither of which is significantly different from zero ($p \approx 0.98$ and $p \approx 0.25$, respectively), suggesting that subjects in these treatments did not do better by following recommendations than by ignoring them. The least–squares lines give additional evidence of these relationships; their slopes are negative and significantly different from zero in the good– and Nash–recommendations treatments ($p < 0.01$ for both treatments, using robust standard errors adjusted for clustering by session), and are not significantly different from zero in the bad– and very–good–recommendations treatments ($p > 0.20$ for both treatments).

Finally, disaggregating by round and according to the recommended action tells a more detailed, but similar, story. Linear panel–data regressions with individual–subject random effects, either with or without session fixed effects, show that a subject's following either type of recommendation in either the good–recommendations treatment or the Nash–recommendations treatment is associated with significant decreases in forgone payoffs, as is following a C recommendation in the bad–recommendations treatment. By contrast, there is no significant association between forgone payoffs and either following D recommendations in the bad–recommendations treatment or following C recommendations in the very–good–recommendations treatment. Finally, following D recommendations in the very–good–recommendations treatment is positively correlated with foregone payoffs; that is, following D recommendations actually lowers a player's payoff in that treatment.[17]

We thus conclude:

Result 4 *In the good and Nash–recommendations treatments, it pays subjects (individually) to follow either type of recommendation. In the bad–recommendations treatment, it pays subjects to follow C recommendations, but there is no statistically significant relationship between following D recommendations and payoffs. In the very–good–recommendations treatment, there is no significant relationship between following C recommendations and payoffs, and it pays subjects not to follow D recommendations.*

5 Summary and discussion

The aim of this paper was to assess the empirical validity of correlated equilibrium, an important generalization of the Nash equilibrium concept. Specifically, we have explored whether subjects make use of known (and publicly announced) distributions of private third–party recommendations as a coordination device in the game of Chicken, the simplest game with which to study a wide variety of correlated equilibria. The treatments in our experiment differ in the distributions of recommendations. Three of our four treatments use distributions that form correlated equilibria; two of these yield symmetric payoffs that are outside the convex hull of Nash equilibrium payoff vectors. In our "good" correlated equilibrium, payoffs are better than any symmetric payoff in the convex hull of Nash equilibrium payoff vectors, while in our "bad" correlated equilibrium, payoffs are worse than any symmetric payoff in the convex hull of

[17] These regressions used the subset of the data in which recommendations were given. The dependent variable is forgone payoff, and the independent variables are C recommendation, D recommendation, followed C recommendation, followed D recommendation. No constant term was used. In the results, p–values were below 0.001 for both types of recommendation in the good– and Nash–recommendations treatment, approximately 0.043 for C recommendations in the bad–recommendations treatment, and approximately 0.77 for D recommendations in the bad–recommendations treatment. In the very–good–recommendations treatment, the p–value was approximately 0.78 for C recommendations and 0.030 for D recommendations, but the coefficient for the latter was positive. Adding session fixed effects to these regressions had little qualitative effect.

Nash equilibrium payoff vectors. A third, "Nash" treatment uses a correlated equilibrium with payoffs in the convex hull of Nash equilibrium payoff vectors, and a fourth, "very good" treatment uses an outcome distribution yielding high payoffs, but which is not a correlated equilibrium.

We find that when subjects do not receive recommendations, their choices can be described fairly well (though not perfectly) by mixed–strategy Nash equilibrium. This result suggests that theoretical rationales for correlated equilibria that do not rely on extrinsic, third–party recommendations (or some other "external event space" in the terminology of Vanderschraaf (2001)) might be difficult to observe in practice—though we acknowledge the possibility that if subjects had interacted in fixed pairings rather than under the random matching protocol we adopted, or had opportunities for communication (as in Moreno and Wooders (1998)), then spontaneously–arising correlated equilibrium might have been more likely to have been observed.

By contrast, giving subjects recommendations nearly always has an effect on behavior, but the effect depends on what recommendations are given. When recommendations are based on an underlying correlated equilibrium, subjects follow them more often than mixed–strategy equilibrium predicts, though far less than 100% of the time, and varying with the correlated equilibrium. When recommendations are not based on a correlated equilibrium, subjects learn to ignore them.

