The final publication is available at http://dx.doi.org/10.1016/j.jebo.2015.06.008.

Does Payoff Equity Facilitate Coordination? A test of Schelling's Conjecture*

Raúl López-Pérez[a], Ágnes Pintér[a ‡], and Hubert J. Kiss[b]

[a] *Department of Economic Analysis, Universidad Autónoma de Madrid, 28049, Cantoblanco, Madrid, Spain.*

E-mail addresses: raul.lopez@uam.es (R. López-Pérez); pinteragnes@gmail.com (Á. Pintér)

[b] *Department of Economics, Eötvös Loránd University, Lágymányosi Campus, 1117 Budapest, Hungary. Kiss is also a research fellow in the Momentum (LD-004/2010) Game Theory Research Group at the MTA KRTK, Hungary.* E-mail address: kisshub@tatk.elte.hu

Abstract: Starting from Schelling (1960), several game theorists have conjectured that payoff equity might facilitate coordination in normal-form games with multiple equilibria –the more equitable equilibrium might be selected either because fairness makes it focal or because many individuals dislike payoff inequities, as abundant experimental evidence suggests. In this line, we propose a selection principle called Equity (EQ), which selects the equilibrium in pure strategies minimizing the difference between the highest and smallest money payoff, if only one such equilibrium exists. Using a within-subjects experimental design, furthermore, we study the relative performance of EQ in twelve simple 2x2 coordination games. In many of these games, we find that EQ explains individual behavior better than a large range of alternative theories, including theories of bounded rationality and several other equilibrium selection principles. Yet we also observe that the frequency of EQ play depends on the payoff structure of the game. For instance, EQ play diminishes when the alternative equilibrium is socially efficient and not very unfair (compared with the EQ equilibrium). Our data suggests that equilibrium selection is affected by several factors and that subjects are heterogeneous in this respect, but also that equity is often a crucial factor to understand coordination.

Keywords: Coordination; equity; experiments; inequity aversion; level-k thinking; payoff and risk dominance.

JEL Classification: C72; C91; D03; D63.

---

* The authors greatly acknowledge financial support from the Spanish Ministry of Education through the research projects ECO2008-00510 and ECO2011-25349. We also want to thank Andreas Leibbrandt, Santiago Sanchez-Pages, Ismael Rodríguez-Lara, Alfonso Rosa-García, one associate editor, and numerous participants at several conferences who provided valuable feedback.

‡ Corresponding author. Phone number: +34 91 4972960 (Á. Pintér)

1. Introduction

Coordination characterizes strategic interaction in many fields of economics and everyday life, and indeed becomes crucial when a group of agents shares a common goal and there are different ways to achieve it. Examples of such type of situations include the choice of an industry standard, deciding where and when to meet somebody, choosing a fast route in the city traffic, or organizing teamwork and division of labor –see Schelling (1960) for additional examples.

A rich experimental literature has found several factors that affect coordination.[1] For instance, pre-play talk often facilitates coordination on Pareto efficient equilibria, especially under certain communication protocols (Cooper et al., 1989 and 1992; Van Huyck et al., 1992; Blume and Ortmann, 2007). Another example is repeated play, in particular when playing with the same partner (Palfrey and Rosenthal, 1994; Brandts and Cooper, 2006). Our paper contributes to this literature, exploring the role of one factor that has not received so much attention, that is, payoff equity.

More precisely, we propose and test experimentally a simple selection principle called Equity or EQ, which selects the equilibrium in pure strategies minimizing the difference between the highest and smallest *pecuniary* payoff, if only one such equilibrium exists. This principle is partially based on a conjecture by Schelling (1960, p. 61) that agents find focal or salient an equilibrium if it is fair. More precisely, Schelling considers a bargaining game in which two players simultaneously make claims over \$100; each player getting his claim if the two claims add to no more than \$100, and zero otherwise. Schelling claims that the 50-50 equilibrium is salient, and provides data from an informal sample showing that almost all subjects achieve coordination on that outcome. EQ generalizes in a simple manner Schelling's conjecture that fairness can facilitate coordination, a point which can be illustrated using the game below, where we depict monetary payoffs.[2] Note that the game has two equilibria in pure strategies –i.e., (R1, C1) and (R2, C2) – if we posit that all players are selfish and hence care only about their own pecuniary payoff. Thus, a coordination problem exists in the typical game-theoretical sense.

	C1	**C2**
R1	75, 140	75, 75
R2	70, 70	110, 110

EQ predicts that both players will act according to equilibrium (R2, C2), which presents the smallest payoff difference. We find the EQ prediction intuitive for several reasons. To start, players might find the EQ equilibrium focal simply because it is fair and equitable, as Schelling suggested.

[1] See Crawford (1997), Camerer (2003), and Devetag and Ortmann (2007) for reviews of some of this literature.
[2] Throughout the paper the Row player's payoff/strategy comes first in any payoff/strategy profile.

In this line, Myerson (1997, p. 112) noted that “welfare properties of *equity* and *efficiency* may determine the focal equilibrium in a game” (in italics in the original). A more subtle, alternative argument is based on the idea that some people may have preferences such that they dislike inequity, that is, payoff differences.[3] In this case, the EQ equilibrium of the original game with selfish players might be *sometimes* the only one that survives a process of iterated deletion of dominated strategies provided that the distribution of types includes a large enough share of types who dislike inequity. This is the case for equilibrium (R2, C2) in the game above if the probability that Row is sufficiently inequity averse is high enough: Since such a player would always play R2 whatever Column chooses, Column would find C2 a best response in expected terms if she believes that Row is very likely so inequity averse.

While the influence of payoff equity on coordination has been suggested by several theorists, the issue has not received much attention from experimental economists. Often, the analyzed coordination games are games with several equilibria where all players get the same payoff (Stag Hunt, Weak-Link, etc.), or games like the Battle of the Sexes with (pure-strategy) equilibria outcomes which are inequitable but just a permutation of each other. Payoff equity cannot facilitate coordination in these games because they do not have a unique equitable equilibrium. Yet one can think of a myriad of games where EQ makes a unique prediction and our experiments study how EQ performs in some of them.

This paper reports data from two experiments (1 and 2). In each, subjects play six different *asymmetric*, 2x2 normal-form games in a within-subjects design. To prevent confounds and assess the relative explanatory power of EQ, the games in both experiments permit separation between EQ and six alternative behavioral rules or ‘types’ (to be formally defined later). These rules include other equilibrium selection principles like *efficiency* and *risk dominance* (Harsanyi and Selten, 1988), but also ‘bounded rationality’ rules such as *maximum payoff*, *security* (or *maximin*), and two rules based on *level-k theory* (Stahl and Wilson, 1995). Most of them have received much attention in the experimental literature on coordination.

Our starting point is a relatively parsimonious model: Players are actually heterogeneous (e.g., some play as EQ predicts, but others are better characterized as maximin types) yet believe that others are of their type and hence are not aware of this heterogeneity.[4] We test this model using the classification procedure by El-Gamal and Grether (1995) and the data from Experiment 1. The results are in line with the idea that subjects are heterogeneous, although a significant share of them

[3] Influential articles like Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) have shown that this hypothesis can organize key experimental phenomena like cooperation in social dilemmas and costly punishment in the ultimatum game.
[4] Level-2 types are the only exception to this last proviso, as they expect the co-player to be a level-1 type (see details in Section 3).

can be best described as equity types. Indeed, we find that EQ and risk dominance (RD) explain significantly more choices than any other rule in Experiment 1, and there is no significant difference in the performance of EQ and RD. If we allow for heterogeneity, the best models with at least two types of players always include a substantial fraction of subjects acting most consistently with EQ. For instance, 35.8% of the subjects are best described by EQ in the optimal model with four types, while 31, 21.8, and 11.4% are respectively best described by RD and two non-equilibrium rules

In summary, the results from Experiment 1 show a good relative performance of EQ in a horse race against six well-known, *simple* rules, thus suggesting that equity is a relevant criterion for play in coordination games. Yet we must also note that our results show significant differences in the explanatory power of EQ across games: The relevance of EQ apparently depends on the payoff structure of the game. In particular, when equity clashes with payoff dominance, it seems that the latter principle triumphs, particularly if the payoff dominant equilibrium yields considerably higher payoffs to the players and is not very unequal. This is incompatible with our simple model in that players do not seem to play always the same simple rule, but change depending on the game. With a regression analysis we identify several game variables that could affect EQ play and explore this point further in Experiment 2.

