A model of procedural and distributive fairness

Michal Wiktor Krawczyk*

February 6, 2007

Abstract *This paper presents a new model aimed at predicting behavior in games involving a randomized allocation procedure. It is designed to capture the relative importance and interaction between procedural justice (defined crudely in terms of the share of one's expected outcome in the sum of all expected outcomes) and distributive justice (reflecting the relation of the actual outcome to the sum of all outcomes). The model is applied to experimental games, including "randomized" variations of simple sequential bargaining games, and delivers qualitatively correct predictions. I also show that in view of the model redistribution of income can be seen as a substitute for (lacking) vertical social mobility. This contributes to the explanation of greater demand for redistribution in European countries vis-a-vis the United States. I conclude with suggestions for further verification of the model and possible extensions.*

1 Introduction

In recent years growing experimental evidence of the low predictive power of models based on the classical homo economicus assumptions has fuelled theoretical advances in economic models admitting deviations from selfish rationality (see e.g. Fehr and Gächter, 2000 for an overview). Two types of models that have attracted special attention are Social Utility Models and intention-based models. Papers in the first group, notable examples of which are Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) assume that (some) decision makers display other-regarding preferences, i.e. they

*CREED, Universiteit van Amsterdam, 11 Reotersstraat, 1018WB Amsterdam. E-mail: m.w.krawczyk@uva.nl. I would like to thank Martin Kocher, Arno Riedl and especially Frans van Winden, as well as participants of the IAREP-SABE 2007 Conference in Paris and 17th International Conference on Game Theory at Stony Brook for their helpful comments.

are assumed to care about other's payoffs. While this approach proved to be helpful in understanding various "anomalies" in the observed behavior, critics were quick to point out its weaknesses. These essentially stemmed from the fact that these models maintain the consequentialist perspective (see Sen, 1979; Loewenstein et al., 2001). They focus exclusively on the outcomes of social interactions (preferably measured in monetary terms), while disregarding the way these outcomes come about. As a result, in experiments manipulating intentionality of the other players' actions, they delivered a rather poor proxy for varying emotional responses and resulting behaviors. Further, it was argued, fairness standards that supposedly regulated behavior were not truly standard, i.e. they varied substantially with rather subtle changes in the environment (see Reuben and van Winden, 2005) and especially with the role assumed in the game (see Babcock and Loewenstein, 1997; and Schmitt, 2004, on self-serving bias).

The other approach that sought to avoid these pitfalls was pioneered by Rabin (1993), followed by Dufwenberg and Kirschteiger (2004) and others. In this line of research perceived intentions are explicitly incorporated and are assumed to trigger corresponding responses (positive or negative reciprocity) mediated by emotional reactions. These models are generally more complicated and often run into problems of multiple equilibria. In the meantime, some interesting attempts to unify the two approaches have been made (see e.g. Falk and Fischbacher, 2001, and Charness and Rabin, 2002).

In this paper I try to develop a model that, while maintaining the simplicity of Social Utility Models, allows modeling at least some instances of importance of intentionality. To this end, I drop the consequentialist perspective by explicitly incorporating a new term into the utility function which captures the way in which outcomes are generated. Thanks to this, procedural considerations, including procedural justice, may be accounted for.[1]

Procedural justice is about transparent and impartial rules ensuring that

[1] I use the term "procedure" quite broadly, to refer to all aspects of the way in which the actual outcome is arrived at. In particular, I disregard here the potentially important distinction between procedural and "interactional" justice, the latter denoting a social aspect of implementing a procedure. As Cropanzano et al. (2002) note, "it is not yet clear whether procedural and interactional justice are separate constructs". For example, Cropanzano and Greenberg (1997) argue for considering procedural and interactional fairness as parts of the same process. This simplification helps to keep the model parsimonious. Ultimately, decision as to whether or not to include interactional fairness as a separate variable should be based on a comparison of the predictive power of respective models.

each of the agents involved in an interaction enjoys an equal opportunity to obtain a satisfactory outcome. To make the model functional and generate predictions for laboratory games, it seems justified to operationalize procedural justice in what can initially appear to be a crude way. Namely, procedural justice is proxied by the expected share of the "pie" for given strategies of all players, where expectation is taken over all possible moves by "nature".[2] Although procedural fairness admittedly goes beyond the neutrality of random procedures, an unbiased lottery is a prototypical example of a procedurally fair interaction (cf. Rawls, 1971, p. 86). It is indeed transparent, impartial and gives everyone a fair share in the long run.

