ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

GAME HARMONY AS A PREDICTOR OF COOPERATION IN 2 X 2 GAMES

Daniel John Zizzo and Jonathan H.W. Tan

Number 151

March 2003

Manor Road Building, Oxford OX1 3UQ

Game Harmony as a Predictor of Cooperation in 2×2 Games

Daniel John Zizzo*
Department of Economics and Christ Church College
University of Oxford

Jonathan H.W. Tan
Lincoln College
University of Oxford

(Version March 8, 2003)

Abstract
This paper presents an experimental test of the relationship between *game harmony* and cooperation in 2×2 games. Game harmony measures describe how harmonious (non-conflictual) or disharmonious (conflictual) the interests of players are, as embodied in the payoffs: coordination games and constant-sum games are examples of games of perfect harmony and disharmony, respectively; most games are somewhere in the middle. Correlation coefficients between measured harmony and cooperation rates were above 0.8 in an experiment with games such as the Prisoner's Dilemma, the Stag-Hunt, a coordination game and three trust games, and above 0.55 in an experiment with random games.
Keywords: game harmony, cooperation, 2×2 games, trust games, social dilemmas.
JEL Classification Codes: C72, C91, H41.

* Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford OX1 3UQ, U.K. (email: daniel.zizzo@economics.ox.ac.uk; jonathan.tan@economics.ox.ac.uk). Please address all correspondence to the first author. We thank Michael Bacharach, David Buschena, David Cooper and participants to presentations in Oxford, Strasbourg and York for advice and Jahirul Islam for experimental assistance. The usual disclaimer applies. This paper is dedicated to the memory of Michael Bacharach.

The opportunity to cooperate underlies a wide spectrum of interactive scenarios. Understanding cooperation provides fundamental insight into many real-world situations, such as the assessment of organizational structures, the design of environmental policies, and defining terms in bilateral trade negotiations. It is important in a host of social dilemmas and in any transaction where economic agents have to trust and fufill trust in order for the Pareto optimal outcome to be achieved. However, different interactive scenarios do not necessarily entail the same extent of interest for players to cooperate: cooperation is likely to vary according to the elements of the game and how the game is perceived by the respective players.

It is then useful to identify common properties of games that are easy to estimate and to use as predictors of how much cooperation one should expect. Such a univocally defined measure allows us to systematically label, along the same dimension, different games. This paper presents an experimental test of the relationship between *game harmony* and cooperation in 2 × 2 games, including familiar ones such as the Prisoner's Dilemma, the Stag-Hunt, the Chicken and three variants of trust games. Game harmony is a generic game property that describes how harmonious (non-conflictual) or disharmonious (conflictual) the interests of players are, as embodied in the payoffs. Consider the coordination game, constant-sum game and Prisoner's Dilemma of Table 1.

(Insert Table 1 about here).

In the coordination game there is perfect harmony of interests between the players: the only problem is one, indeed, of coordination. In the constant-sum game, the gain of a player is the loss of another, which means that there is perfect *dis*harmony of interests. Most games, such as the Prisoner's Dilemma, do not fall in either of these two extremes: rather, they are somewhere in the middle, in terms of harmony of interests, between coordination games and constant-sum games. Game harmony measures, being intuitively straightforward numerical indices, capture this "somewhere in the middle" more precisely. Zizzo's (2003b) most basic measure is an average of Pearson correlation coefficients among payoff pairs: in the case of a 2 × 2 game, it reduces to the simple Pearson correlation coefficient between the payoffs of the players for each state of the world. Another measure replaces Pearson with Spearman correlations, and so a cardinal with an ordinal measure of game harmony. The ordinal measure is less sensitive than the cardinal one to the particular payoff values being chosen, and might therefore be more appealing when one seeks to compare game harmony between different game structures; for robustness, we consider both measures. In two-player games, both measures are bounded between –1 and +1, with +1 the value for coordination games and –1 the value for constant-sum games: higher numerical values indicate greater game harmony.

We compute game harmony values relying only on monetary payoffs. We argue that our simple game harmony measures can be used as effective predictors of cooperation in 2×2 games: the greater the harmony of a game, the more the cooperation we should expect in a game. We find that game harmony can explain between 57% and 93% of the variance in cooperation rate in a sample of well-known games, and between 31% and 45% in a sample of randomly generated games. These results are interesting, as they rely on minimal informational requirements on the part of the observer: all that we need to know is the monetary payoffs in order to have, at least, a benchmark cooperation rate.

We are not aware of other attempts in measuring game harmony as a numerical index in a general setting. Our correlation-based measures of game harmony are closest to the "index of correspondence" (IC) informally presented for 2×2 games by two social psychologists, Harold Kelley and John Thibaut (1978); IC has been rather neglected by economists. However, their measure is more complex than the basic measures of game harmony reviewed in this paper, and is applicable only for 2-player games. Furthermore, IC is not a pure measure of game harmony as it confounds harmony with scope for cooperation, i.e. with the amount that players stand to gain from cooperating. We show that at least in our game samples IC does not gain explanatory power by doing so.

Anatol Rapaport and Albert Chammah (1965) devised specific cooperation indices in relation to 2×2 Prisoner's Dilemmas. Their measures are very *ad hoc* and can only be applied to this narrow, if important, class of games. Zizzo (2003b) shows that their most successful measure appears to work as a proxy, in this specific setting, for our general measure of cardinal game harmony; such a measure can explain some 80% of the variance in cooperation rates in the Rapaport and Chammah's dataset. Furthermore, Zizzo (2003a) shows the predictive success of basic game harmony measures in a sample of mostly randomly generated 3×3 games.

Although it is trivial to show that game harmony is not the only thing that matters for cooperation (and we do so in section 4), we nevertheless posit that it plays a significant predictive role. In section 1 we review some relevant results of Zizzo (2003b) and present our experimental hypothesis. Sections 2 and 3 describe the experimental design and results, respectively. Section 4 contains the discussion and conclusions. The experimental instructions are in the appendix.

I. Game Harmony and Experimental Hypothesis

Following Zizzo (2003b), let Γ be a finite N-person game in normal form, and let N be the set of players such that $|n| = N$. Denote W_i the actions available to player i, so $W = W_1 \times W_2 \times ... \times W_n$ is the set of possible outcomes or states of the world, each of which we label by w in W. Payoffs are defined by x_{iw}: $W \rightarrow \mathbb{R}$, a standard Von Neumann-Morgernstern utility function, and so x_{iw} is the payoff for player i ($i \in N$) in state of the world w. For n players, there are $C = ½\, n\, (n\text{-}1)$ player pairs. Let us label the payoffs x_{iw} for each pair c as a_{cw}, b_{cw} for $w \in W$. Under some weak technical assumptions guaranteeing existence (see Zizzo, 2003b), one can then define the cardinal harmony G(Γ) of game Γ as the arithmetic mean of the Pearson correlations between the C pairs of Γ:

(1) $$G(\Gamma) = G(x_i, W) = \frac{1}{C}\sum_{c=1}^{C} r_c(a_{cw}, b_{cw}) = \frac{1}{C}\sum_{c=1}^{C} \frac{Cov(a_{cw}, b_{cw})}{\sigma_a \sigma_b}$$