Our results from Experiments 1 and 2 are consistent with the hypothesis that the frequency of EQ play depends negatively on the social efficiency of the alternative equilibrium (e_A) and positively on the unfairness (i.e., payoff difference) of e_A. The results from Experiment 2 (but not Experiment 1) also suggest that EQ play depends negatively on the degree of risk dominance of e_A. Furthermore, we investigate whether having strictly equal payoffs in the EQ equilibrium increases EQ play, motivated by earlier studies like Engelmann and Strobel (2004) and Güth et al. (2001) that point out how the existence of strictly equal payoffs might particularly reinforce inequality considerations. Yet our *preliminary* evidence suggests that *ceteris paribus* having strictly equal payoffs does not enhance EQ play. Finally, some primary evidence also hints that inequity aversion is not the principal reason behind the relevance of the EQ principle, at least if we posit that preferences for equity are stable across games.

The rest of the paper is organized as follows. The next section reviews some related experimental literature on coordination and discusses how our evidence contributes to the understanding of several results of this literature, whereas Section 3 presents in detail the EQ principle and the other rules that we consider in this paper. In turn, Section 4 describes and motivates Experiment 1 and analyzes the obtained data. Finally, Section 5 presents Experiment 2 and reports the corresponding data and results, while Section 6 concludes.

2. Related Literature

Our paper complements a large experimental literature on coordination that analyzes the performance of selection principles like payoff dominance, risk dominance, and security (Cooper et al., 1990; Van Huyck et al., 1990; Straub, 1995; and Schmidt et al., 2003). In this regard, our results indicate that subjects often fail to coordinate on the risk dominant equilibria or play the secure strategy when EQ predicts otherwise. We also complement a growing literature (Haruvy and Stahl, 2007; Crawford and Iriberri, 2007a) that considers as well theories of non-equilibrium behavior. In our simple games, most subjects do not act as these theories forecast if EQ makes a different prediction. Yet our classification analysis indicates that the behavior of a minority of subjects can be best explained by some of these theories. In this respect, our study is closely related to Haruvy and Stahl (2007), as both compare the relative performance of several models with heterogeneous agents. In line with our results, Haruvy and Stahl (2007) show that heterogeneous models can explain play in coordination games significantly better than models of homogeneous players. When comparing their results to ours, yet, one has to keep in mind that Haruvy and Stahl (2007) only consider symmetric games and that they study 3x3 games while our focus is on simpler 2x2 games.[5] In a related line, Costa-Gomes et al (2001) also investigate strategic sophistication, assuming different types of strategic and non-strategic players. Four of the 7 theories that we use here appear in their study as well. Their results suggest that a considerable share of players behaves according to a non-equilibrium prediction. However, they do not study the effect of payoff equity on coordination, our main goal in this paper.

Two papers have analyzed how payoff equity affects coordination. Chmura et al. (2005) investigate several 2x2 games with two equilibria. One of them is payoff dominant (or at least efficient, in the sense employed in our study), whereas the other gives the same payoff to both players. The authors find that coordination on the first, efficient equilibrium becomes more complicated when the distance in the players' payoffs in that equilibrium increases. Still, the majority of the subjects play the efficient equilibrium in most games. This is consistent with our finding that payoff dominance trumps equity when both principles clash. In contrast to Chmura et al. (2005), we also study games without a payoff dominant equilibrium, and moreover compare its relative performance to alternative theories like risk dominance and security. Our results suggest that EQ is the best predictor in games where payoff dominance and equity do not contradict each other.

[5] This seems important. Crawford and Iriberri (2007b) note that sophistication among subjects depends on the game played. In our games, the fraction of subjects who play in a boundedly rational manner might be smaller than in more complex 3x3 games. Further, the results from Ho and Weigelt (1996) in extensive-form games suggest that an equilibrium is played less often if it is more complex to compute than others, even if it has low payoff disparity.

On the other hand, Crawford et al. (2008) study how salient strategy labels help to achieve coordination, showing that the effect of labels is much reduced in games with even minutely asymmetric payoffs, and suggesting an explanation based on level-k thinking and "team reasoning". It is worth clarifying that Crawford et al. (2008) study games akin to the Battle of the Sexes without a unique equitable equilibrium, so that EQ has no bite then. Yet our paper complements and qualifies their claim in the title that "even minute payoff asymmetry may yield large coordination failures". In effect, EQ indicates how the existence of an equitable equilibrium (i.e., one where there are small or null payoff *asymmetries*) may facilitate coordination, and we provide evidence in this line.

Finally, our paper provides some evidence on how coordination depends on both inequality and social efficiency and it is thus related to several earlier papers that suggest a trade-off between fairness and efficiency (although in individual decision problems). For instance, Kritikos and Bolle (2001), Charness and Rabin (2002), and Engelmann and Strobel (2004) provide evidence from dictator games and claim that social welfare concerns can sometimes be stronger than equity ones.

3. Theories

In this section we present the EQ principle and six additional theories to explain behavior in one-shot coordination games. For simplicity, we focus on two-player, normal-form games, although extensions to n-player games are often direct. To facilitate the comparison of the different theories, we assume that all players are selfish and risk neutral; their utility equals their monetary payoff.[6] For expositional purposes, we keep using the names 'Row' and 'Column' to refer to the players in what follows.

Let E denote the set of equilibria in pure strategies of the game. We say that equilibrium e ∈ E is selected by the EQ selection principle if the difference between the highest and the smallest *monetary* payoff of e is minimal among all equilibria in E. To clarify matters, consider the games below, both with two equilibria in pure strategies: (R1, C1) and (R2, C2). EQ clearly predicts play of (R2, C2) in both.

	C1	**C2**
R1	75, 140	75, 75
R2	70, 70	110, 110

	C1	**C2**
R1	90, 100	0, 20
R2	20, 0	40, 40

[6] Since EQ selects the equilibrium with smallest differences in *monetary* payoffs, this point is immaterial for its description. As mentioned in the introduction, however, play of the EQ equilibrium could be *motivated* on the grounds that some individuals are inequity averse and find strictly dominant to choose the EQ strategy. Appendix B of the online supplement illustrates this point in more detail by using the model by Fehr and Schmidt and discussing the conditions required for a player to have a dominant strategy in the two games used for illustration in this section and others.

Besides the EQ principle, we consider two other equilibrium selection principles, defined as follows:

i. *(Social) efficiency (EF): selects the equilibrium that maximizes the sum of the players' payoffs.*

   This seems a natural extension of payoff dominance, which selects equilibrium e* if players receive strictly higher payoffs at e* than at any other equilibrium. We focus on efficiency because payoff dominance often selects no equilibrium (e.g., left-hand game above). Note anyway that the efficient equilibrium coincides with the payoff-dominant one if it exists. *Example:* efficiency selects equilibrium (R2, C2) in the left-hand game above and (R1, C1) in the right-hand game.

ii. *Risk dominance (RD): selects the equilibrium maximizing the product of players' losses from unilateral deviations.*[7]

   Example: (R2, C2) and (R1, C1) are risk-dominant in the left-hand and right-hand games above, respectively.

The remaining theories to be considered here relax the assumption of equilibrium play even in static games like ours.

iii. *Level-k: level-0 agents choose with uniform probability between their available strategies, while level-k agents best respond to a level-(k-1) strategy.*[8]

   Thus, level-1 agents play a best response to a level-0 strategy, level-2 agents play a best response to a level-1 strategy, and so on. Given previous evidence (see Haruvy et al., 1999), we only consider *level-1 (L1)* and *level-2 (L2)* agents.

   Example: level-1 predicts outcome (R2, C1) in the left-hand game above and (R1, C1) in the right-hand game; while level-2 predicts (R1, C2) in the left-hand game and (R1, C1) in the right-hand game.

iv. *Maximum payoff (MP): each player chooses the action that potentially gives her the highest monetary payoff in the game.*

   Haruvy et al. (1999) refer to this as the "optimistic" rule. An analogous "pessimistic" rule could be easily defined, but we abstain from that as Haruvy et al. (1999) and Costa-Gomes et al. (2001) found little evidence for it.

[7] This definition stands for 2x2 games with two pure-strategy equilibria, but it can be extended to n-player games (see for instance Haruvy and Stahl, 2007), although this is not necessary for our purposes. See also Harsanyi and Selten (1988, p. 355-357) for a motivation of payoff and risk dominance.

[8] See Stahl and Wilson (1995) for more details.