The expected share in the game will typically be influenced by the choices of the fellow players. Thus, from an individual's viewpoint, the expected share of the pie can be thought of as what the other person was willing to award them (on average), a proxy for intentions. Hence, participants in the interaction are allowed to react to "good" or "bad" intentions displayed in the behavior of others. In this way my model, which may be viewed as a parsimonious extension of the social utility model by Bolton and Ockenfels, allows for indirect modelling of reciprocity-type behavior. The predictive power is thus substantially improved.

To name a few examples, the model explains the importance of procedural considerations in dictator games (Karni et al., 2001) and in shrinking-pie sequential-bargaining games, where fair random role assignment creates an entitlement to a large portion of the pie (Sonnegard 1996).[3] Still, it shows that distributive justice *per se* matters as well (Kroll and Davidovitz, 2003). It also accounts for the correlation between conditional gifts and expectations in solidarity game (Selten and Ockenfels, 1998).

Perhaps most important, though, the model explains some interesting results in Blount (1995), Kramer et al. (1995), Falk et al. (2000) and Bolton et al. (2005) where subjects conditioned their behavior in simple bargaining games on the fairness of the mechanism producing initial offers. Morover, they did it in ways that are difficult to explain within current

[2]Needless to say, this definition is relevant only for games that indeed involve random moves. Otherwise the expected share coincides with the actual share.

[3]In related studies, Hoffman & Spitzer (1985) compare the random distribution of bargaining roles with the assignment based on results of game of NIM, which, however is not truly a game of skill, as the first mover can guarantee winning by using simple backward-induction. Hoffman et al. (1994) only find a substantial ($p < 0.01$) effect of contest entitlement (based on a trivia quiz) combined with a set of instructions presenting the problem in a market context and underlying that the winner of the contest has earned the right to take the favorable position. As Bearden (2001) sums up "The results on P1 [proposer] legitimacy and ultimatum offers are mixed."

models incoporating the perceived intentions of the interaction partner.

An interesting field application of the model pertains to the observed relationship between perceived equality of opportunity in the society and the support for redistribution via taxation (Alesina et al., 2004)

The rest of the paper is organized as follows. In Section 2 the formal model, using the Psychological Game Theory (PGT) framework, is formulated. Section 3 shows how several experimental and empirical findings can be explained using the model. Section 4 contains a short review of the relevant theoretical literature, mostly developed by psychologists, which provides background information and justification for the assumptions incorporated in the model. Section 5 concludes with some suggestions for future research. Details of the PGT approach are explained in Appendix A whereas all proofs can be found in Appendix B.

2 The model

Adapting the concept of self-centered fairness (see e.g. Bolton and Ockenfels 2000), I assume that a "sufficient statistic" for an agent to assess the value of particular play of a game involving a lottery is a triple of terms (capturing three different motives): his actual payoff (representing self interest), share of his payoff in the sum of all payoffs (as a proxy for the perceived distributive fairness) and share of his expected payoff in the sum of all expected payoffs, determined by strategies of all players (as a proxy for the perceived procedural fairness). The two latter terms must then be compared to their reference points – values that people perceive as appropriate in given situation.