In the case of 2 × 2 games, $C = 1$, so G(Γ) = $r(a_w, b_w)$, which is obviously bounded between -1 and +1. We shall mention three of the properties that can be proven in relation to G(Γ). G(Γ) is scale-independent, meaning that G(Γ) = G(x_i, W) = G($k_i\, x_i + h_i, W$) for any finite $k_i > 0$ and h_i. G(Γ) is also invariant to transformations of Γ into a game Ψ that is the aggregation of any number of replicas of Γ, so G(Γ) = G(Ψ). Further, an equivalence result can be shown to hold between perception of a game as more or less harmonious and (psychological or otherwise) payoff transformation. In a non-psychological interpretation, the payoff transformation could simply be a way of comparing game harmony and cooperation rates in any two given payoff matrices where, at least at a first approximation, such a relationship between payoffs were to hold. Let the payoff transformation for player i in relation to the other players indexed by $j \neq i$ be defined by the function:

(2) $$V_{iw} = x_{iw} + \sum_{j \neq i}^{n} \beta_i x_{jw}$$

where $\beta_i \in [-1,1]$ is the weight on j's utility component. Denote Γ^u the untransformed payoff matrix and Γ^v the payoff matrix transformed according to V.

PROPOSITION (Equivalence Result). *Assume that the material payoffs* a_{cw} *and* b_{cw} *are given and that* $G(\Gamma^v)$ *and* $G(\Gamma^u)$ *are defined. Then* $G(\Gamma^v) \gtreqless G(\Gamma^u) \leftrightarrow \beta_i \gtreqless 0 \;\forall\; i$, *and, if* $\beta_i > 0$, $\partial\beta_i / \partial G(\Gamma^v) > 0$.

The proof of this result is in Zizzo (2003b). The greater game harmony associated to a positive payoff transformation will imply greater cooperation in all games where such simple payoff transformation will help to attain it (such as in the Prisoner's Dilemma). We do not believe, however, that this is the only way in which a relationship between cooperation and game harmony can be established. This paper tries a different approach, experimental rather than theoretical: we verify whether game harmony can indeed be used as a predictor of cooperation in games, and in what sense this is the case.

A closely related measure of game harmony can be obtained by considering payoff orderings rather than their cardinal values. Let X_i be the set of all payoff values for player i in W, and let $x^{\rho}_{iw} = rank(x_{iw} | X_i)$, which can be mapped into rank payoff pairs $a^{\rho}_{cw}, b^{\rho}_{cw}$. Then:

$$(3) \qquad G_{\rho}(\Gamma) = G_{\rho}(x_i, W) = \frac{1}{C}\sum_{c=1}^{C} r_c(a^{\rho}_{cw}, b^{\rho}_{cw})$$

In the case of 2 × 2 games, this reduces to $G_{\rho}(\Gamma) = r_c(a^{\rho}_{cw}, b^{\rho}_{cw})$. Both $G(\Gamma)$ and $G_{\rho}(\Gamma)$ implicitly assign equal weight to all states of the world. A more general class of game harmony measures in Zizzo (2003b) relaxes this assumption: in order for these more general measures to have any bite, however, one would need to specify how the different weights are determined. A natural way of doing this would be to assume that, if a subject follows an algorithm $\mathcal{L}$ (e.g., Nash) in solving Γ, a weight of 0 should be assigned to the $\mathcal{L}$-dominated outcomes, and game harmony should be defined only over the $\mathcal{L}$-undominated outcomes. While this may have some normative appeal, it also leads to a "uniqueness paradox": if $\mathcal{L}$ has a unique solution, then there is only one $\mathcal{L}$-undominated outcome, and so game harmony cannot be computed, since no correlation coefficients among payoffs can be computed on the basis of just one observation. This is paradoxical, as it means that, the more successful $\mathcal{L}$ is in pinpointing a unique solution, the less the range of games over which $\mathcal{L}$-weighted game harmony can be computed. This makes $\mathcal{L}$-weighted game harmony unusable in our context, since many well-known 2 × 2 games have unique solutions with algorithms such as L1 or L2[1] or Nash, and even a weak algorithm such as rationalizability would imply the non-computability of game harmony in the Prisoner's Dilemma.

Let there be only two players, 1 and 2, and let σ_1 and σ_2 be the respective standard deviations of their payoffs. One can then define Kelley and Thibaut's IC measure as:

(4) $$IC = G(\Gamma)\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}$$

a measure applicable only for two-player games. In their general applicability and lack of degrees of freedom, we believe that G(Γ) and $G_\rho(\Gamma)$ have the virtues of simplicity and parsimony. It will be on them that we shall focus in the rest of our paper to make a case for game harmony as a predictor of cooperation in 2 × 2 games. Nevertheless, we shall also analyze whether the extra term in (4) provides any additional explanatory power to IC in our samples relative to G(Γ).

We make the parsimonious assumption that there is a strong link between perceived harmony and harmony value as can be estimated from the game matrix: if so, one can estimate G(Γ) and $G_\rho(\Gamma)$ from the material payoff matrix, treating the material payoffs as the utility values x_{iw}, and use G(Γ) and $G_\rho(\Gamma)$ as predictors of cooperation in Γ. Our experimental hypothesis can then be formulated as follows: *whenever a unique cooperative action is defined in* Γ*, the mean cooperation rate (i.e., the mean probability of choosing this action) is increasing in* G(Γ) *and* $G_\rho(\Gamma)$.

Obviously, this is a strong assumption, but it is one that is convenient to make as a strong and parsimonious test of the practical relevance of our measures of game harmony. Of course subjects may perceive the game as being more or less harmonious than what its material payoff structure entails, depending on situational factors such as the way the instructions are phrased or a history of play of similar games, and more individual-specific factors such as the player's personality. To check for history-dependence, in one experiment (Experiment A) we varied the game harmony of the games assigned in the practice stage. In addition, Zizzo and Tan (2002) presents the results of the last stage of Experiment A, where subjects were asked to judge the similarity of a given game to either a pure coordination (perfectly harmonious) or constant-sum (perfectly disharmonious) game. This was an attempt to try to gather extra information on how agents perceived the harmony of the game, but will not be discussed here.

---

[1] As defined by Dale Stahl and Paul Wilson (1995) and Miguel Costa-Gomes et al. (2001), L0 corresponds to random play, L1 to best reply against random play, and L2 to best reply against the best reply against random play.

II. The Experimental Design

A. *Experiment A*

Fifteen experimental sessions were run in Oxford in the March of 2002. Each session had four subjects, for a total of an overall 60 subjects. The experiment was divided in three stages (plus the payment stage), all of which involved 2×2 games. The games were never exactly symmetric, though they were sometimes approximately so (as will be explained below), so one could play each game in one of two roles. It may be tempting to label these roles as that of "row player" and that of "column player", but subjects always had game matrices displayed on the screen in such a way that they would be row players: this was achieved by suitably transposing game matrices as needed. We thus find more appropriate to define the role according to whether, in any given round, subjects saw the "direct" (standard) or "transposed" presentation of the game on the computer display, and label their roles as that of d-players and t-players, respectively. All game payoffs were provided as numbers between 0 and 100.