Example: Maximum payoff predicts outcome (R2, C1) in the left-hand game above and (R1, C1) in the right-hand game.

v. *Security (S) or maximin: players choose the strategy with the largest minimum payoff.*

Note that this behavioral rule leads to equilibrium play if all available strategies are part of at least one equilibrium and the games are symmetric (as in Van Huyck et al., 1990), but it may lead to a non-equilibrium outcome in asymmetric games like ours.

Example: The prediction of security in the left-hand and right-hand games above is respectively (R1, C2) and (R2, C2).

4. Experiment 1

4.1 Design and Procedures

A total of 126 subjects participated in this experiment, which consisted of 4 sessions run at the Universidad Autónoma de Madrid. No participant attended more than one session. Subjects were undergraduate students with different majors (Economics, Business Administration, or Psychology). They earned on average 7.7 Euros from their decisions, and there was no show-up fee. In every session, subjects played six 2x2 normal-form games with the following payoff matrices (payoffs are presented in points at the rate 1 point = 0.1 Euro; we discuss our choice of games later):

Game 1	C1	C2
R1	75, 140	75, 75
R2	70, 70	110, 110

Game 2	C1	C2
R1	120, 120	85, 85
R2	80, 80	90, 160

Game 3	C1	C2
R1	110, 100	0, 90
R2	90, 0	90, 90

Game 4	C1	C2
R1	120, 105	102, 102
R2	100, 110	115, 115

Game 5	C1	C2
R1	190, 95	80, 80
R2	120, 90	120, 100

Game 6	C1	C2
R1	90, 100	0, 20
R2	20, 0	40, 40

Each session was conducted as follows. Before it started, the instructions and the decision sheets (dependent on role) were randomly distributed in conveniently separated seats across the room so as to avoid communication between subjects. The sheets were initially covered and the subjects could freely choose their seat; in that way, subjects were randomly assigned either role 'A' (Row) or 'B' (Column) for the whole session. Subjects first read the instructions at their own pace and then the experimenter read them aloud so that common knowledge was ensured. Questions were privately clarified.[9]

All six games were presented on the same decision sheet, and subjects were allowed to make their decisions in the order they wanted and revisit their choices at any moment before the experiment ended. As noted in Haruvy and Stahl (2007), this feature theoretically ensures that each subject plays each game in the same conditions, thus preventing potential order effects. Moreover, it increases the robustness of our results, as subjects could make all their choices with the maximum information possible.[10] To control for a potential order effect caused by the listing of the games in the decision sheet, however, the ordering in which the games were presented in the decision sheet was different across sessions (see also Haruvy and Stahl, 2007).

Participants were never told their counterparts' choices in any game, to prevent repeated game effects. After the subjects had made their decisions in the six games, they were asked to answer a brief questionnaire and once everybody finished, their decision sheets were collected. Then one game was randomly selected for payment with the roll of a dice, each subject was anonymously and randomly matched with a subject of her opposite role, her payoff was computed for the payment-relevant game, and she was paid in private. All previous features of our design were common knowledge among the participants. The sessions –including paying the subjects individually– lasted about 60 minutes.

We chose our games to evaluate the extent to which the EQ principle can explain play in coordination games, and how it performs compared with the other theories presented in section 3. Our games offer evidence in this regard because they allow us to discriminate between EQ and the alternative theories. Table 1 shows the strategies predicted by each theory (including EQ) in each of our six games. As the reader can confirm, we have variability in the predictions of each pair of theories. For instance, efficiency and risk dominance predict different actions in games 2 and 3. More precisely, there is a total of 12 predictions (one for Row and Column in each game) and the mean of different predictions across any two theories is around 6, the lowest difference being 3 (MP vs. L1, L1 vs. RD, and L2 vs. RD) and the highest 9 (EF vs. S, and MP vs. S).[11] Observe that EQ differs in at least four predictions (out of 12) from any other theory; moreover, we further discriminate between EQ and the other theories in Experiment 2.

[9] The instructions for player A can be found in Appendix A of the online supplement. The instructions for player B are analogous and are available upon request.

[10] A potential issue when presenting all games at the same time is that subjects might wish to achieve consistency across games. While this factor might affect subjects' behavior, we nevertheless believe that this is immaterial for our research goal, which is evaluating the relative performance of EQ. In particular, it is not clear how a search for consistency could explain that a subject follows EQ and not another selection principle.

[11] For comparison, the lowest difference in predictions in Costa-Gomes et al. (2001) is 2 out of 18 games.

	Theory						
	EQ	**EF**	**RD**	**L1**	**L2**	**MP**	**S**
Game 1	(R2, C2)	(R2, C2)	(R2, C2)	(R2, C1)	(R1, C2)	(R2, C1)	(R1, C2)
Game 2	(R1, C1)	(R2, C2)	(R1, C1)	(R1, C2)	(R2, C1)	(R1, C2)	(R1, C2)
Game 3	(R2, C2)	(R1, C1)	(R2, C2)	(R2, C2)	(R2, C2)	(R1, C1)	(R2, C2)
Game 4	(R2, C2)	(R2, C2)	(R2, C2)	(R1, C2)	(R2, C1)	(R1, C2)	(R1, C1)
Game 5	(R2, C2)	(R1, C1)	(R1, C1)	(R1, C1)	(R1, C1)	(R1, C2)	(R2, C1)
Game 6	(R2, C2)	(R1, C1)	(R1, C1)	(R1, C1)	(R1, C1)	(R1, C1)	(R2, C2)

Table 1: Prediction of (Row, Column) play by each theory

We finish with some miscellaneous remarks on our choice of games. First, we have excluded from our design some classical games like the Battle of the Sexes, which do not easily permit discrimination between the theories. Yet note that games 3 and 6 are variations of the well-studied Stag Hunt game (see for instance Cooper *et al*., 1990). Observe also that our games are very simple 2x2 coordination games, thus helping subjects to understand the decision tasks and the underlying strategic considerations. We remark as well that the payoffs are also chosen so that the average payoff across all games is similar for the row and column player (around 90 points). Further, our choice of games prevents confounds with motivations like altruism (Andreoni and Miller, 2002; Charness and Rabin, 2002) or pure reciprocity (Rabin, 1993) because these motivations do not predict a *unique* equilibrium in our games. In fact, altruism predicts the same two equilibria as the standard, selfish model in all games: (R1, C1) and (R2, C2). In turn, Rabin (1993) predicts the same equilibria in all games except one (see proposition 3 in Rabin, 1993).[12]

4.2. Results

This section is divided into two parts. In 4.2.1, we present some aggregate results for each game and compare the aggregate performance of each theory across games and sessions. Part 4.2.2 is devoted to the study of individual behavior using a classification procedure, in order to further analyze the relative performance of the theories and control for heterogeneity among subjects.

4.2.1 Aggregate Results

We first present the frequency of each outcome in each game. For this, we compute the frequencies considering all possible matches between row and column players across sessions. Since 126 subjects participated in the experiment, this means that instead of 63 observations we use 63*63 = 3969 matches in each game. In this line, Table 2 presents the payoff matrices of the six games (in

[12] The only exception is game 4, where (R1, C2) and (R2, C2) are fairness equilibria, but (R1, C1) is not, at least if the players are sufficiently reciprocal (consult appendix A in Rabin, 1993).

bold the outcome predicted by the EQ principle), together with the relative frequency of each outcome (in parenthesis in the corresponding cell), using the data from all sessions in Experiment 1.

Game 1	C1	C2
R1	75, 140 *(5%)*	75, 75 *(17%)*
R2	70, 70 *(19%)*	**110, 110** ***(59%)***

Game 2	C1	C2
R1	**120, 120** **(68%)**	85, 85 *(18%)*
R2	80, 80 *(11%)*	90, 160 *(3%)*

Game 3	C1	C2
R1	110, 100 *(16%)*	0, 90 *(22%)*
R2	90, 0 *(27%)*	**90, 90** ***(35%)***

Game 4	C1	C2
R1	120, 105 *(8%)*	102, 102 *(29%)*
R2	100, 110 (13%)	**115, 115** ***(50%)***

Game 5	C1	C2
R1	190, 95 *(17%)*	80, 80 *(25%)*
R2	120, 90 *(23%)*	**120, 100** ***(35%)***

Game 6	C1	C2
R1	90, 100 *(39%)*	0, 20 *(24%)*
R2	20, 0 *(23%)*	**40, 40** **(14%)**

Table 2: Frequency of outcomes (in parenthesis) for all possible matches across sessions

The data suggests that EQ has a significant predictive power in our games, as most people play as predicted by EQ in all games except game 6. In effect, the frequency of the EQ outcome (in bold) is larger than 0.5 x 0.5 = 0.25 in games 1 to 5, which is only possible if the majority of subjects plays the strategies selected by EQ in those games. Further, the frequency of the EQ outcome is almost always significantly higher at the 5% level than the frequency of any other outcome in games 1 to 5 when the test of proportions is considered.[13] In game 6 the EQ outcome is significantly less frequent than the other outcomes.