My central hypothesis based on the insights from psychology is that people care about fairness of both procedures and outcomes and that these dimensions of justice are interconnected (I shall refer to it as interaction hypothesis). In a way that will be formalized in the following subsections, expected payoffs and actual payoffs are substitutes to each other when judging overall fairness of the situation (that is, of the procedure, given its results): the more given play of a game is procedurally favorable to some player – the higher his expected share – the less he is concerned about increasing his actual share, holding monetary incentives fixed.[4]

The model predicts in particular that people will deem a physical division of an object (if it results in no or only modest efficiency loss) most fair (it is ideal on both fairness dimensions); will trade-off procedural justice

[4] Some psychological justification of this assumption is offered in Section 4.

for distributive justice (e.g. even a risk-neutral individual might prefer an almost-equal physical split to a fair lottery), will generally accept a bad outcome more easily if it results from a fair or advantageously biased procedure (than from a disadvantageously biased one); conversely, will consider a good outcome more fair if it resulted from a disadvantageously biased procedure.

Before I develop the formal model, let us look briefly at some predictions resulting from it, which, on qualitative level, can be derived from the underlying interaction hypothesis. When adapted to preference for redistribution, it suggests that members of open societies with high inter- and intragenerational mobility, thus offering plenty of opportunities for moves up (as well as down) the social ladder will be in favor of relatively low taxes and social benefits. This is because a bad outcome (belonging to a low-income, low-status social class) is acceptable if it results from a fair "procedure". On the contrary, a lower class member who is believed to have been bound to become one, due to unfair constraints on social mobility existing in a society, deserves substantial support from the state (so, indirectly, from his fellow citizens facing more attractive opportunities). This fits stylized facts on differences between the USA and Europe (Alesina et al., 2004, Alesina and Angeletos, 2005).

With regard to experimental evidence, my model predicts that intentions of other players signalled in their choices affecting expected payoff, will influence attractiveness of particular outcomes and thus will affect behavior, even if they are unimportant from strategic point of view. The most clear example of this are the results on perturbed ultimatum game due to Kramer, et al. (1995), where bad outcome was more acceptable if resulted from a more fair procedure – i.e. from one in which the proposer intended to be relatively generous.

Similarly, interaction between procedural and distributive justice predicted by the model is apparent in Bolton et al. (2005) and Cox and Deck (2005). Here, again, an explicitly just procedure, i.e. a procedure giving the responder a fair share on average, substantially lowered frequency of latter's rejections, even if resulting allocation was highly unequal.

Generally, I believe that the model generates correct predictions in a wide range of settings, both in the field and in the lab. Particularly, even though parsimonious form of social utility models is kept[5] instances where "intentions mattered" may be captured.

[5] Contrary, e.g., to the model by Rabin (1993), beliefs (except for behavior strategies modeled here as first-order beliefs) do not enter the utility function.

2.1 Actual and expected shares

To approach the problem of relative importance of actual and expected share, I shall define for every terminal history z and every agent $i = 1, 2, ...n$ the actual share σ_i in the total material payoff (TMP) at z:

$$\sigma_i = \sigma_i(z) = \begin{cases} y_i/c \ , \ \text{if } c > 0 \\ 1/n \ , \ \text{if } c = 0, \end{cases} \tag{1}$$

where y_j is player j's material payoff (assumed to be non-negative) and $c = \sum_{j=1}^{n} y_j(z)$ is the TMP at z. Further note that behavior strategy vector (composed of players' strategies δ and nature's strategy δ_0 and defined in Appendix A) determines the vector of expected payoffs of all players $(y_1^E, y_2^E, ...y_n^E)$ in the game, whereas expectation is taken over all possible choices of nature:

$$y_j^E = y_j^E(\delta, \delta_0) = \sum_{z'} P(z'|\delta, \delta_0) y_j(z'), \tag{2}$$

where $P(z'|\delta, \delta_0)$ denotes the probability of arriving at terminal history z' given the behavior strategies δ_0 and δ. Thus y_j^E is what player's j expected payoff at the onset of the game was, given behavior of all the players and nature.

Let now σ_i^E denote the ratio of the expected payoff of player i to the sum of all expected payoffs:

$$\sigma_i^E = \sigma_i^E(\delta, \delta_0) = \begin{cases} y_i^E/c^{E,i} & \text{,if } c^E > 0 \\ \quad 1/n & \text{, if } c^E = 0, \end{cases} \tag{3}$$

where $c^E(\delta, \delta_0) = \sum_{j=1}^{n} y_j^E(\delta, \delta_0)$ is the expected value of c.

