Preliminary Experimental Results on the Similarity Function in 2×2 and 3×3 Games

Daniel John Zizzo*
Department of Economics and Christ Church College
University of Oxford
daniel.zizzo@economics.ox.ac.uk

(Version October 11, 2002)

Abstract

This paper employs two experimental datasets to assess along what dimensions agents evaluate whether two 2×2 or 3×3 games are similar. The Euclidean distance between the payoffs, the absolute difference in game harmony and the difference in mean and standard deviation between the games are all significant components of the estimated similarity function, and so is whether the two games have zero, unique or multiple pure Nash equilibria. However, our explanatory power is limited: substantial extra information can be gathered by collecting similarity evaluation data rather than relying just on our observable proxies.

JEL Classification Codes: C72, C91.
Keywords: similarity, game harmony, cognition.

* Thanks to Jahirul Islam and Jonathan Tan for experimental assistance. The usual disclaimer applies. Financial assistance from the George Webb Medley Fund, Oxford, is gratefully acknowledged. The experimental instructions can be found in Zizzo and Tan (2002) in relation to the 2 × 2 games experiment and in Zizzo (2002) in relation to the 3 × 3 games experiment; they are also available on request.

1. Introduction

This paper presents some preliminary results from experiments where subjects were asked to rate on a numerical scale the similarity of 2×2 or 3×3 games to other games. It may be of interest to know when we should expect a decision problem to be perceived as similar to another. For example, Gale et al. (1995) suggested that play in ultimatum games in the short run may be driven by familiarity with games the agent has had lots of exposure to outside the lab. Zizzo and Tan (2002) found that similarity ratings could be used as partial predictors for cooperative behavior between 2×2 games, depending on how similar subjects evaluated a game to a cooperative game relative to a constant-sum game. Jehiel (1999) recently built a model capable of explaining, among other things, the role of poor outside options in a non-cooperative game-theoretical context, by assuming that agents reason by analogy between several situations. In relation to individual choice, Gilboa and Schmeidler (2001) described a model of case-based reasoning where agents choose an action according to the similarity of the decision problem to others faced in the past. Rubinstein (1988) applied the idea of similarity to explain the Allais paradox in an individual choice setting, and Buschena and Zilberman (1995) presented experimental evidence where this approach outperformed many theories of decision under risk. Yet, any model based on similarity requires a specification of the similarity function in order to be truly testable and falsifiable; no attempt of this kind has been made in relation to games, albeit one in relation to choices between binary lotteries exists (Buschena and Zilberman, 1999).

This short note considers the similarity choices made by subjects in Stage 3 of the 2×2 games experiment described in Zizzo and Tan (2002) and of the 3×3 games experiment described in Zizzo (2002a, Treatments 1-7). The question that we address here and not in those papers is to try to explain how similarity judgements are formed.

If we find that we are able to explain a large portion of the variance in similarity judgements using observables rather than actually asking for them, then we would gain in our ability to predict when two games are similar and so when we may expect assimilation of a game to another to occur. If, conversely, this were not the case, this would point to the importance of employing similarity judgements to gather empirical data that can be used to test similarity and analogy-based theories. It would also suggest the need for more conceptual work on what are the relevant dimensions of similarity between games.

Our analysis is preliminary in many respects. We mention three. First, the experiment with 2×2 games and that with 3×3 games were not fully homogenous, as it will be apparent discussing the experimental design: this may have reduced the power of our regression analysis. Second, and only in part because of this, our estimated specification of the similarity function can explain only about 10% of the variance in similarity evaluations. Third, particularly in the light of the existing literature, it would be interesting to expand the gathering of similarity evaluations to game-theoretical contexts other than 2×2 or 3×3 games. Hopefully this note will help stimulating further research in the determination, estimation and usefulness of similarity functions.

Section 2 describes the experimental design. Section 3 sets up our regression analysis and section 4 analyzes the results and concludes.

2. Experimental design

The experiments were run between May 2001 and May 2002 in Oxford, in session of four subjects each.[1] The 3×3 games experiment had 112 subjects;[2] the 2×2 games experiment had

[1] Full details can be found in Zizzo and Tan (2002) and Zizzo (2002).

60 subjects. The two experiments had a common structure. Payoff values were always specified as numbers between 0 and 100. In Stage 1, subjects did practice by choosing actions with (2×2 or 3×3) games for (ten or six, respectively) rounds; each player was matched with the same coplayer throughout the practice, and received feedback on the play outcome after each round. In Stage 2, subjects chose actions in newly faced (2×2 or 3×3) games, for (twenty or ten, respectively) rounds. They received no feedback after each round and knew they were paired with a different coplayer relative to that in the practice stage. Stage 3 is the focus of this paper, as subjects were asked to evaluate the similarity between games.