While play of the EQ strategies is frequent in most of our games, the frequency of EQ play differs across games. This can be seen in Table 3, which presents the percentage of hits of each theory in each game –i.e., if one participant chooses in a game the strategy predicted by a given

[13] The exception is the (R2,C1) outcome in game 3 (p-value = 0.0849). When taking into account multiple testing using the Bonferroni correction, the outcomes (R1,C2) and (R2,C1) in game 5 (p-values: 0.1248 and 0.0537, respectively) are not significantly less frequent than the EQ outcome. This is the most stringent method to avoid type I errors. The test results with the Bonferroni corrections are to be found in Table C1 of the online supplement.

theory, we count that as a hit for that theory; the frequencies for the best-performing theories in each game are presented in bold.[14]

Game	EQ	EF	RD	L1	L2	MP	S
1	**76.98%** [a, b]	**76.98%**	**76.98%**	50.79%	49.21%	50.79%	49.21%
2	**82.54%** [a]	17.46%	**82.54%**	53.17%	46.83%	53.17%	53.17%
3	**59.52%** [c]	40.48%	**59.52%**	**59.52%**	**59.52%**	40.48%	**59.52%**
4	**71.43%** [b]	**71.43%**	**71.43%**	57.94%	42.06%	57.94%	28.57%
5	**59.52%** [c]	40.48%	40.48%	40.48%	40.48%	**50.79%**	**49.21%**
6	37.30% [d]	**62.70%**	**62.70%**	**62.70%**	**62.70%**	**62.70%**	37.30%
All	**64.55%**	51.59%	**65.61%**	54.10%	50.13%	52.65%	46.16%

Note: 756 observations overall (126 in each game). EQ frequencies with the same superindex (a, b, c, d) are not significantly different at the 5% level (test of proportions; see Table C3 in the online supplement for the exact p-values).[15]

Table 3: Frequency of hits of each theory, for each game and across all games

Succinctly speaking, we observe three groups of games depending on the frequency of EQ play. Games 1, 2, and 4 show the highest frequency (over 70%). In games 3 and 5 EQ play is still most frequent. Game 6 has the lowest level. Hence, the performance of EQ in a game seems to depend on the whole payoff constellation. In particular, the existence of a payoff dominant equilibrium different than the EQ one seems to deter coordination on the equitable equilibrium, as apparently occurs in games 3 and 6.[16] The results from game 6 also hint that coordination on the EQ equilibrium is hindered as the difference in payoffs between the two equilibria grows larger.

We now compare the performance of the EQ principle to that of other theories across games. In Table 3 we highlight the statistically best theory (or theories) in bold for each game –note that the percentage of hits coincides in some games for some theories; this is because these theories share predictions in those games. If in a given game the percentage corresponding to the best theory is not significantly higher than that of the next best theory, then both are highlighted in bold, e.g. in game 5. We can see that the behavior of the majority of the subjects is consistent with the best theory in each game, as the highest percentage of hits is significantly above 50 percent in all games except 3 and 5 (t-test corrected for multiple testing; p-values < 0.015 for games 1, 2, 4, and 6 and p-value = 0.112 for games 3 and 5; Table C4 in the online supplement reports exact results). The most

[14] The hits of each theory separately for row and column players are available in Table C2 in the online supplement.

[15] If the Bonferroni correction is used, there are not significant differences in the frequency of EQ play between game 4 and games 2, 3, and 5.

[16] The performance of EQ in game 5 might be also affected by the fact that equilibrium (R1, C1) is 'almost payoff dominant' in this game.

important observation on the performance of EQ is summarized as follows (see Table C5 in the online supplement for more details):

Result 1: *EQ and RD perform significantly better than any other theory in aggregate (t-test; $p < 0.0001$ always).* [17] *In games 1 to 4 EQ performs significantly better than any other theory that predicts a different outcome (t-test; $p \leq 0.05$). Overall, the performance of EQ and RD is not statistically different.*

In game 5 EQ is the best theory, although it shows no statistical difference with MP and S.[18] Game 5 is of particular interest because the predictions of EQ there differ from those by the other theories. We find reassuring that EQ performs rather well in this game compared to the alternatives, even though the EQ prediction implies asymmetric payoffs. This point suggests that the driving force behind the good performance of EQ is not just the symmetry of payoffs *per se*, a point that we analyze in more detail using the data from Experiment 2. With respect to game 6, EQ performs here significantly *worse* than the best theories (t-test; $p < 0.001$ always), again signaling that EQ tends to fail when it conflicts with payoff dominance. Finally, leaving aside game 6, it is noteworthy that the alternative theories to EQ perform best when they share predictions with EQ. Otherwise they are often unable to explain even 50 percent of the choices. All this seems an indication of the relevance of EQ.

For completeness, we can also compare the performance of those theories different than EQ. In this respect, we find that risk dominance outperforms any other theory on aggregate (t-test; $p < 0.001$ always; for more details see Table C6 in the online supplement). Yet it performs rather badly in game 5, where its predictions are different than those of EQ (it is very successful in game 6, though). In turn, security, level-2, and MP hardly explain half of the aggregate choices. The most successful of the non-equilibrium theories regarding hits is level-1 theory, but it does not perform significantly better than level-2 theory or maximum payoff (t-test, p-values in Table C7 of the online supplement).

In summary, our results suggest that EQ describes behavior in our games better than any other theory different than RD (especially if we leave game 6 aside). Yet, considerable uncertainties remain, both because an aggregate analysis might conceal a high level of heterogeneity, and because there is still much evidence that is left unexplained even by the EQ principle. Therefore, after

[17] Unless otherwise noted, all the differences reported in the rest of section 4.2.1 are calculated using the Bonferroni correction.

[18] When correcting for multiple testing using the Bonferroni method, in game 4 L1 and MP are not significantly less frequent anymore than EQ at the 5% level (p-value = 0.0632 in both cases), just like EF, RD, L1 and L2 in game 5 (p-value = 0.096 in each case).

understanding the aggregate tendencies in our data, we move now to thoroughly study individual behavior and subjects' heterogeneity.

4.2.2 Individual analysis

We offer here an analysis on the *individual level* of the motives behind the players' choices in our coordination games. For this we use the classification procedure from El-Gamal and Grether (1995).[19] This procedure circumvents the multicollinearity problems that would appear in a classical regression analysis if the motives were treated as independent variables and allows appropriate inferences even when testing all possible motives –no matter how similar their predictions are– at the same time. In the application of the procedure, the key concept is that of *behavioral rule*, defined as a vector of strategies for a player, specifying one strategy for each of our six games. We focus on seven behavioral rules based on the homonymous theories presented in section 3 (they are implicitly described in Table 1), as these are the theories that have attracted most of the attention in previous studies. Note that the behavioral rules are not always symmetric, so that we must sometimes distinguish between rules for Row and Column. For example, the level-1 rule for Row is (R2, R1, R2, R1, R1, R1), while the same rule for Column is (C1, C2, C2, C2, C1, C1).

We first consider the case with no heterogeneity, i.e., we posit that all subjects follow the same behavioral rule ***r*** in each game. Yet we also assume that each subject may commit an error and deviate from ***r*** in a game with some equal probability ε.[20] This probability can be estimated by maximum likelihood – just by finding the frequency of actual choices unexplained by ***r*** – and suggests how well ***r*** fits our data. Based on this process we can find the best single behavioral rule, which is the one with the minimal error. Table 4 below contains the results of the classification analysis; the first line of this table presents the results of the best single rule, together with its corresponding error rate ε, its log-likelihood (LL) value, and the penalized log-likelihood (PLL).[21] Unsurprisingly, the best single rule (RD) coincides with the rule with the highest number of aggregate hits (see Table 3).

---

[19] For a detailed description of the procedure, see Appendix D of the online supplement.

[20] The assumption that any subject trembles with the same probability in any game merits two clarifications. First, it seems realistic because (i) our games are rather simple and (ii) subjects were given the possibility to revisit their choices at any moment and hence no change of ε through time (due to learning effects) should be expected. Second, the assumption implies that the probability of error is independent across games. In other words, the probability of deviating from ***r*** at any game is not conditioned on how the subject behaved in any other game. This seems reasonable in our experiment because, due to reason (ii) above, no order effects due to learning are expected.