As in previous efforts to implement correlated equilibria in the laboratory, our results cast a bit of doubt on the usefulness of this solution concept as a descriptive notion, as the correlated equilibrium point predictions are not observed. On the other hand, our study reveals several new and important empirical findings about the correlated equilibrium concept. First, the lesson of our good– and Nash–recommendations treatments is that it is not necessary for recommendations always to be followed in order for them to have an effect. Recommendations in these treatments were followed only roughly 70–75% of the time, but this was enough to have a significant effect on the distribution of outcomes.[18] (There was also a positive, but insignificant, effect on average payoffs.) Second, the lesson of our very–good–recommendations treatment is that correlated equilibrium is likely a *necessary condition* for recommendations to have any substantial effect on behavior. In particular, we found that, consistent with the theoretical prediction, subjects were not blindly following recommendations in the very–good-recommendations treatment (as they would have if they were, for example, simply trying to please the experimenters, or choosing high–payoff outcomes irrespective of the outcomes' strategic properties). Third, our bad–recommendations treatment shows that it is particularly difficult to induce subjects to follow recommendations based on correlated equilibria that are Pareto *inferior* to the available Nash equilibria. This finding would seem to greatly limit the class of empirically relevant correlated equilibria to those that Pareto improve upon the set of Nash equilibria.[19]

Future theoretical and empirical work on the topic of correlated equilibria might relax the assumption

[18]This frequency is comparable to that found by Cason and Sharma (2007), whose baseline treatment resembles our good–recommendations treatment. Unlike Cason and Sharma, we do not attempt here to disentangle among competing explanations for subjects' failure to follow recommendations (see Note 8), such as social preferences, uncertainty about whether the opponent will follow recommendations, or simply decision errors; however, we expect that their finding—that all three of these factors have some impact—should apply to our subjects as well.

[19]A referee has pointed out an alternative explanation for our results: that recommendations are more likely to be followed when the underlying outcome distribution allows easy application of Bayes's rule. Posterior probabilities in the good– and Nash–recommendations treatments are always 0, 0.5, or 1, while they can be $\frac{1}{3}$ or $\frac{2}{3}$ in the bad–recommendations treatment and $\frac{1}{9}$ or $\frac{8}{9}$ in the very–good–recommendations treatment; the increased cognitive requirements involved in calculating expected payoffs using these latter fractions might have pushed frequencies of following recommendations toward one–half.

that recommendations arise from a non–strategic third party according to deterministic (and commonly known) probabilities. In place of this construct, a self–interested "mediator" player might repeatedly choose recommendations to make to the players of the stage game. In such an environment, the mediator's payoff could be based on the payoffs earned by the stage–game players: for example, it might be proportional to their average payoff. In this setting, the researcher could explore whether the mediator's frequencies of recommendations to players were consistent with any correlated equilibrium, and if so, which one: good, bad, Nash, or some other one.[20]

A second useful extension would be to consider some "language" issues as they apply to correlated equilibrium. For instance, one might wonder whether the form of recommendations matters: for example, whether subjects are told, "It is recommended that you play C", as in our design, or they simply see the message "C" on their screens. The salience and literal meanings of recommendations are also of interest: must the message space for recommendations correspond precisely to the action space, or might it be larger (for example, including also "no message"), or consist of a set of messages with no clear mapping to the action space (such as the message space {@, & })?[21]

We leave these extensions to future research.

References

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

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

Blume, A., D. DeJong, Y.-G. Kim, and G.B. Sprinkle (1998). "Experimental evidence on the evolution of meaning of messages in sender–receiver games," *American Economic Review* 88, pp. 1323–1340.

Blume, A., D. DeJong, Y.-G. Kim, and G.B. Sprinkle (2001). "Evolution of communication with partial common interest," *Games and Economic Behavior* 37, pp. 79–120.

Brandenburger, A. and A. Friedenberg (2008), "Intrinsic correlation in games," *Journal of Economic Theory* 141, pp. 28–67.

Brandts, J. and B. McLeod (1995), "Equilibrium selection in experimental games with recommended play," *Games and Economic Behavior* 11, pp. 36–63.