Obviously $0 \leq \sigma_i \leq 1$ and $0 \leq \sigma_i^E \leq 1$. Further $\sigma_i^E(\delta, \delta_0) = 1$ implies that $\sigma_i(z) = 1$ for every terminal history z which obtains under δ, δ_0 with positive probability. Similarly $\sigma_i^E(\delta, \delta_0) = 0$ implies that $\sigma_i = 0$ for relevant end-nodes. Note also that, generally expected share is *not* equal to the expectation of the actual share.

2.2 Motivation function

The question that remains to be answered is how individuals evaluate and aggregate different dimensions of the allocation process and its outcomes to arrive at the final assessment representing preferences. In general, it is assumed that every agent aims at maximizing the expected value of the motivation function:

$$v_i = v_i(y_i, \sigma_i, \sigma_i^E). \tag{4}$$

For mathematical convenience I shall assume that v_i is twice continuously differentiable. Further, I shall assume that, holding the actual and the expected share constant, the agents prefer more money to less and (may) display risk aversion:

Assumption 1 *Self-interest.*

$$\begin{aligned} v_{i1}(y_i, \sigma_i, \sigma_i^E) &> 0, & (5)\\ v_{i11}(y_i, \sigma_i, \sigma_i^E) &\leq 0. & (6) \end{aligned}$$

The next Assumption reflects the fact that individuals have preference for both distributive and procedural justice. It closely resembles one of Bolton and Ockenfels' (2000) assumptions and readers are referred to the short discussion there. Let $RP_i \in [0,1]$ ($RP_i^E \in [0,1]$) denote the actual (expected) share in the division problem that is considered most appropriate by player i. If not stated otherwise, I shall assume that $RP_i^E = RP_i = \frac{1}{n}$.

Assumption 2 *Fairness Consideration. Motivation function is concave in the actual share and in the expected share. If the expected (actual) share is equal to its reference point, RP_i (RP_i^E), motivation function is maximized when the actual (expected) share is also equal to its reference point:*

$$\begin{aligned} \frac{\partial^2 v_i(y_i, \sigma_i, \sigma_i^E)}{\partial \sigma_i^2} &< 0, & (7)\\ \frac{\partial^2 v_i(y_i, \sigma_i, \sigma_i^E)}{\partial (\sigma_i^E)^2} &< 0, & (8)\\ \frac{\partial v_i(y_i, \sigma_i, \sigma_i^E)}{\partial \sigma_i}\Big|_{\sigma_i^E = RP^E} &= 0 \textit{ if } \sigma_i = RP, & (9)\\ \frac{\partial v_i(y_i, \sigma_i, \sigma_i^E)}{\partial \sigma_i^E}\Big|_{\sigma_i = RP} &= 0 \textit{ if } \sigma_i^E = RP. & (10) \end{aligned}$$

The next Assumption is crucial for the model. It reflects the interaction effect.

Assumption 3 *Procedural-Distributive Justice Interaction. Marginal motivation value of the actual share is non-increasing in the expected share:*

$$\frac{\partial^2 v_i(y_i, \sigma_i, \sigma_i^E)}{\partial \sigma_i \partial \sigma_i^E} = v_{i23}(y_i, \sigma_i, \sigma_i^E) \leq 0, \tag{11}$$

Remark 4 *Marginal motivation value of expected share is non-increasing in the actual share.*

The last assumption places natural restrictions on interaction between material interest and preference for justice.