[21] The PLL follows El-Gamal and Grether (1995) and other studies (e.g., Shachat and Walker, 2004), introducing a penalty for the use of each extra rule which will help us to decide whether the increased complexity of a model improves substantially the understanding of the data. Appendix D of the online supplement explains in more detail the logic behind the penalty term.

# rules	Best Model	% of subjects per rule	ε	LL Value	PLL Value
1	RD	100%	0.344	-486.560	-489.199
2	RD; EQ	51.6%; 48.4%	0.267	-438.820	-531.435
3	EQ; RD; MP	39.3%; 36.5%; 24.2%	0.233	-410.238	-556.580
4	EQ; RD; MP; L2	35.8%; 31%; 21.8%; 11.4%	0.209	-387.542	-572.771
5	EQ; RD; MP; L2; L1	33.1%; 27.0%; 17.4%; 11.4%; 11.0%	0.194	-372.408	-588.393
6	EQ; RD; MP; L1; S; L2	27.4%; 26.5%; 16.7%; 9.9%; 9.8%; 9.7%	0.185	-362.249	-603.845
7	EQ; RD; MP; L1; EF; S; L2	26%; 22.1%; 13.8%; 9.9%; 9.9%; 9.8%; 8.7%	0.185	-362.249	-625.907

Table 4: Results of the classification analysis; across all sessions

To allow for heterogeneity, consider now any combination of our seven rules (a *model*). We can classify subjects by assigning each one to the rule that best fits her behavior among those present in the model.[22] Given this assignment, we can compute the corresponding error in an analogous manner as before, and hence find the model with two, three, etc. rules that best accounts for the data in our games. In this vein, Table 4 reports information for the best model with up to seven rules. We also indicate the proportion of subjects assigned to each respective rule (in the model with just one rule this is obviously 100%). We find that the best model with two rules is composed by EQ and RD: 48.4% of the subjects are here assigned to EQ. With three rules the optimal model includes EQ and RD plus MP. From this case on, EQ is the rule to which most subjects are assigned. Even with seven rules, more than a quarter of all subjects are best described by EQ. Our most important findings about heterogeneity are summed up as follows:

Result 2: *Subjects are heterogeneous, but the behavior of a large fraction of them is best explained by the EQ principle whatever the number of rules considered. Regarding the remaining subjects, they are best depicted either by risk dominance or by non-equilibrium theories like level-k or MP.*

As an implication of the fact that a large fraction of the subjects tend to play EQ, those subjects who follow EQ often get a larger monetary payoff than those who do not follow EQ. In fact, playing according to EQ yields the highest *ex post* payoffs (considering the actual play of the opponent players) in games 1-4. In games 5 and 6, however, we note that playing according to EQ

[22] If k rules fit best the behavior of one subject, we assign 1/k 'subjects' to any such rule. Note further that any such model implicitly assumes for tractability that players are actually heterogeneous (i.e., play different rules), but are *not* aware of this heterogeneity. In addition, the models assume that any player always follows the same rule in all games. These two assumptions simplify the analysis but are yet capable to explain a large part of the evidence.

does not pay off. In game 5, both row and column players actually lose a small amount of money (0.25 and 3.5 points, respectively, that is, some cents) if they play EQ, while in game 6 the losses are more heavy, amounting on average to 28 and 36 points for the row and column players (which are equivalent to 2.8 and 3.6 Euros, respectively).

Although subjects are clearly heterogeneous, we have to examine whether the increased accuracy of heterogeneous models is worth their increased complexity. Naturally, the models with more rules have lower error rates, but they are also less parsimonious, and we would like to achieve a compromise in this respect. Although there are many criteria to judge whether models with additional rules improve substantially our understanding of the data, one possibility is to introduce a penalty for the use of each extra rule; this is what the PLL does. According to this criterion, the introduction of an additional rule in a model is recommended only if the PLL value grows with the addition of that rule. Our data hence suggests that one rule is enough to provide a parsimonious depiction of the subjects' behavior, pointing thus to the importance of RD in our games (yet note that a model with EQ alone has also a very low error rate of 0.354 and according to the overall hits presented in Table 3 the difference between them is not significant).

To infer whether heterogeneous models substantially improve the error rate, we also use maximum-likelihood (ML) tests to compare the best models with 1 and 2 rules, 2 and 3 rules, and so on up to 7 rules as all of them are respectively nested.[23] Each of these tests indicates whether the net improvement obtained with the more complex model is significant. In this respect, we observe that the best model is the one with five rules, as adding rules sequentially to the models and comparing them pairwisely, we reject the model with six rules in favor of that with five (degrees of freedom = 13; statistic value = 20.3; $p > 0.05$).

Result 3: *If parsimony is our main goal, a model assuming that all players follow the RD rule is optimal – note yet that an 'EQ-only' model has a very similar error rate. In terms of accuracy, the models with five, six, or seven rules are not significantly different. The simplest among the most accurate models has therefore five rules (EQ; RD; MP; L2; L1); and according to it, a third of all subjects are best described by EQ.*

---

[23] We say that the optimal model with k rules is nested if it is a restriction of the model with k+1 rules. Let $\lambda = L_R/L_U$ denote the likelihood ratio, where L_R and L_U are respectively the values of the likelihood function for the restricted model and the more complex model. Since the statistic $-2\cdot Ln(\lambda)$ is asymptotically distributed as a chi-squared with degrees of freedom equal to the number of restrictions imposed, we reject the restricted model if $-2\cdot Ln(\lambda)$ is very large, as this indicates in turn very small values of λ. For a final clarification, note that in a model with k rules we must determine for each subject the rule that he/she follows and those that he/she does not follow. This means k parameters per subject, taking value zero (the subject does not follow the corresponding rule) or 1 (the subject follows the rule). Let then $\boldsymbol{r}^*$ denote the rule not present in the model with k rules and n* the number of subjects who are assigned rule $\boldsymbol{r}^*$ in the model with k+1 rules. As we move from the optimal model with k+1 rules to that with k rules, we impose n* restrictions, as the parameters corresponding to rule $\boldsymbol{r}^*$ are restricted to take value zero for these n* subjects.

We have also performed the classification analysis with the data from games 1, 2, 4, and 5. This is interesting in order to further test the idea that EQ explains behavior better if it does not clash with payoff dominance. Table C8 in the online supplement indicates the best models with 1, 2, 3, and 4 rules in the mentioned games and across all sessions, using the above described classification analysis, and provides some evidence consistent with the cited idea: (i) whatever the number of rules, EQ is always included in the best model and the (relative) majority of the subjects is assigned to it; (ii) the error rate of the best model in games 1, 2, 4, and 5 is always smaller than the error rate of the same model in games 1 to 6; (iii) we observe that the best model with 4 or more rules explains more than 90 percent of the subjects' choices in these games.

Result 4: *EQ explains individual behavior particularly well when we focus on those games where EQ and payoff dominance do not conflict (games 1, 2, 4, and 5).*

Finally, we have also performed a regression analysis that investigates several potential determinants of compliance with EQ (such as the subject's role, the order of the games on the decision sheet, demographic differences, or the subject's degree of risk aversion based on Holt and Laury, 2002). Our dependent variable is the number of times (*hits*) that a subject chooses according to the EQ principle in all our games. We have run two types of regressions for this count data: OLS and negative binomial.[24] The marginal effects for two different specifications of the negative binomial model and the results of the OLS model are shown in Table C9 in the online supplement. We find that demographic characteristics are not significant, and with respect to potential effects of the subject's role or the order of display of the games in the decision sheet, the regressions reassuringly show no systematic effect on EQ play. In both regressions we find a highly significant effect of risk aversion on the play of EQ, a point which we will discuss later –we anticipate however that this effect turns out not to be robust, as it is not replicated with the data of Experiment 2. The rest of the variables are not significant in any regression.