In the experimental instructions games were labeled as "decision tables", players as "participants" and coplayers as "coparticipants". In order to check understanding of the instructions, subjects filled questionnaires at the start of each stage. Their answers were checked by experimenters, and, if any was incorrect or missing, the relevant points were explained individually. Instructions and questionnaires were on paper, but the experiment was otherwise computerized.

Stage 1. In the first stage ("Practice Stage"), subjects did practice by playing six 2×2 games twice, namely once as d-players and once as t-players. They were matched with a single co-player throughout the stage, played games in random order and received feedback about the outcome of each round. Practice Stage points did not count towards final winnings. Games had been chosen according to the following procedure: 1) payoff values were generated randomly, by drawing payoff values from a uniform distribution between 0 and 100; 2) games without a unique pure Nash equilibrium were discarded; 3) according to the experimental condition, games with high, medium or low game harmony were selected (High, Medium and Low condition, respectively). Games with a unique pure Nash equilibrium were chosen because we wanted games to be, at least roughly, of the same strategic complexity notwithstanding differences in game harmony. Furthermore, in order to ensure that any distortion in behavior in the later stages were not due to a particular reinforcement learning history from the practice stage, we tried to use different game samples in different sessions, albeit games of about the same level of game harmony for sessions of the same condition.[2] The mean G(Γ) [$G_\rho(\Gamma)$] values were 0.985 (0.943), 0.020 (0.013) and -0.980 (-0.915) for the games used in the High, Medium and Low conditions, respectively. Each of the two practice stage pairs in each session played the games in random order.

Stage 2. In the second stage, subjects played ten 2 × 2 games twice, again once as d-players and once as t-players. Games were played in random order. No feedback was received after each round in this stage. At the end of the experiment, a round was randomly picked by the computer to determine the "action payment". The subject's action in this round was matched with that of a different "coparticipant" from that of the practice stage. Each payoff point earned in the payment round was converted into 0.06 U.K. pounds. Subjects could therefore earn between zero and six pounds as their action payment, depending on their choice of action, that of their coplayer and the game played in the payment round. The game matrices are displayed in Table 2.

(Insert Table 2 about here).

Most Table 2 games are payoff-perturbed versions of familiar 2 × 2 games: Prisoner's Dilemma (PD), Stag-Hunt (StH) and Chicken (ChK), plus a constant-sum game (CSG) and a coordination game (CDG). The Envy (Altruism) Game (EG and AG, respectively) is a game with a strictly dominant solution in its material payoff values: deviations from the strictly dominant solution can be interpreted as due (if not to trembling) to envy (altruism) or other negative (positive) interdependence in preferences. The remaining games are instances of trust games as defined in James Coleman (1990) and Michael Bacharach et al. (2001): the d-player is the truster, with a choice whether to trust (by playing top) or withhold trust (by playing bottom); the t-player is the trustee, with a choice whether to fulfill or violate the trust. Bacharach et al. (2001) employ both the Kind Trust Game (KTG) and the Needy Trust Game (NTG): the difference between the two is that, in the latter, the truster is "more in need", i.e. she ends up in a worse fate if she does not trust. UTG

[2] Originally 4 sessions were planned for each condition and so four sets of games were prepared for each condition. When a fifth session was run at about the same time because of a surplus of subjects, one of the sets of games was used again for each condition. Thus, in each condition one game sample was used twice. Twelve different game samples were used in a total of fifteen sessions.

stands for Unequitable Trust Game and is equivalent to the KTG, but with about 20 points added to the payoff values of the truster and 30 points subtracted from the payoff values of the trustee.

Small payoff perturbations have been used before in experiments with 2 × 2 games (e.g. Frederick Rankin et al., 2000). They were used in our experiment for two reasons: 1) we wanted to reduce the likelihood that subjects would realize that they were playing each game in both roles; 2) we wanted games in Stage 2 not to appear "different" from Stage 1 games because of the symmetric nature of some of them and the higher frequency of the same numbers being used as payoff values. Payoff perturbation explains why games were never exactly symmetric, although four out of ten (PD, StH, ChK, CDG) were basically so (i.e., were symmetrical up to the payoff perturbation).

Stage 3, on similarity evaluations, followed, and is described in Zizzo and Tan (2002). In playing Stage 2, subjects knew that Stage 3 exclusively involved individual choice tasks, i.e. no interactions among the players. At the end of the experiment subjects were paid the action payment (up to U.K. £ 6), the payment from Stage 3 (up to £ 12), plus £ 4 for participation. Average payments were about £ 10 for around one hour of work.

B. *Experiment B*

Experiment B was run in Oxford in November 2002 in five sessions of four subjects each, for a total of twenty subjects. It consisted of two stages only, Stage 1 and Stage 2: there was no "similarity evaluations" stage. Stage 1 was for practice and was identical to the Medium condition of Experiment A (but different sets of randomly generated games were used). In Stage 2, subjects had to play thirty randomly generated games twice, once as d-players and once as t-players, for a total of sixty rounds. The games are reproduced in Table 3, and were again presented to the subjects in random order. Payoffs were drawn from a uniform distribution between 0 and 100 in all cases, and the only filtering rule in choosing the games was that we required the Nash bargaining solution action (i.e, the action associated to the outcome with the highest *product* of payoffs) to coincide with the utilitarian action (i.e., the action associated to the outcome with the highest *sum* of payoffs). We did not compute game harmony values of the random games ahead of running the experiment.

(Insert Table 3 about here.)

Each experimental point was worth £ 0.08 at the end of the experiment.[3] Mean payments were about 8.5 pounds for around 45-50 minutes of work.

III. Experimental Results

A. *Definition of cooperative action*

In order to determine whether game harmony was associated with a higher likelihood of play of the cooperative action, we need to define what the cooperative action is for each game in Stage 2. In Experiment A, in the light of the easy interpretability and often well-known features of the games at hand, this should not be too controversial: unique cooperative actions, when defined, are illustrated in bold in Figure 1. No cooperative action is defined in the constant-sum game, nor (ignoring the payoff perturbation) is there a clear one for the t-player in the EG and AG (although payoff perturbation implies a strictly dominant solution in both cases). For the d-player, the altruistic action in the AG and the non-envious action in the EG are considered as the cooperative actions; in the case of the AG, this can be justified, for example, on the basis of either a utilitarian social welfare function or the Nash bargaining solution. For the trust games, the (*trust*, *fulfillment*) pair is always identified as the cooperative outcome. In the coordination game the payoff (and Pareto) dominant solution is taken as the cooperative outcome. The cooperative action are obvious in the remaining cases. Overall, a cooperative action is defined for 960 Stage 2 observations.

This game-by-game-procedure to determine cooperative outcomes might be criticized for not being systematic enough. In the light of this, we devised an alternative, rule-based procedure to determine whether an outcome can be considered uncontroversially cooperative: we treated an action as cooperative if, over the (non-payoff-perturbed) game matrix,

(a) it corresponds to both the utilitarian solution (it has the highest sum of payoffs) and the Nash Bargaining solution (it has the highest product of payoffs);

(b) it does not correspond to a strictly dominant strategy.