Assumption 5 *Justice-Payoff Interaction. Absolute value of the marginal motivation value of the expected share is non-decreasing in material payoff. However, a pure transfer from other player does not increase the marginal motivation value of the expected share:*

$$\begin{aligned} v_{i31}(\cdot) &\geq 0 \text{ } \textit{if } v_{i3}(\cdot) > 0, && (12) \\ v_{i31}(\cdot) &\leq 0 \text{ } \textit{if } v_{i3}(\cdot) < 0, && (13) \\ v_{i31}(\cdot) + \frac{1}{TMP^E} v_{i32}(\cdot) + \frac{1}{TMP^E} v_{i33}(\cdot) &\leq 0. && (14) \end{aligned}$$

As a justification of the seemingly peculiar conditions in Assumption 5, consider a game with no randomization involved. Suppose the total size of the pie is increased, while expected share of player i (and, consequently, actual share, since they coincide) stays constant. I postulate that an individual will care about procedural fairness in such "scaled up" problem at least as strongly as she did in the "scaled down" problem.[6] This is exactly what inequalities (12) and (13) in Assumption 5 guarantee.[7] What regards inequality (14), consider a game with no randomization involved. Consider change in strategies resulting in a pure transfer from some player j to player i (i.e. size of the pie is held constant). This should not make the benefiting player i demand more of the expected share. Thus the first term in (14) is

[6] It says nothing however about the relative important of justice vs. self interest.

[7] Similar assumption regarding importance of distributive justice is, in my view, also plausible.

either negative or smaller or equal in absolute value then the sum of the last two (non-positive, see Assumptions 2 and 3) terms.

A very simple example of function satisfying the assumptions can be obtained by slight modification of a linear-quadratic function proposed by Bolton and Ockenfels (2000):

$$v_i(y_i, \sigma_i, \sigma_i^E) = v_i - b(w\sigma_i + (1-w)\sigma_i^E - 1/2)^2, \tag{15}$$

$v_{i2}(y_i, \sigma_i, \sigma_i^E) = v_i - 2bw(w\sigma_i + (1-w)\sigma_i^E - 1/2),$
$v_{i23}(y_i, \sigma_i, \sigma_i^E) = v_i - 2bw(1-w)\sigma_i^E,$

Here, the $w \in (0,1)$ parameter measures relative importance of distributive considerations vis a vis procedural justice, while $b(>0)$ captures importance of fairness in general when compared to the self-interest motive. Clearly, $v_{i1} > 0, v_{i11} = 0$ (Assumption 1), $v_{i22}, v_{i33} < 0$, $v_{i2} = v_{i3} = 0$ when $, \sigma_i = \sigma_i^E = 1/2$ (Assumption 2), $v_{i23} < 0$ (Assumption 3) and $v_{i31} = 0$ combined with previous results trivially satisfies Assumption 5.

It can be easily seen that assumption 3 plays an important role in establishing further results regarding effects of procedural and distributive justice

Lemma 6 *For fixed actual (expected) share and given monetary payoff unique optimal expected (actual) share exists, that is:*

(1) for each σ_i and y_i there exists $\sigma_i^{E*} = \sigma_i^{E*}(\sigma_i, y_i) \in [0,1]$ such that $\frac{\partial v_i(y_i,\sigma_i,\sigma_i^E)}{\partial \sigma_i^E} > 0 \ \forall \sigma_i^E \in [0, \sigma_i^{E*}]$ and $\frac{\partial v_i(y_i,\sigma_i,\sigma_i^E)}{\partial \sigma_i} < 0 \ \forall \sigma_i^E \in [\sigma_i^{E*}, 1]$.

Obviously, σ_i^{E*} maximizes motivation function for given y_i and σ_i. Similarly:

(2) for each σ_i^E and y_i there exists $\sigma_i^* = \sigma_i^*(\sigma_i^E, y_i) \in [0,1]$ such that $\frac{\partial v_i(y_i,\sigma_i,\sigma_i^E)}{\partial \sigma_i} > 0 \ \forall \sigma_i^E \in [0, \sigma_i^*]$ and $\frac{\partial v_i(y_i,\sigma_i,\sigma_i^E)}{\partial \sigma_i} < 0 \ \forall \sigma_i \in [\sigma_i^*, 1]$

and again σ_i^* maximizes motivation function for given y_i and σ_i^E. Further,

(3) optimal expected share does not depend on y_i

$$\sigma_i^{E*}(\sigma_i, y_i) = \sigma_i^{E*}(\sigma_i, y_i')\forall_{y_i, y_i'} \text{ and} \tag{16}$$

(4) optimal actual (expected) share is non-increasing in expected (actual) share. In particular, optimal actual (expected) share is not greater than RP (RP^E) if expected (actual) share is greater than RP^E (RP) and not smaller than RP (RP^E) if expected (actual) share is greater than RP^E (RP).