5. Disentangling some uncertainties: Experiment 2

The previous results indicate that EQ helps to organize our data, at least under the assumption that players follow any of the simple seven rules described in Section 3. At the same time, however, the varying performance of EQ across games signals that the payoff structure of the game matters. That is, our prior assumption that subjects determine whether some equilibrium is focal or prominent –in the sense of Schelling, 1960– based on a single criterion (e.g., equity) is wrong. For instance,

[24] There is no over-dispersion in our data, so that the negative binomial model yields quantitatively the same results as the Poisson model.

equity or fairness might be a relevant criterion, but also payoff dominance, risk dominance, and social efficiency. When some of these criteria conflict, however, subjects may have to trade off between the relevant criteria, and evaluate the attractiveness of each equilibrium. In a game, for instance, some equilibrium F may be fair, and another E socially efficient. If players find both criteria relevant for equilibrium selection, they may solve their dilemma by judging how fair and socially efficient each equilibrium is: If equilibrium F is not 'too' inefficient while equilibrium E is rather unfair or inequitable, they may opt for F. A potential complication here is that players possibly differ in how sensitive they are to each criterion, that is, in their 'switching points'. To follow with our previous example, a player may consider that equilibrium E is focal if it is a "bit" more efficient than F, even if E is very unfair. If he moreover believes (maybe inaccurately) that other players share his perceptions on what makes an equilibrium prominent, the efficient equilibrium may be the logical option for him. Yet others may find that only a large difference in efficiency makes the efficient equilibrium focal. An obvious implication of the previous discussion is that the frequency of play of a fair, EQ equilibrium will depend on the payoff structure.

To pursue and investigate this line of reasoning, we suggest three potential measures of the attractiveness of the EQ equilibrium of a game, all of them well defined in any 2x2 normal-form game with two equilibria in pure strategies: (I) DDP = difference of payoffs in the non-EQ equilibrium minus difference of payoffs in the EQ equilibrium, (II) DSE = difference in social efficiency between the non-EQ and the EQ equilibrium, and (III) DPU = difference in the products of payoffs from unilateral deviations between the non-EQ and the EQ equilibrium. Each measure respectively corresponds to one of the following natural selection criteria: (I) fairness, (II) social efficiency, and (III) risk dominance. Table 5 illustrates these variables with the games of Experiment 1. For example, in Game 1 DDP is computed as $(140-75)-(110-110)=65$. DSE is simply $(75+140)-(110+110)=-5$, while DPU is calculated as $(75-70)\cdot(140-75)-(110-75)\cdot(110-70)=-1075$.

	DDP	**DSE**	**DPU**
Game 1	65	-5	-1075
Game 2	70	10	-1000
Game 3	10	30	-7900
Game 4	15	-5	-5
Game 5	75	65	650

Game 6	10	110	4000

DSE = Difference in social efficiency between the non-EQ and the EQ equilibrium; DPU = difference in the products of payoffs from unilateral deviations between the non-EQ and the EQ equilibrium; DDP = difference in the difference of payoffs between the non-EQ and the EQ equilibrium.

Table 5: Attractiveness of the EQ equilibrium of each game in Experiment 1, according to three criteria

The question here is which of these three variables significantly affects the frequency of EQ play in a game. To obtain some hints in this respect, we run several OLS and probit regressions where the independent variable is always a dummy variable taking value 1 if EQ was played in a game of Experiment 1 by a subject. The dependent variables are some or all of the three variables that appear in Table 5. Table 6 reports OLS regression results; probit results are qualitatively similar.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS
DSE	-0.00327***			-0.00300***
	(0.000458)			(0.000564)
DPU		-1.36e-05***		9.50e-07
		(4.01e-06)		(5.36e-06)
DDP			0.00284***	0.00179***
			(0.000617)	(0.000578)
Constant	0.757***	0.633***	0.529***	0.676***
	(0.0230)	(0.0181)	(0.0317)	(0.0377)
Observations	756	756	756	756
R-squared	0.081	0.010	0.030	0.093

Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 6: Determinants of EQ play in the games of Experiment 1

Note that each variable has explanatory power when used as a sole regressor (columns 1-3). When using all the variables in the same regression (column 4), however, the coefficient of DPU is not significant. This, together with the signs of the estimated coefficients of the other variables, hints at a hypothesis which we aim to explore further with the games of Experiment 2:

H1: The frequency of EQ play depends positively on DDP and negatively on DSE.

More precisely, the games considered in Experiment 2 are presented below (with the prediction of EQ in bold):

Game 1	C1	C2
R1	20, 10	120, 100
R2	**50, 40**	10, 20

Game 2	C1	C2
R1	10, 30	**70, 70**
R2	80, 60	20, 5

Game 3	C1	C2
R1	85, 55	20, 5
R2	10, 30	**75, 65**

Game 4	C1	C2
R1	10, 20	**50, 50**
R2	70, 55	20, 10

Game 5	C1	C2
R1	50, 15	70, 40
R2	**75, 75**	15, 50

Game 6	C1	C2
R1	60, 90	40, 10
R2	10, 40	**65, 50**

In these games, H1 makes the following predictions:

Prediction 1: The frequency of EQ play across Games 1 to 6 in Experiment 2 satisfies the following ranking: Game 5 > Game 3 ≈ Game 2 > Game 4 > Game 6 > Game 1.

This prediction is based on Table 7 below, which lists the DDP and DSE values of each game in the homonym columns (the games are ranked as in prediction 1) and indicates variable DPU as well for further clarification. Observe that Games 2 and 3 are identical in variables DSE and DDP.

	DDP	DSE	DPU
Game 5	30	-40	750
Game 3	20	0	1825
Game 2	20	0	1850
Game 4	15	25	1800
Game 6	15	35	3750
Game 1	10	130	9300

Table 7: Attractiveness of the EQ equilibrium of each game in Treatment 2

The previous discussion on the frequency of EQ play across games should not obscure one important goal of this paper, which is ascertaining the relevance of EQ as a selection principle in comparison with the other theories presented in Section 3. In this respect, our games have been designed so that risk dominance, security, L1, and L2 predict *always* differently than EQ.[25] Moreover, efficiency and EQ also predict differently in games 1, 4, and 6 –observe also that in games 2 and 3 both equilibria are socially efficient, and that in these games there is no payoff dominant equilibrium. Furthermore, in games 1 and 4 there is also no confound between EQ and the

[25] In game 5, the L1 rule is not defined for the column player (both strategies give the same expected payoff) and henceforth L2 is not defined for Row.

maximum payoff rule. Table 8 summarizes the predictions by each rule. In short, play of the EQ equilibrium in our games seems strong evidence in favor of the EQ selection principle.

	Theory						
	EQ	**EF**	**RD**	**L1**	**L2**	**MP**	**S**
Game 1	(R2, C1)	(R1, C2)	(R1, C2)	(R1, C2)	(R1, C2)	(R1, C2)	(R1, C2)
Game 2	(R1, C2)	(R1, C2), (R2, C1)	(R2, C1)	(R2, C1)	(R2, C1)	(R2, C2)	(R2, C1)
Game 3	(R2, C2)	(R1, C2), (R2, C1)	(R1, C1)	(R1, C1)	(R1, C1)	(R1, C2)	(R1, C1)
Game 4	(R1, C2)	(R2, C1)	(R2, C1)	(R2, C1)	(R2, C1)	(R2, C1)	(R2, C1)
Game 5	(R2, C1)	(R2, C1)	(R1, C2)	(R1, NA)	(NA, C1)	(R2, C1)	(R1, C2)
Game 6	(R2, C2)	(R1, C1)	(R1, C1)	(R1, C1)	(R1, C1)	(R2, C1)	(R1, C1)

Table 8: Prediction of (Row, Column) play by each theory in Experiment 2

Our games permit as well discrimination between EQ and a much narrower criterion which selects equilibria where payoffs are *exactly* equal, as some players may find a fair equilibrium focal only when it involves perfect equality, but not otherwise. If this was true, one should expect the frequency of EQ play to be diminished when the EQ equilibrium gives unequal payoffs to the players. In this respect, however, one must be careful to control for other potential variables that might affect the frequency of EQ play, like the ones discussed before. Hence, we get a more complex hypothesis qualifying H1 above:

H2: Other things equal (i.e., DSE, DPU, DDP), EQ play is significantly higher if the EQ equilibrium involves equal payoffs: Perfect equality improves coordination significantly.

Since Games 2 and 3 are equal or arguably very similar with respect to DDP, DSE, and DPU, and the EQ equilibrium gives equal payoffs to the players in Game 2 but not in Game 3, we conclude from H2 that:

Prediction 2: The frequency of EQ play in game 2 is significantly higher than in game 3.

Before discussing the results from Experiment 2, we remark that experimental procedures were *identical* to those in Experiment 1. A total of 94 subjects participated in this second experiment, which consisted of 5 sessions run in September and October of 2014 at the Universidad Autónoma de Madrid. As in Experiment 1, the order of display of the six games in the decision sheets differed across sessions. Subjects earned on average 7.81 Euros from their decisions.