Criterion (b) follows the intuition that, if an action is strictly dominant, then it is best for a player to play that action irrespectively of what the other player does, and therefore it may not make much sense to talk in this case of "co-operation", interpreted as "operating with", working together with, the other player towards a common goal.

---

[3] The marginal rate was slightly higher than in Experiment A to compensate for the longer length of Stage 2.

There are only three differences in practice between the two procedures. The game-by-game procedure defines a cooperative outcome for the d-player in the EG and the NTG, but, as this action corresponds to the strictly dominant strategy, it is ruled out by the rule-based procedure. The rule-based procedure is also undefined in relation to the t-player in the Chicken game, as the actions associated to the Nash Bargaining and utilitarian outcomes are different in the two cases. We shall refer to cooperation as determined by the game-by-game procedure and rule-based procedure as *g-cooperation* and *r-cooperation*, respectively.

The rule-based procedure is particularly helpful in Experiment B, where, because of the random nature of the games, it is necessary to have a clear-cut and well-defined criterion for defining an action as cooperative.

B. *Results*

In Experiment A mean g-cooperation (r-cooperation) rates were 0.469 (0.389), 0.531 (0.458) and 0.509 (0.454) in the Low, Medium and High condition, respectively. There is no robust evidence for an effect of having been faced with games of different levels of game harmony during the practice on later cooperative play. For example, a nonparametric Kruskal-Wallis test yields χ^2 (2) = 1.505 (1.787) in relation to g-cooperation (r-cooperation); in both cases, $P > 0.1$.

(*Insert Tables 4 and 5 about here*).

Table 4 presents the the mean cooperation rate for each game and the corresponding game harmony (and Kelley and Thibaut's IC) values in relation to Experiment A. Table 5 provides information on the correlation coefficients between game harmony (and IC) values and mean g-cooperation and r-cooperation rates in relation to both experiments. Cooperation rates are defined not only *by game* but also *by game role* (i.e., by whether an agent is d-player or t-player for any given game matrix). It is interesting to analyze cooperation rates by role for two reasons. First, basic (i.e., unweighted) game harmony measures (or the IC) cannot explain differences in cooperation rates between different roles in asymmetric games: by grouping cooperation rates by game and so averaging out cooperation rates in different roles in the asymmetric games, we are ignoring this source of unexplained variation. Second, in particular cases (such as for the AG), cooperation rates are computed from half of the observations that we have for the other games of Figure 1, since they are computed on the basis of d-players' choices only (the only ones for which, as we explained above, we can define a cooperative choice for these games). This is not taken into account in the correlation analysis by game.

(*Insert Figures 1-3 about here*).

The main finding from Table 5 is one of a strong and always statistically significant positive correlation between game harmony, both in its cardinal and its ordinal measure, and mean cooperation rate by game c_Γ and by role c_r: the correlation coefficients involving G(Γ) and $G_\rho(\Gamma)$ range between 0.557 and 0.962 (with $P < 0.05$ or better). The figures illustrate this correlation in relation to G(Γ). Our basic game harmony measures can explain between 31% and 45% of the variance in mean cooperation rate in Experiment B, depending on the specific game harmony and the correlation measures used. They can explain between 57% and 93% of the variance in mean cooperation rates in Experiment A. The correlations are higher in Experiment A, with mostly well-known and classic games, than in Experiment B, with randomly generated games. Still, with either dataset, the evidence suggests that *basic game harmony measures can be used as predictors of cooperation in 2 × 2 games*. The result is noteworthy in the light of the simplicity and parameter-free nature of our measures. It conforms to our prediction of an association between cooperation and basic game harmony measures.

Other findings also emerge. G(Γ) performs better than $G_\rho(\Gamma)$ in Experiment B; it also performs better than IC. In Experiment A, the correlation coefficients between IC and mean cooperation rates appear roughly the same as those involving G(Γ) and $G_\rho(\Gamma)$. There is no evidence suggesting that using a more complex concept than game harmony, such as IC, adds any predictive power on our datasets.

Of course, there are outliers. For example, as shown by Figure 1, in Experiment A the main outliers for the (g-cooperation, G(Γ)) pair are KTG and UTG: while these games are only just slightly more disharmonious than the Chicken game, subjects are only about half as likely to cooperate in them. In moving from g-cooperation to r-cooperation, in Experiment A the goodness-of-fit deteriorates because (a) the EG observation has a good fit in the g-cooperation dataset and (b) the NTG observation, if reflecting only the choices of the t-player as for the r-cooperation dataset, becomes an outlier. In relation to Experiment B, five games present OLS residuals higher than 0.3. Games 2, 10 and 24 are negative outliers. In games 2 and 24 the action that is associated with the L0, L1 and L2 strategy, and with strict dominance, do not coincide with the cooperative solution for both players. In game 10, the L0, L1 and L2 solutions again are not the same as the cooperative solution (in addition, for the t-player the non-cooperative solution present less risk). Games 3 and 6 are positive outliers. In both cases, the L0, L1 and L2 solutions are the same as the cooperative solutions.

Cases such as those of games 3 and 6 in Experiment B suggest that some subjects may have played the "cooperative" action without any intention to co-operate, and so point to the difficulty of defining cooperation unambiguously when one deals with not well-known games. But, more generally, the anomalies are only such if we were to believe what cannot be believed, namely that game harmony is the only thing that matters in determining cooperation. As our discussion of outliers suggested, and an example in the next section also will drive home, there are strategic features of games that are not fully by the game harmony values.

IV. Discussion and Conclusions

Simple game harmony measures can be used to predict the occurrence of cooperation across a range of well-known and randomly generated 2 × 2 games. Knowledge of the material payoff matrix is all that is required to compute these measures. We presented the results from two experiments, one where subjects played mostly well-known 2 × 2 games, such as the Prisoner's Dilemma, the Stag-Hunt, the Chicken and three variants of trust games (Experiment A), and the other where subjects played randomly generated games (Experiment B). Even when due attention is given to the meaningful asymmetry of roles in some of these games, we find Pearson and Spearman correlations above 0.8 between our measures of game harmony and mean cooperation rates in Experiment A, and above 0.55 in Experiment B.

Obviously, it would be naïve to think that game harmony is the only thing that matters for cooperation, for at least two reasons. First, because of the scale invariance property, it does not take into account how much is at stake in cooperating when everything else is proportionately the same (e.g., Bacharach, in press). This is what is implicitly taken into account by Kelley and Thibaut's (1978) "index of correspondence", though in our datasets we find that it has no additional explanatory power. Second, two games can have the same $G(\Gamma)$ and $G_\rho(\Gamma)$ and yet be likely to lead to different cooperation rates, because of differences in the strategic structure of the game. Table 6 contains a pair of games exemplifying this limitation.

(Insert Table 6 about here).

Our results however point to the conclusion that simple game harmony measures can be powerful tools in assessing cooperation rates in 2 × 2 games, including well-known and important formulations of social dilemmas and games where coordination is an issue.