Cason, T.N. and T. Sharma (2007), "Recommended play and correlated equilibria," *Economic Theory* 33, pp. 11–27.

Crawford, V.P. (1998), "A survey of experiments on communication via cheap talk," *Journal of Economic Theory* 78, pp. 286–298.

[20] We have begun to explore this possibility (Duffy and Feltovich (in preparation).

[21] To this last question, we note that the recommendations in this setup are a special case of games with cheap talk. Blume et al. (1998, 2001) found that in a number of cheap–talk games, messages typically did eventually acquire literal meanings even when they didn't initially have them. See also Crawford (1998) for a more general survey of games with cheap talk.

Duffy, J. and E.O'N. Fisher (2005), "Sunspots in the laboratory," *American Economic Review* 95, pp. 510–529.

Duffy, J. and N. Feltovich (2008), "Endogenous correlated equilibria: Theory and evidence," in preparation.

Einy, E. and B. Peleg (1995), "Coalition–proof communication equilibria," in W.A. Burnett et al., eds., *Social Choice, Welfare and Ethics*, Cambridge: Cambridge University Press, Chapter 13.

Feltovich, N. (2005), "Critical values for the robust rank–order test," *Communications in Statistics—Simulation and Computation* 34, pp. 525–547.

Fischbacher, U. (2007), "z–Tree: Zurich toolbox for ready–made economic experiments," *Experimental Economics* 10, pp. 171–178.

Foster, D. and R.V. Vohra (1997), "Calibrated learning and correlated equilibrium," *Games and Economic Behavior* 21, pp. 40–55.

Fudenberg, D. and D.K. Levine (1998), *The Theory of Learning in Games*, Cambridge, MA: MIT Press.

Fudenberg, D. and D.K. Levine (1999), "Conditional universal consistency," *Games and Economic Behavior* 29, pp. 104–130.

Gul, F. (1998), "A comment on Aumann's Bayesian view," *Econometrica* 66, pp. 923–927.

Hart, S. and A. Mas-Colell (2000), "A simple adaptive procedure leading to correlated equilibrium," *Econometrica* 60, pp. 1127–1150.

Hofbauer, J. and K. Sigmund (1998), *Evolutionary Games and Population Dynamics*, Cambridge: Cambridge University Press.

Luce, R.D. and H. Raiffa (1957), *Games and Decisions: Introduction and Critical Survey*, New York: Wiley.

Moreno, D. and J. Wooders (1996), "Coalition–proof equilibrium," *Games and Economic Behavior* 17, pp. 113–128.

Moreno, D. and J. Wooders (1998), "An experimental study of communication and coordination in noncooperative games," *Games and Economic Behavior* 24, pp. 47–76.

Moulin, H. and J.-P. Vial (1978), "Strategically zero–sum games: the class of games whose completely mixed equilibria cannot be improved upon," *International Journal of Game Theory* 7, pp. 201–221.

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

Seely, B., J.B. Van Huyck and R.C. Battalio (2005), "Credible assignments can improve efficiency in laboratory public goods games," *Journal of Public Economics* 89, pp. 1437–1455.

Selten, R. and R. Stoecker (1986), "End behaviour in sequences of finite prisoner's dilemma supergames: a learning theory approach," *Journal of Economic Behavior and Organization* 7, pp. 47–70.

Sharma, T. and R. Torres (2004), "Coordination in teams," working paper, Instituto Tecnológico Autónomo de México.

Siegel, S. and N.J. Castellan, Jr. (1988), *Nonparametric Statistics for the Behavioral Sciences*, 2nd Ed., New York: McGraw Hill.

Skyrms, B. (1996), *Evolution of the Social Contract*, Cambridge: Cambridge University Press.

Van Huyck, J.B., A.B. Gilette and R.C. Battalio (1992), "Credible assignments in coordination games," *Games and Economic Behavior* 4, pp. 606–626.

Vanderschraaf, P. (2001), *Learning and Coordination: Inductive Deliberation, Equilibrium, and Convention*, New York: Routledge.

Vanderschraaf, P. and B. Skyrms (2003), "Learning to take turns," *Erkenntnis* 59, pp. 311—348.

Young, P. (2005), *Strategic Learning and Its Limits*, Oxford: Oxford University Press.