5.1 Experimental Results

Table 9 presents the payoff matrices of our six games (in bold the outcome predicted by the EQ principle), together with the relative frequency of each outcome (in parenthesis in the corresponding cell), using the data from all sessions in Experiment 2. We observe substantial differences in the frequency of EQ play across games. For instance, the frequency of the EQ outcome is larger than 25% in games 2 and 5, which means that most subjects play the strategies selected by EQ in those games. In contrast, only a minority of subjects played according to EQ in the rest of the games. As we will discuss, these variations are basically in line with Hypothesis 1 above.

Game 1	**C1**	**C2**
R1	20, 10 *(5%)*	120, 100 *(88%)*
R2	**50, 40** ***(0%)***	10, 20 *(7%)*

Game 2	**C1**	**C2**
R1	10, 30 (29%)	**70, 70** ***(29%)***
R2	80, 60 *(21%)*	20, 5 *(21%)*

Game 3	**C1**	**C2**
R1	85, 55 *(31%)*	20, 5 *(16%)*
R2	10, 30 *(35%)*	**75, 65** ***(18%)***

Game 4	**C1**	**C2**
R1	10, 20 *(15%)*	**50, 50** ***(9%)***
R2	70, 55 *(48%)*	20, 10 *(28%)*

Game 5	**C1**	**C2**
R1	50, 15 *(29%)*	70, 40 *(19%)*
R2	**75, 75** ***(31%)***	15, 50 *(21%)*

Game 6	**C1**	**C2**
R1	60, 90 *(74%)*	40, 10 *(14%)*
R2	10, 40 *(10%)*	**65, 50** ***(2%)***

Table 9: Frequency of outcomes (in parenthesis) for all possible matches across sessions in Experiment 2

Table 10 below is the correlate of Table 3 and presents the percentage of hits of each theory in each game, across all sessions of Experiment 2.

Game	EQ	EF	RD	L1	L2	MP	S
1	6.38% [c]	93.62%	93.62%	93.62%	93.62%	93.62%	93.62%
2	54.26% [a]	100%	45.74%	45.74%	45.74%	45.74%	45.74%
3	45.74% [a]	100%	54.26%	54.26%	54.26%	41.49%	54.26%
4	28.72% [b]	71.28%	71.28%	71.28%	71.28%	71.28%	71.28%
5	55.32% [a]	55.32%	44.68%	28.72%	24.47%	55.32%	44.68%
6	13.83% [c]	86.17%	86.17%	86.17%	86.17%	41.49%	86.17%
All	34.04%	76.60%	65.96%	63.30%	62.59%	58.16%	65.96%

Note: 564 observations overall (94 in each game). EQ frequencies with the same superindex (a, b, c) are not significantly different at the 5% level (two-sided test of proportions; see Table C10 of the online supplement for statistical results).[26] Recall that in games 2 and 3 both equilibria are socially efficient, thus explaining why any choice is consistent with EF in those games.

Table 10: Frequency of hits of each theory, for each game and across all games in Experiment 2

Result 5: *The frequency of EQ play differs across games, with the following ranking in order of significance: Game 5 ≈ Game 2 ≈ Game 3 > Game 4 > Game 6 > Game 1 (one-sided t-test, p <0.046 always for inequalities; see Table C11 in the online supplement for details and for p-values using Bonferroni correction).*

Noticeably, the actual ranking coincides rather well with that in Prediction 1; the only exception is that H1 predicted game 5 > games 2 and 3, while these differences do not turn out to be significant in our data (one-sided t-test, p-values = 0.4421 and 0.0956, respectively). Hence, Hypothesis 1 does not seem to fare badly. We can comment, however, on other alternative hypotheses, particularly the simplest ones. First, the difference between games 4 and 6 points against the assumption that only the variable DDP affects (positively) EQ play. Regarding DPU, second, the values suggest that there should not be significant difference between Games 2, 3, and 4 (and if any, in favor of Game 4). However the findings show that in Game 4 EQ is worse than in the other two games. Finally, the hypothesis that only DSE matters predicts the same ranking as Prediction 1 and hence is also roughly in line with Result 5 above. Further research should discriminate more accurately between H1 and the 'DSE-only' hypothesis.[27]

To test Prediction 2, we compare the frequency of EQ play in Games 2 and 3 of Experiment 2. The results of the t-test indicate that we cannot reject the null hypothesis that they are equal (p-value = 0.2455). This suggests that strictly equal payoffs have only a diminished role in fostering EQ. Since this statement comes from a single comparison, however, it should be taken with some caution. Further research should investigate how the existence of equal payoffs affects EQ play, *other factors constant*.

A comparison of the data from Experiments 1 and 2 sheds some more light on H1. In effect, we observe that the frequency of EQ play is overall smaller in the games of Experiment 2 than in those of Experiment 1 (chi-square test, p < 0.0001). This accords well with H1, as some games in Experiment 1 dominate several others in Experiment 2 (i.e., have lower values of DSE and larger of

[26] When correcting for multiple testing using the Bonferroni method, the difference between games 3 and 4; 4 and 6 ceases to be significant.
[27] This latter hypothesis is somehow at odds with our data from Experiment 1, where game 5 exhibits a large DSE but also a rather substantial frequency of EQ play

DDP). For instance, Game 1 of Experiment 1 dominates any game of Experiment 2 except Game 5, and Game 2 of Experiment 1 dominates Games 1, 4, and 6 of Experiment 2. Further evidence comes when we apply the classification analysis of Section 4.2.2 to our games in Experiment 2, as we observe that the role of EQ is much diminished, clearly a signal that most games there are not conducive to EQ play (as H1 suggests).[28]

# of rules	Best Model	% of subjects per rule	ε	LL Value	PLL Value
1	EF	100%	0.323	-354.69	-357.33
2	EF&RD	54.3%; 45.7%	0.215	-293.22	-363.66
3	EF&RD&L2	46.3%; 41.5%; 12.2%	0.202	-283.88	-395.06
4	RD&EF&MP&EQ	42.5%; 30.9%; 15.2%; 11.4%	0.176	-262.00	-402.87
5	RD&EF&MP&EQ&L2	38.8%; 26.4%; 13.7%; 10.8%; 10.3%	0.163	-250.87	-415.35
6	RD&EF&MP&L1&EQ&L2	27.8%; 25.9%; 13.5%; 11.9%; 10.6%; 10.3%	0.163	-250.87	-435.13
6	EF&RD&S&MP&EQ&L2	25.9%; 19.9%; 19.9%; 13.5%; 10.6%; 10.3%	0.163	-250.87	-435.13
7	EF&RD&S&MP&EQ&L2&L1	19.1%; 12%; 12%; 10%; 7.9%; 7.7%; 6%	0.163	-250.87	-452.26

Table 11: Results of the classification analysis in Experiment 2

A comparison across the two experiments also provides some hints regarding one possible explanatory cause behind the EQ principle (i.e., inequity aversion, as explained in the introduction). In effect, the non-EQ strategies in many of the games in Experiment 1 are dominated under the assumption that some types have reasonable levels of inequity aversion (taking into account the distribution suggested by Fehr and Schmidt, 1999; see Appendix B of the online supplement). In the games of Experiment 2, in contrast, dominance requires extreme levels of inequity aversion or is simply impossible (this is formally proved in Appendix B of the online supplement). While inequity aversion does not make a unique prediction of play in these games, one may assume therefore that if inequity aversion explains coordination on the EQ equilibrium then its role as a coordinating device should be hindered in the games of Experiment 2. That is, the frequency of EQ play should be *consistently* lower in this set of games. Yet this is not what we observe. For instance, if we compare EQ play in game 5 across experiments, then we cannot reject the null hypothesis that EQ play is equal in both treatments (chi-square test, $p = 0.532$). This comparison is relevant because the non-EQ strategies are dominated in game 5 of Experiment 1 even under the assumption that the types show rather low levels of inequity aversion, while they are undominated in the homonym game of Experiment 2 whatever the distribution of inequity averse types. Still, we insist that a test of inequity aversion using the games of Experiment 2 is not very clean because there are no dominated strategies

[28] Note that in some games of Experiment 2 some rules do not yield unambiguous predictions (e.g., efficiency in games 2 and 3). To deal with that problem in the classification analysis we assigned 0.5 points in these cases. For example, whatever a player chose in game 2, it augmented the efficiency score by 0.5 point, a compromise between assigning 0 or 1 point. Note also that Risk Dominance (RD) and Security (S) share predictions in all games, so that any model where RD is present but S is not performs as well as an identical model where RD is substituted by S.

in them and hence two equilibria still remain even if players are (reasonably) inequity averse. Although with many caveats, therefore, the evidence does not seem much in favor of an inequity aversion argument: The role of payoff equality in improving (sometimes) coordination is not apparently caused by agents' dislike of inequity.