In the 2×2 games experiment, they had to rate how similar a game was to one of two comparison games ("comparison decision tables" or CDTs) on a scale between 0 and 9 (with 0 as maximum and 0 as minimum, following Buschena and Zilberman, 1999). There were fourty rounds. In the payment stage a round was randomly chosen and a "similarity payment" determined on the basis of the subject's decision in it: subjects were paid £ 12 to get the evaluation exactly exactly right, with a penalty of £ 5 per each point of error (with a minimum payment of £ 0). The instructions contained a table with details on the amount of payment for any given level of error, so subjects were not required to make any significant computation; also, a couple of questions in the practice questionnaire checked their understanding of the mechanism. Payment was determined at the end of the experiment and subjects received no feedback on the outcome of their similarity choices during Stage 3. For this reason, the determination of the "correct" similarity answer was an issue that had to be practically addressed in order to determine payments, but not one with serious bearing on the experiment: the risk of contamination from an arbitrary choice of "correct" answers was avoided.[3] The games and CDTs are contained in Table 1. The similarity tasks were presented in random order. Games 1-8 and the CDTs (as defined in Table 1) corresponded to games already faced in Stage 2, and some of the games were payoff-perturbed versions of games well-known to economists, such as the Prisoner's Dilemma, the Stag-Hunt and the Chicken. Subjects faced these games twice, once as in Table 1 and once in its transposed form. They also compared the CDTs once with the transpose of each CDT and once with games 9-10. There were fourty rounds overall. The CDTs themselves were a coordination game and of a constant-sum game.

In the 3×3 games experiment, subjects rated similarity on a scale between 0 and 100 (with 100 as maximum and 0 as minimum). For the randomly determined winning round, subjects were paid £ 3 to get the guess exactly right, with a penalty of £ 0.12 per each point of error (with a minimum payment of £ 0). Sixteen games were presented and compared to a single CDT, as specified in Table 1, and so there were sixteen rounds overall. Games were new to the subjects (as was the CDT), and all of them were determined by randomly drawing payoff numbers from a uniform distribution between 0 and 100. Everything else was as for the 2×2 games experiment.

3. Regression specification and variables

Table 2 contains the results of a random-effects regression analysis on *S*, a normalized similarity value. *S* is equal to the similarity value assigned by the subjects in the 2×2 games

---

[2] Two subjects are excluded from the analysis: one was later discovered to be an Oxford M.Phil. Economics student, and another one had done the 2×2 games experiment just a few weeks earlier.

[3] The correct answer was chosen by giving equal weight to the Euclidean distance between payoffs (EDP) and to whether the game had a unique pure Nash equilibrium (PNE), unlike the CDTs, or otwherwise. Let *Unique* be equal to 1 when there is a PNE, and let *max*(*EDP*) and *min*(*EDP*) the maximum and minimum EDP between a game and a CDT. Then the "correct" answer was determined as the nearest integer to $4.5 \times Unique + 4.5 \times [(EDP - min\,(EDP)] / [max(EDP) - min(EDP)]$.

experiment, and to (100 - *similarity*) × 0.09 in the 3×3 games experiment, in order to assign them the same metric as for the 2×2 games experiment (from 0 to 9, with 0 equal to maximum and 9 to minimum similarity). A Breusch-Pagan LM test for random effects is significant [χ^2 (1) = 1594.34, $P < 0.001$] while a Hauman specification test is insignificant [χ^2 (7) = 0.03, *n.s.*], so the usage of a random-effects regression specification appears appropriate.

nEDP is equal to the Euclidean distance between the payoffs[4] (EDP) for 3×3 games, and to EDP for 2×2 games, normalized for comparability to make the maximum value of EDP among 2×2 games (59.835) be the same as the maximum value of EDP among 3×3 games (229.806), and similarly in relation to the minimum values (11.704 and 78.835, respectively).