Appendix

The experimental instructions are provided for Experiment A. Those for Experiment B had the changes indicated in <…> brackets.

<u>Instructions for Stage 1</u>

You are about to participate in an experiment on decision-making. The experiment will be conducted in four <three> stages. Stage 1 (the Practice Stage) is for practice only, while in the Payment Stage you are paid whatever amount you have earned in Stages 2 and 3 <Stage 2>, plus additional 4 pounds for participation.

In the Practice Stage you will be asked to choose *actions* for twelve rounds. Each round your action will be paired with that of one other participant (your *coparticipant*), and this will determine the outcomes both for you and your coparticipant. The nature of the decision in the Practice Stage is discussed below. You will always be matched with the same coparticipant in the Practice Stage. After each round you will be told what actions were chosen by you and your coparticipant, and how many *experimental points* you and your coparticipant earned in the round as the result of your actions. You will receive *no information* about the actions of and points earned by the participants that are not your coparticipant, and similarly they will not be informed about your actions or the points you have earned. In the later stages of the experiment, you will *not* be matched with the same coparticipant as in the Practice Stage.

You should try to make the best decisions you can in the Practice Stage: by doing so you can get the greatest understanding on how to do well in the rest of the experiment.

The Decision Table in the Practice Stage

Each decision that you face will be described by a *Decision Table* consisting of eight numbers arranged in two rows and two columns. Decision Tables will appear also in Stage 2 and Stage 3, and so it is quite important that you get a good understanding of what they represent.

An example (namely, the Decision Table for round 1) is currently on display on the computer screen. You and your coparticipant have two available actions, A and B. A yellow and a blue cell, in pairs, are placed in a grid in correspondence of each of the nine combinations of possible actions by you and your coparticipant. Two numbers, one in the yellow cell and one in the blue cell, are placed in correspondence of each combination of possible actions. The number in the *yellow* cell is the amount of experimental points that *you* would get for each combination of possible actions; the number in the *blue* cell is the amount of experimental points that your *coparticipant* would get for each combination of possible actions. To make some examples based on the Decision Table on the computer display: if you choose A and your coparticipant chooses A, you get [*number of points*] points and your coparticipant gets [*number of points*]; if you choose B and your coparticipant chooses B, you get [*number of points*] points and your coparticipant gets [*number of points*]; finally, if you choose B and your coparticipant chooses A, you get [*number of points*] points and your coparticipant gets [*number of points*]. The *point numbers* in all cells are always between 0 and 100.

To choose an action, you need to click one of the buttons labelled A and B. You should click A if you want to choose action A, and B if you want to choose action B. A message window will then appear asking you to confirm your choice. To do so, click OK on the window and then click the

Confirm button. If you want to cancel your choice, click OK on the window and then click the Cancel button.

You will not get to know the choice of your coparticipant for the round until your coparticipant has chosen as well, and similarly he or she will not learn about your action until he or she has made his or her choice. In making your choices, you are not allowed to speak to other participants or communicate in any other way.

Before starting the practice, we ask you to answer a brief questionnaire, with the only purpose of checking whether you have understood the instructions. Raise your hand when you have completed the questionnaire.

Many thanks for your participation to the experiment, and good luck!

Please raise your hand if you have any questions.

Instructions for Stage 2

In Stage 2, you are asked to choose actions for twenty <sixty> rounds in relation to Decision Tables.

At the start of Stage 2, you will be matched with a *different* coparticipant from the one you will have played the Practice Stage with. You will have to take decisions for twenty <sixty> rounds, but you will receive no feedback on their outcome until the end of the experiment.

This is the last interactive stage of the experiment: your Stage 3 earnings will depend only on your choices, not on combinations of choices by you and some other participant, while Stage 4 is just for payment. *<Paragraph omitted in Experiment B.>*

Your choices

You can choose an action exactly as you have done in the Practice Stage, first by clicking on the A or B button and then by confirming. You and your coparticipant will earn point numbers as the result of your actions, exactly as in the Practice Stage.

You will not receive any feedback about the outcome of your choices after each round. No communication of any kind with the other participants is allowed.

Your winnings

The computer will randomly choose a payment round to determine the *action payment*. This payment round will be the same for you and your coparticipant.

The action payment depends on the point numbers you earn in this round, and so it depends on the actions by you and your coparticipant. More specifically, *each point earned in this round is worth 0.06 <0.08>* pounds *in the Payment Stage* (so, for example, 100 points are worth 6 <8> pounds).

Please do not take decisions in a hurry: you can improve your chances to do well by thinking carefully about each Decision Table.

Before starting making decisions, we ask you to answer a second brief questionnaire, once again with the only purpose of checking whether you have understood the instructions. Raise your hand when you have completed the questionnaire.

Please raise your hand if you have any questions.

REFERENCES

Bacharach, Michael. *Beyond the individual: The theory of cognition and cooperation in games*. Princeton: Princeton University Press, in press.

Bacharach, Michael; Guerra, Gerardo and Zizzo, Daniel J. "Is Trust Self-Fulfilling? An Experimental Study." Mimeo, University of Oxford, Oxford, 2001.

Coleman, James S. *Foundations of social theory*. Cambridge, MA: Harvard University Press, 1990.

Costa-Gomes, Miguel; Crawford, Vincent P. and Broseta, Bruno. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica*, September 2001, *69*(5), pp. 1193-1235.

Kelley, Harold H. and Thibaut, John W. *Interpersonal relations: A theory of interdependence*. New York: Wiley, 1978.

Rankin, Frederick W.; Van Huyck, John B. and Battalio, Raymond C. "Strategic Similarity and Emergent Conventions: Evidence from Similar Stag Hunt Games." *Games and Economic Behavior*, August 2000, *32*(2), pp. 315-37.

Rapoport, Anatol and Chammah, Albert M. *Prisoner's dilemma: A study in conflict and cooperation*. Ann Arbor: University of Michigan Press, 1965.

Stahl, Dale O. and Wilson, Paul W. "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, July 1995, *10*(1), pp. 218-54.

Zizzo, Daniel J. "Game Perception and Harmony in 3×3 Games: An Experiment in Cognitive Game Theory." Mimeo, University of Oxford, Oxford, 2003a.

Zizzo, Daniel J. "Harmony of Games in Normal Form." Mimeo, University of Oxford, Oxford, 2003b.

Zizzo, Daniel J. and Tan, Jonathan H. W. "Game Harmony as a Predictor of Cooperation in 2×2 Games." Mimeo, University of Oxford, Oxford, 2002.