As for Experiment 1, finally, we have performed a regression analysis that investigates several other potential determinants of compliance with EQ (such as the subject's role, the order of the games on the decision sheet, demographic differences, or the subject's degree of risk aversion based on Holt and Laury, 2002). Again, our dependent variable is the number of times (*hits*) that one subject chooses according to the EQ principle in all our games. We have run two types of regressions for this count data: OLS and negative binomial. We find that risk aversion has a significantly negative coefficient, so the more risk averse a player is the less likely is that she/he plays EQ –recall that we got the opposite effect in Experiment 1. In turn, our variable 'selfish', taking value 1 if the subject chooses always the allocation maximizing her own payoff in 4 hypothetical games that appeared in the Questionnaire, has also a significantly negative coefficient, so subjects who care more about own payoffs tend to play less EQ. Both findings appear to be in line with intuition because playing EQ in most games of Experiment 2 was hardly conducive to the maximal earnings (given how most others acted) or is a rather risky strategy if one wants to achieve large payments.[29] Furthermore, there is some evidence that men were more likely to play EQ, though here the coefficient is barely significant. Although age was not very much dispersed (it ranges from 17 to 31), there is a significant positive relationship in our data between age and playing EQ. Finally, a player's role (row vs. column), her/his ideology, exposure to game theory, the session in which she/he played or the order of presentation of the games in the decision sheet do not have an effect on playing EQ.[30]

7. Conclusion

We suggest a criterion for equilibrium selection, based on the idea that the existence of a unique equitable equilibrium facilitates coordination. In a horse race with another six simple rules, most of them well studied in the literature on coordination, EQ outperforms any other rule except risk dominance. Indeed, EQ explains the behavior of a substantial share of the subjects, even if we allow for heterogeneity. In this respect and although parsimony recommends otherwise, the assumption that players are heterogeneous is advisable in order to obtain a more accurate explanation of our results. In addition, we also observe that the explanatory power of EQ is limited in some

[29] We are very grateful to an anonymous referee for pointing out this explanation.

[30] For these results we have only 81 observations because some players did not answer some of the questions in the questionnaire and they were eliminated from the regressions. The detailed statistical results can be found in Table C12 of the online supplement.

games, particularly in those where there is an alternative equilibrium that is payoff dominant. In this line, we get some evidence consistent with the hypothesis that the frequency of EQ play depends negatively (positively) on how socially efficient (unfair) the alternative equilibrium is. In turn, some preliminary evidence suggests that strict equality of payoffs at the EQ equilibrium does not increase EQ play, *ceteris paribus*. In any case, the picture revealed by our data is a complex one, as players appear to be heterogeneous on what factors they consider when choosing a strategy, and possibly these factors change from game to game.

References

- Andreoni, J., Miller, J., 2002. Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica* 70, 737-753.
- Blume, A., Ortmann, A., 2007. The Effects of Costless Pre-Play Communication: Experimental Evidence from Games with Pareto-Ranked Equilibria. *Journal of Economic Theory* 132, 274-290.
- Bolton, G. E., Ockenfels, A., 2000. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review* 90(1), 166-93.
- Brandts, J., Cooper, D. J., 2006. A Change Would Do You Good .... An Experimental Study on How to Overcome Coordination Failure in Organizations. *American Economic Review* 96(3), 669-693.
- Camerer, C., 2003. *Behavioral Game Theory-Experiments in Strategic Interaction*, Princeton University Press.
- Charness, G., Rabin, M., 2002. Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics* 117, 817-869.
- Chmura, T., Kube, S., Pitz, T., Puppe, C., 2005. Testing (Beliefs about) Social Preferences: Evidence from an Experimental Coordination Game. *Economics Letters* 88, 214-220
- Cooper, R., DeJong, D., Forsythe, R., Ross, T., 1989. Communication in the Battle of the Sexes Game: Some Experimental Results. *RAND Journal of Economics* 20(4), 568-587.
- Cooper, R., D DeJong, D., Forsythe, R., Ross, T., 1990. Selection Criteria in Coordination Games: Some Experimental Results. *American Economic Review* 80, 218-233.
- Cooper, R., DeJong, D., Forsythe, R., Ross, T., 1992. Communication in Coordination Games. *Quarterly Journal of Economics* 107, 739-771.
- Costa-Gomes, M., Crawford, V. P., Broseta, B., 2001. Cognition and Behavior in Normal-Form Games: An Experimental Study. *Econometrica* 69(5), 1193-1235.
- Crawford, V. P., 1997. Theory and Experiment in the Analysis of Strategic Interaction. In: Kreps, D., Wallis, K. (Eds.). *Advances in Economics and Econometrics: Theory and Applications*, Vol. I. Cambridge: Cambridge University Press.
- Crawford, V. P., Iriberri, N., 2007a. Level-*k* Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions? *Econometrica* 75(6), 1721-1770.

- Crawford, V. P., Iriberri, N., 2007b. Fatal Attraction: Salience, Naiveté, and Sophistication in Experimental "Hide-and-Seek" Games". *American Economic Review* 97(5), 1731-1750.
- Crawford V. P., Gneezy, U., Rottenstreicher, Y., 2008. The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures. *American Economic Review* 98 (4), 1443-1458.
- Devetag. G., Ortmann, A., 2007. When and Why? A Critical Survey on Coordination Failure in the Laboratory. *Experimental Economics* 10, 331-344.
- El-Gamal, M., Grether, D., 1995. Are People Bayesian? Uncovering Behavioral Strategies. *Journal of the American Statistical Association* 90(432), 1137-1145.
- Engelmann, D., Strobel, M., 2004.Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review* 94(4), 857-869.
- Fehr, E., Schmidt, K., 1999. A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics* 114(3), 817-68.
- Güth, W., Huck, S., Müller, W., 2001. The relevance of equal splits in ultimatum games. *Games and Economic Behavior* 37(1), 161-169.
- Harsanyi, J. C., Selten, R., 1988. *A General Theory of Equilibrium Selection in Games*. MIT Press.
- Haruvy, E., Stahl, D. O., Wilson, P., 1999. Evidence for Optimistic and Pessimistic Behavior in Normal-Form Games. *Economic Letters* 63, 255-260.
- Haruvy, E., Stahl, D. O., 2007. Equilibrium Selection and Bounded Rationality in Symmetric Normal-Form Games. *Journal of Economic Behavior and Organization* 62, 98-119.
- Ho, T. H., Weigelt, K., 1996. Task Complexity, Equilibrium Selection and Learning: An Experimental Study. *Management Science* 42, 659-679.
- Holt, C.A., Laury, K., 2002. Risk Aversion and Incentive Effects. *American Economic Review* 92, 1644-1655.
- Kritikos, A., Bolle, F., 2001. Distributional concerns: equity-or efficiency-oriented? *Economics Letters* 73(3), 333-338.
- Myerson, R. B., 1997. *Game Theory: Analysis of Conflict*. Cambridge: Harvard University Press (3rd edition).
- Nagel, R., 1995. Unraveling in Guessing Games: An Experimental Study. *American Economic Review* 85, 1313-1326.
- Palfrey, T.R., Rosenthal, H., 1994. Repeated Play, Cooperation and Coordination: An Experimental Study. *Review of Economic Studies* 61, 545-565.

- Rabin, M., 1993. Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83, 1281-1302.
- Schelling, T., 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.
- Schmidt, D., Shupp, R., Walker, J. M., Ostrom, E., 2003. Playing Safe in Coordination Games: The Role of Risk Dominance, Payoff Dominance, Social History, and Reputation. *Games and Economic Behavior* 42, 281-299.
- Shachat. J., Walker, M., 2004. Unobserved Heterogeneity and Equilibrium: An Experimental Study of Bayesian and Adaptive Learning in Normal Form Games. *Journal of Economic Theory* 114, 280-309
- Stahl, D., Wilson, P., 1995. On Players' Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior* 10, 213-254.
- Straub, P. G., 1995. Risk Dominance and Coordination Failures in Static Games. *The Quarterly Review of Economics and Finance* 35 (4), 339-363.
- Van Huyck, J. B., Battalio, R., Beil, R., 1990. Tacit Coordination Games, Strategic Uncertainty, and Coordination Failures. *American Economic Review* 80, 234-248.
- Van Huyck, J. B., Gillette, A., Battalio, R., 1992. Credible Assignments in Coordination Games. *Games and Economic Behavior* 4, 606-626.