DistanceGH is the absolute difference (distance) in cardinal *game harmony* values between the games. Game harmony measures how harmonious (non-conflictual) or disharmonious (conflictual) the interests of the players are, as embodied in the game payoffs (Zizzo, 2002a, 2002b; Zizzo and Tan, 2002). At one extreme, pure coordination games are games in which the interests of the players are perfectly aligned; at the other extreme, in a two-player zero-sum game the gain of a player is always at the expense of her coplayer. All other games are somewhere in the middle. Zizzo (2002b) defines and discusses the properties of a class of game harmony measures based on the Pearson or on the Spearman correlation between the payoffs of the players. For two-player games, the simplest measure of cardinal harmony of game Γ is indeed just the Pearson correlation between the payoffs a_w and b_w of players *A* and *B*, respectively, for each possible game outcome *w*, so $G(\Gamma) = r\ (a_w, b_w)$. The ordinal measure $G_\rho(\Gamma)$ can be found simply by replacing Pearson with Spearman. It is obvious that these measures are bounded between -1 and 1, with -1 for any constant-sum game and 1 for any coordination game; Table 1 reproduces $G(\Gamma)$ for all the games listed.

SameNumberNE is a dummy equal to 1 if the game has zero, unique or multiple PNE depending on whether the CDT has zero, unique or multiple PNE, respectively. *DMean* (*DSD*) is the difference between the means (standard deviations) of the two games. *DNumber* is equal to 1 minus the fraction of payoff numbers that are in common between the game and the CDT.

Three variables control for the experimental context. *Experiment* is a dummy equal to 1 for the the 2×2 games experiment. The 3×3 games experiment had treatments where, in Stage 2, some payoff cells were hidden and in relation to which subjects were asked to guess incentive-compatibly what the payoff values were, before playing each game; there was also a treatment where some payoff cells were hidden but there was no payoff-guessing. *HiddenCell* is equal to 0 whenever the game matrix was fully displayed (always in the 2×2 games experiment, sometimes in the 3×3 games experiment), to 1 when some payoff cells were hidden but there was no guessing task, and to 2 when there was also payoff-guessing. In both experiments, the average harmony of the practice stage games [as measured by $G(\Gamma)$ or $G_\rho(\Gamma)$] was high, intermediate or low according to the treatment: *PracticeGH* is 0 for the low game harmony treatment, 1 for the intermediate case and 2 for the high case.

There are also some demographical variables: *DAge* is equal to age minus the mean age in the subject sample; *Sex* is 1 for males, and 0 otherwise; *HardSciences*, *Humanities* and *Economics* are 1 respectively for subjects with a hard sciences, humanities and economics (or business) background, and 0 otherwise.

[4] Buschena and Zilberman (1999) use Euclidean distances. An Euclidean distance is the square root of the sum of the squared differences between items over all the relevant dimensions.

4. Results and conclusions

The experimental context variables and all the demographical variables except for a humanities background (which led to higher similarity evaluations) are insignificant. *nEDP* is significant: subjects look at how different the payoffs look like in order to determine the similarity between the games. This result mirrors an analogous finding, in relation to lotteries, by Buschena and Zilberman (1999). Subjects also appear sensitive to differences both in the mean and in the standard deviations between the games.

The most interesting significant variables are probably *SameNumberNE* and *DistanceGH*. Economists sometimes classify games by the availability of pure Nash equilibria, and so, apparently, do subjects. Subjects also appear to be sensitive to the distance in game harmony between the games (replacing the distance in $G(\Gamma)$ with the distance in $G_\rho(\Gamma)$ makes no estimation difference): they consider two games as more different the greater their difference in terms of game harmony.

However, one needs to recognize that the R^2 is very low: we can only explain some 10% of the variance in similarity evaluations.[5] In part, this may be due to the fact that we are putting together the 2×2 and 3×3 games datasets: by running separate regressions on the two datasets, one can raise the R^2 to 0.17-0.20, but this solution looks clearly *ad hoc*, and anyway only mildly successful. Obviously, further experiments with homogenous protocols for the 2×2 and the 3×3 games experiments would help. Even so, it is a safe conclusion that we can gather substantial extra information by collecting similarity evaluation data rather than relying just on observable proxies. More conceptual work is required on what are the relevant dimensions of similarity between games, and this work must pass the acid test of systematic experimentation.

References

Buschena, D., and D. Zilberman (1995) "Performance of the Similarity Hypothesis Relative to Existing Models of Risky Choice" *Journal of Risk and Uncertainty* **11**, 233-262.

Buschena, D., and D. Zilberman (1999) "Testing the Effects of Similarity on Risky Choice: Implications for Violations of Expected Utility" *Theory and Decision* **46**, 251-276.

Gale, J., K.G. Binmore, and L. Samuelson (1995) "Learning to Be Imperfect: The ultimatum game" *Games and Economic Behavior* **8**, 856-90.

Gilboa, I., and D. Schmeidler (2001) *A Theory of Case-Based Decisions*, Cambridge University Press: Cambridge.