TABLE 1 – EXAMPLES OF GAMES WITH DIFFERENT HARMONY

Coordination Game		Constant-Sum Game		Prisoner's Dilemma	
3, 3	0, 0	0, 3	3, 0	2, 2	0, 3
0, 0	1, 1	3, 0	0, 3	3, 0	1, 1

TABLE 2 – GAMES USED IN EXPERIMENT A IN THEIR "DIRECT" PRESENTATION

Prisoner's Dilemma		Envy Game		Altruism Game	
92, 11	38, 37	*61, 72*	*59, 73*	59, 28	61, 29
64, 63	10, 93	50, 28	48, 29	*48, 72*	*50, 73*

Stag-Hunt		Chicken	
10, 51	**92, 93**	92, 36	10, 11
52, 53	52, 9	**63, 62**	38, 94

Trust Games					
Kind Trust Game		Unequitable Trust Game		Needy Trust Game	
33, 34	34, 35	52, 3	53, 4	3, 34	4, 35
81, 82	14, 100	**100, 51**	33, 69	**81, 82**	14, 100

Also Comparison Decision Tables in Stage 3			
Constant-Sum Game		Coordination Game	
71, 31	18, 84	74, 75	32, 31
26, 76	89, 13	13, 12	**85, 86**

Notes: Bold letters stand for a unique cooperative solution. Italics denote the unique cooperative action for the row player in the Envy Game and the Altruism Game. Obviously in the actual experiment the cooperative actions and outcomes were not highlighted in any way. Stage 2 t-players saw the game matrices in their transposed form.

TABLE 3 – GAMES USED IN EXPERIMENT B IN THEIR "DIRECT" PRESENTATION

	Column Player	
Row	A, B	C, D
Player	E, F	G, H

Game Code	A	C	D	E	F	G	H	I
1	6	86	5	74	70	64	26	55
2	36	8	71	20	24	40	66	48
3	63	90	15	14	13	76	32	80
4	42	69	55	53	72	84	44	25
5	87	71	91	76	32	42	76	3
6	52	2	70	82	10	94	91	49
7	35	90	6	80	7	84	40	74
8	61	30	83	35	76	95	80	95
9	90	24	87	82	11	21	75	37
10	66	93	27	84	35	15	83	63
11	21	80	91	71	80	62	43	69
12	25	94	30	77	74	2	67	70
13	92	86	16	5	67	22	75	43
14	93	24	3	12	46	95	81	27
15	8	4	62	82	16	86	41	42
16	71	80	31	30	46	53	20	70
17	29	7	81	79	95	20	83	99
18	72	18	21	16	57	15	97	19
19	24	72	80	92	40	81	65	25
20	78	28	7	66	24	45	22	41
21	85	70	3	91	91	62	80	26
22	83	47	24	41	59	80	0	98
23	95	51	51	57	10	46	17	87
24	84	39	41	0	11	14	47	77
25	35	89	78	16	63	7	49	53
26	73	76	4	82	24	11	58	10
27	20	38	20	50	60	94	98	9
28	86	5	39	90	92	30	92	80
29	80	3	77	22	71	29	8	85
30	66	8	27	23	100	69	34	26

Notes: Games are defined row-by-row: to obtain the payoff matrix corresponding to a given round and condition, replace the A, B, C, D… values in the generic payoff matrix with the corresponding value on the row for that round and condition. Stage 2 t-players saw the game matrices in their transposed form.

TABLE 4 – GAME HARMONY AND MEAN COOPERATION RATES IN EXPERIMENT A

					Direct game	Tranposed game
Game (Γ)	G(Γ)	Gρ(Γ)	IC	c_{Γ}	c_r	c_r
Experiment A (Game Name)						
Prisoner's Dilemma	-0.817	-0.8	-0.817	0.2	0.217	0.183
Coordination Game	1	1	0.999	0.825	0.783	0.867
Envy Game	0.98	0.6	0.468	0.883	{0.883}	
Altruism Game	-0.98	-0.6	-0.468	0.2	0.2	
Kind Trust Game	0.066	-0.2	0.065	0.308	0.3	0.317
Unequitable Trust Game	0.066	-0.2	0.065	0.308	0.383	0.233
Needy Trust Game	0.503	0.8	0.5	{0.608} [0.3]	{0.917}	[0.3]
Stag-Hunt	0.488	0.513	0.488	0.633	0.65	0.617
Chicken	0.16	0.2	0.16	0.6	{0.583}	0.617
Experiment B (Game Code)						
2	0.098	0	0.096	0.15	0.15	
3	0.599	0.8	0.554	1	1	1
6	-0.332	-0.4	-0.327	0.725	0.8	0.65
8	0.406	0.258	0.205	0.7	0.65	0.75
10	0.258	0	0.248	0.2	0.25	0.15
11	-0.691	-0.4	-0.3	0.4	0.25	0.55
13	0.857	1	0.855	0.8	0.75	0.85
14	0.004	0.2	0.004	0.45	0.45	
15	0.5	0.4	0.453	0.8	0.7	0.9
18	0.752	0.8	0.086	0.825	0.9	0.75
19	-0.07	0.4	-0.069	0.4	0.45	0.35
20	-0.861	-0.8	-0.694	0.275	0.35	0.2
21	-0.647	-0.4	-0.593	0.05	0.05	
22	-0.479	-0.4	-0.458	0.05	0.05	
23	-0.341	0.2	-0.264	0.05		0.05
24	0.363	0.6	0.36	0.175	0.05	0.3
25	-0.904	-0.8	-0.716	0.375	0.15	0.6
26	-0.106	-0.4	-0.103	0.525	0.4	0.65
29	-0.975	-1	-0.975	0.025	0.05	0

Notes: only games where a cooperative action is defined for at least one player are included. G(Γ) and $G_\rho(\Gamma)$ are our cardinal and ordinal game harmony measures, respectively; IC is Kelley and Thibaut's (1978) "index of correspondence" measure. c_Γ is the mean cooperation rate by game. Whenever defined, the "direct game" c_r is the mean cooperation rate of the row players facing the game matrices as presented in Table 2; the "transposed game" c_r is the mean cooperation rate of the row players facing the transpose of the game matrices as presented in Table 2. In relation to Experiment A, values in […] brackets refer to g-cooperation rates only, and values in {…} brackets refer to r-cooperation rates only; g-cooperation and r-cooperation rates coincide otherwise.

TABLE 5 – CORRELATIONS BETWEEN GAME HARMONY AND IC VALUES AND COOPERATION RATES

		Cooperation by Game		
		Experiment A		Experiment B
		G-Cooperation (*n* = 9)	R-Cooperation (*n* = 8)	R-Cooperation (*n* = 19)
Pearson	$G(\Gamma)$	0.921***	0.793*	0.649**
	$G_\rho(\Gamma)$	0.923***	0.766*	0.565*
	IC	0.867**	0.806*	0.611**
Spearman	$G(\Gamma)$	0.962***	0.752*	0.673**
	$G_\rho(\Gamma)$	0.911***	0.752*	0.583**
	IC	0.878**	0.752*	0.573**

		Cooperation by Role		
		Experiment A		Experiment B
		G-Cooperation (*n* = 16)	R-Cooperation (*n* = 13)	R-Cooperation (*n* = 33)
Pearson	$G(\Gamma)$	0.817***	0.832***	0.619***
	$G_\rho(\Gamma)$	0.834***	0.840***	0.557***
	IC	0.784***	0.844***	0.579***
Spearman	$G(\Gamma)$	0.864***	0.864***	0.668***
	$G_\rho(\Gamma)$	0.838***	0.864***	0.580***
	IC	0.816***	0.864***	0.540***

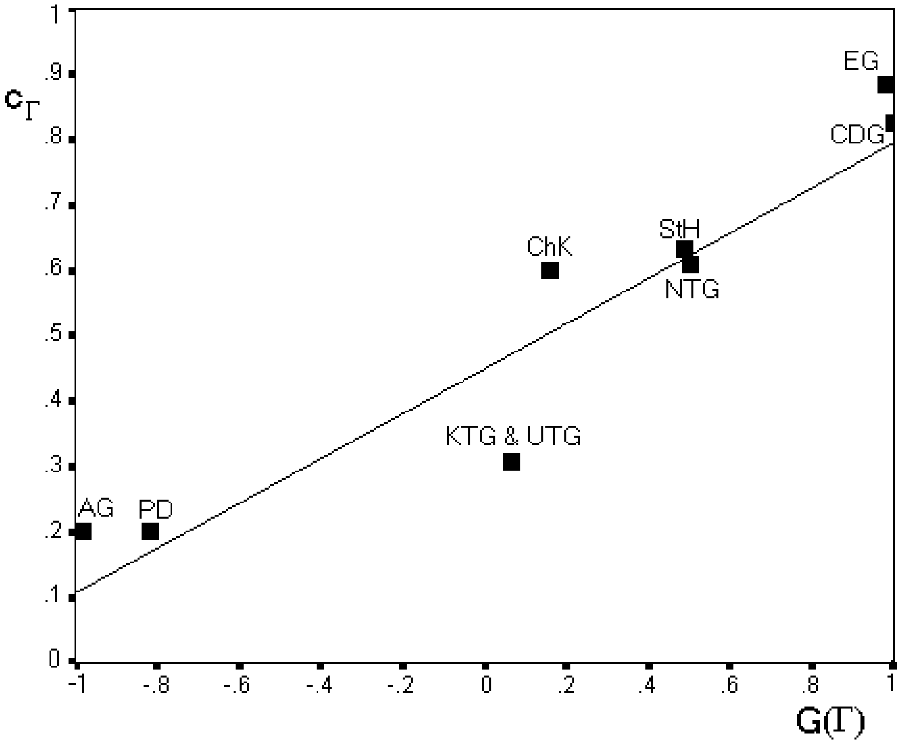
* Significant at the 5-percent level.
** Significant at the 1-percent level.
*** Significant at the 0.1-percent level.

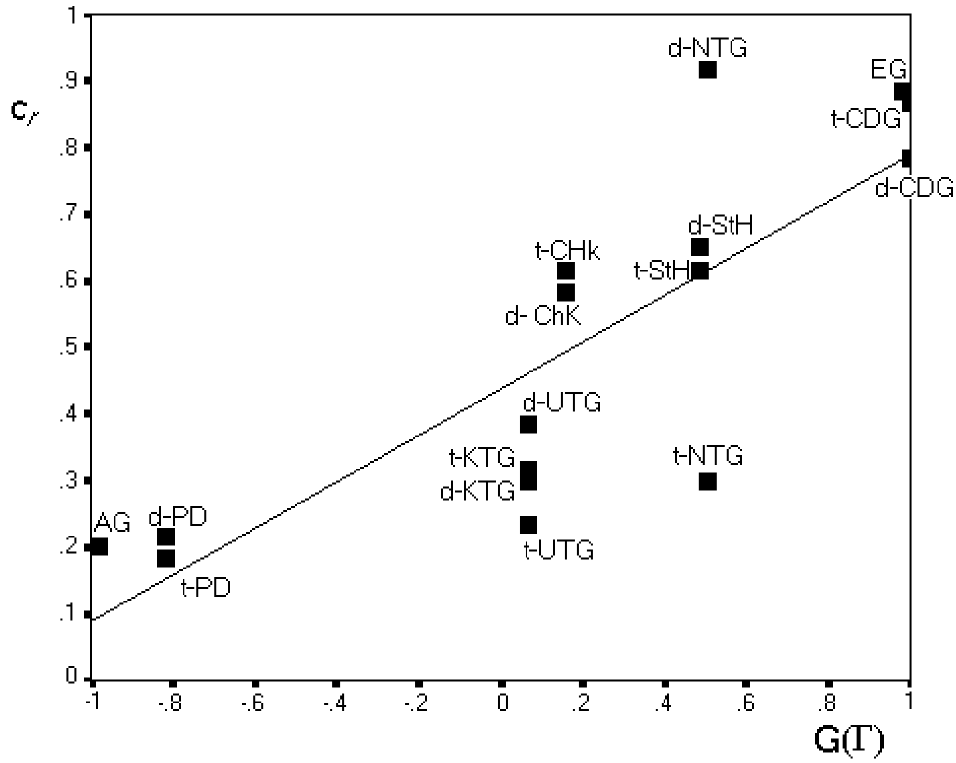
TABLE 6 – PAIRS OF GAMES WHERE GAME HARMONY DOES NOT WORK

Prisoner's Dilemma			PD with Swapped Cells	
2, 2	0, 3		2, 2	0, 3
3, 0	1, 1		1, 1	3, 0

Notes: the game to the right is obtained by swapping the second row of the Prisoner's Dilemma (PD). The two games have identical $G(\Gamma)$ and $G_\rho(\Gamma)$ values, but in the game to the right there is no strictly dominant strategy or pure Nash equilibrium, suggesting that higher cooperative (*Top, Left*) play may be expected.