Jehiel, P. (1999) "Predicting by Analogy and Limited Foresight in Games" mimeo.

Rubinstein, A. (1988) "Similarity and Decision-Making Under Risk (Is There a Utility Theory Resolution to the Allais Paradox?)" *Journal of Economic Theory* **46**, 145-153.

Zizzo, D.J. (2002a) "Game Perception and Harmony in 3 × 3 Games: An Experiment in Cognitive Game Theory" Department of Economics discussion paper n. 118, University of Oxford.

Zizzo, D.J. (2002b) "On the Measurement of Harmony in Normal-Form Games" Department of Economics discussion paper n. 116, University of Oxford.

Zizzo, D.J., and J.H.W. Tan (2002) "Game Harmony as a Predictor of Cooperation in 2 × 2 Games" Department of Economics discussion paper n. 117, University of Oxford.

[5] 0.182 of this explained variance is due to the random effects component of the regression model.

Table 1: Experimental games and game harmony values $G(\Gamma)$

	Column Player		
Row Pl.	A, B	C, D	E, F
	G, H	I, J	K, L
	M, N	O, P	Q, R

	Number	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	$G(\Gamma)$
CDT 2x2	1	74	75	32	31			13	12	85	86									1
	2	71	31	18	84			26	76	89	13									-1
2x2 Games	1	92	11	38	37			64	63	10	93									-0.817
	2	61	72	59	73			50	28	48	29									0.98
	3	59	28	61	29			48	72	50	73									-0.98
	4	33	34	34	35			81	82	14	100									0.066
	5	52	3	53	4			100	51	33	69									0.066
	6	3	34	4	35			81	82	14	100									0.503
	7	10	51	92	93			52	53	52	9									0.488
	8	92	36	10	11			63	62	38	94									0.16
	9	61	62	17	16			4	3	70	71									1
	10	36	82	93	25			81	37	30	88									-1
CDT 3x3	1	5	56	88	52	72	43	36	61	15	25	57	41	86	22	23	23	77	28	-0.12
3x3 Games	1	94	21	97	51	80	29	44	72	72	43	46	5	60	71	20	23	96	91	0.278
	2	29	84	39	1	56	78	92	64	33	29	31	93	1	74	25	82	52	45	-0.138
	3	99	75	55	75	83	69	69	46	4	93	76	71	74	82	74	1	64	37	-0.292
	4	28	65	64	83	25	12	85	92	78	74	41	49	74	56	8	42	13	30	0.766
	5	28	80	11	85	72	25	93	48	39	86	26	67	83	39	37	70	85	32	-0.858
	6	19	77	67	22	95	77	65	7	9	93	23	23	90	17	50	64	6	45	-0.343
	7	90	84	4	42	9	64	92	17	99	65	73	25	5	64	60	10	2	69	-0.228
	8	46	98	64	47	61	48	13	77	32	17	78	54	52	18	44	18	62	30	-0.149
	9	86	50	94	84	53	34	75	60	91	88	23	21	29	49	73	70	40	13	0.814
	10	50	45	1	86	15	75	74	60	95	94	51	67	23	99	60	92	13	93	-0.199
	11	85	64	97	74	22	56	1	71	48	83	55	97	88	88	88	65	34	59	0.237
	12	28	60	27	20	25	52	33	96	87	10	94	87	58	7	25	93	47	58	-0.207
	13	46	98	64	47	61	48	31	49	32	17	78	54	86	36	44	18	62	30	0.063
	14	43	82	95	78	49	5	28	88	24	100	72	27	93	81	18	73	76	34	-0.26
	15	12	93	84	22	93	62	1	100	59	59	47	62	15	93	81	57	87	34	-0.879
	16	18	12	69	42	42	13	28	86	16	41	36	24	14	5	6	74	73	90	0.266

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. CDT stands for Comparison Decision Table.

Table 2. Random-effects regression on normalized similarity values *S* (*n* = 4120)

	Coef.	SE
Experiment	-0.099	0.221
PracticeGH	-0.039	0.104
HiddenCell	0.022	0.112
SameNumberNE	-0.818*	0.075
nEDP	0.014*	0.001
DistanceGH	0.713*	0.058
DMean	0.034*	0.007
DSD	0.05*	0.007
DNumber	0.211	0.188
DAge	0.017	0.017
Sex	0.067	0.163
HardSciences	0.059	0.224
Humanities	0.717*	0.24
Economics	0.191	0.25
Constant	2.552*	0.336
Within R^2	0.121	
Between R^2	0.082	
Overall R^2	0.116	

* indicates significance at $P < 0.001$.