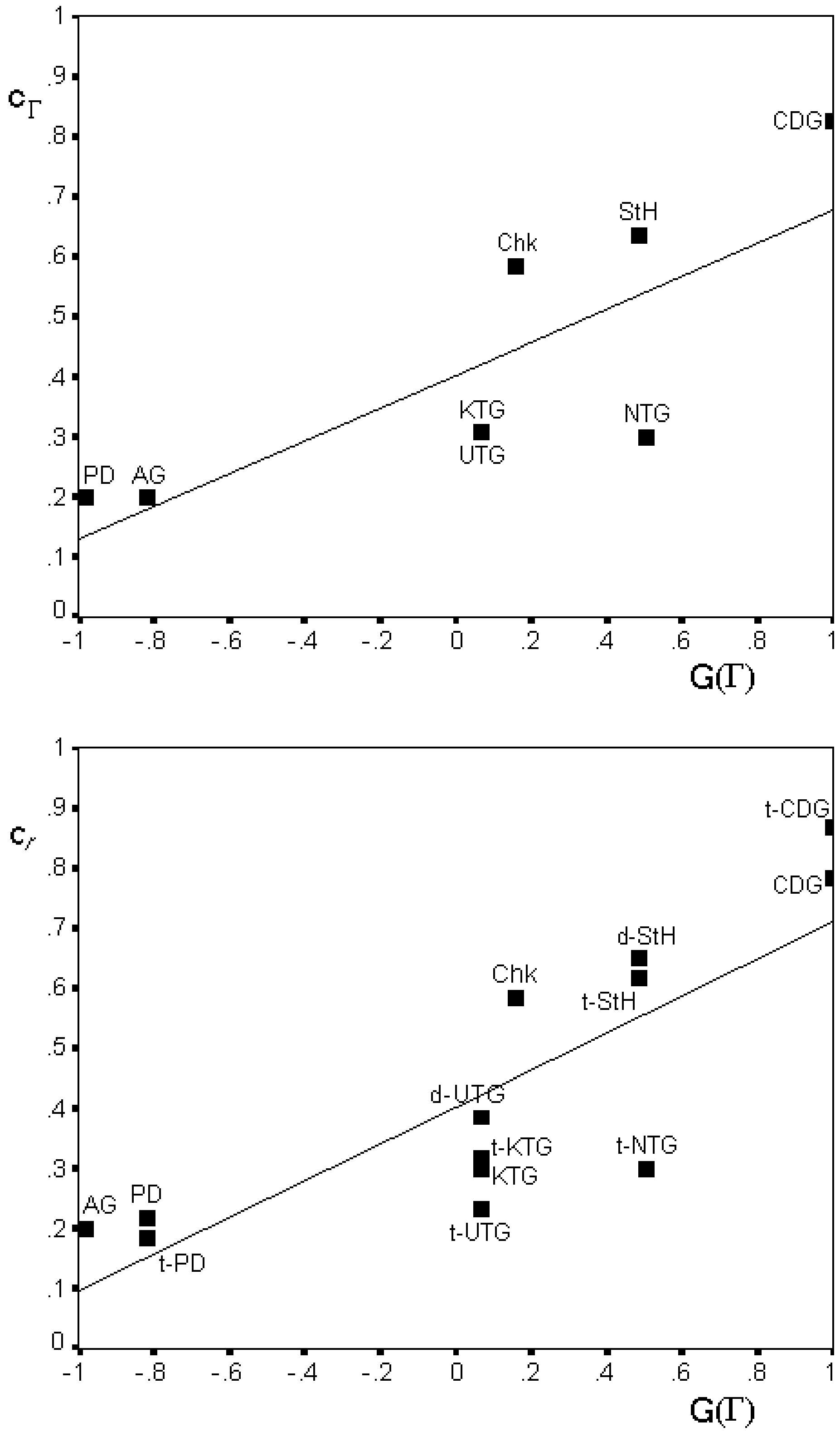
FIGURE 1. GAME HARMONY AND MEAN G-COOPERATION IN EXPERIMENT A





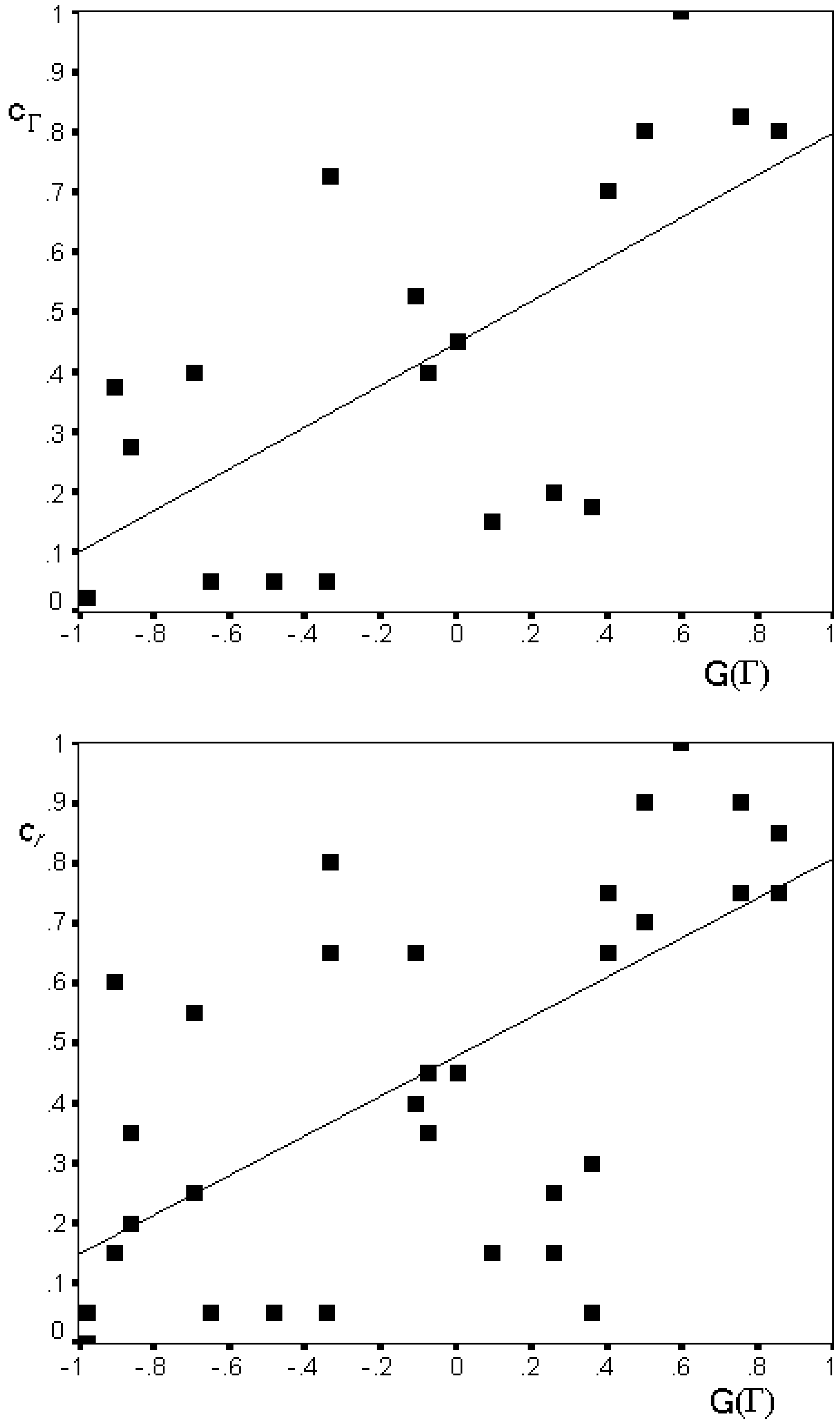
Notes: c_Γ (c_r) is the mean g-cooperation rate for each game (role). AG: Altruism G.; CDG: Coordination G.; ChK: Chicken; EG: Envy G.; PD: Prisoner's Dilemma; KTG/NTG/UTG: Kind/Needy/Unequitable Trust G., respectively; StH: Stag-Hunt. In the data by role, when cooperation rates are defined both for d-players and t-players in game Γ, d-Γ (t-Γ) is the notation for the datapoint for d-players (t-players).

FIGURE 2. GAME HARMONY AND MEAN R-COOPERATION IN EXPERIMENT A



Notes: c_Γ (c_r) is the mean r-cooperation rate for each game (role). All other notation is as for Figure 1.

FIGURE 2. GAME HARMONY AND MEAN R-COOPERATION IN EXPERIMENT B



Notes: c_Γ (c_r) is the mean r-cooperation rate for each game (role).