Proof. See Appendix B. ■

Before applying the theory to particular examples I should note that existence of equilibrium is warranted.

Theorem 7 *(Theorem 9 in BD). There exists at least one PDF equilibrium.*

Proof. Results immediately given continuity of the PDF function. ■

3 Applications of the model

Note that if a player is concerned solely with the distributive justice or when there is no risk involved in the game (so $\sigma_i = \sigma_i^E$), the model essentially reduces to Bolton and Ockenfels' ERC. Therefore the predictive power of this concept carries over to my model. More important, it captures some other regularities observed in laboratory games involving risk. Below I mention some of them.

Of course, in every case, rules imposed by the experimenter define only the material payoff game. I implicitly construct a dynamic psychological game by equipping subjects with motivation function satisfying assumptions discussed in the previous section. Equilibrium analysis is fairly simple and intuitive in the case of sequential ultimatum-type games. I am particularly concerned with responder's strategies. What regards (simultaneous move) solidarity game, I stick to non-equilibrium analysis of the link between actions and beliefs.[8]

The results are organized in accordance with the type of game to which they relate. I first analyze ultimatum and other simple bargaining games, then solidarity game and finally, and finally, the issue of redistribution considered mostly in non-experimental literature.

3.1 Ultimatum and related games

Consider the ultimatum game in which in the first stage one of the players (called proposer) makes an offer regarding division of the pie of size c. I allow for the possibility that the offer be randomly modified (or, in the

[8] We have to note that most experiments discussed here used a strategy method, which is innocuous in my model due to dynamic inconsistency that may arise in psychological games. Subjects may use this method as a commitment device and provide a strategy that they would not want to follow if they could, say, actually respond to a particular offer in Ultimatum Game. Because allowing for this kind of strategic commitment substantially complicates the analysis and because there is only limited evidence that use of the strategy method may alter experimental results, for the time being I neglect this consideration. All propositions are thus formulated and proved under the simplifying assumption that (following experimenters' best efforts) subjects behave *as if* they were actually facing given offer. We note that explicitly accounting for the use of strategy method does not invalidate the results. Proofs however are more tedious and are omitted here

extreme case, generated randomly, call it a Random Offer Ultimatum Game, ROUG). In the second stage the other player (respondent) learns the offer (and possibly the pre-perturbation offer) and either accepts or rejects it, setting rewards to 0 for both. I will denote the share claimed for proposer by Xc (capital letter indicating a random variable, $0 \leq X \leq 1$), while remaining $(1-X)c$ is to be offered to the respondent.

I first derive some predictions for the UG and then discuss corresponding empirical findings. My first result determines the general pattern of responder's behavior in ROUG

Statement 8 *In ROUG for each random mechanism and each responder there is a set $A = (1-\overline{x}, 1-\widetilde{x})$ with $0 < 1-\overline{x} < 0.5 < 1-\widetilde{x} \leq 1$ such that offers of from A are always accepted, offers of $1-\overline{x}$ and $1-\widetilde{x}$ are accepted with probabilities $p, q \in [0,1]$ respectively and other offers are rejected.*

Proof. See Appendix B. ■

In the following I will sometimes refer to $1-\overline{x}$ as the (lower) rejection threshold and to $1-\widetilde{x}$ as the upper rejection threshold.

Following statement shows that responders are sensitive to the random mechanism used – the higher offers it generates, the lower the rejection threshold.

Statement 9 *Consider two random offers generators that can be associated with random variables $1-X$ and $1-X'$ If $1-X$ first-order stochastically dominates $1-X'$, then rejection treshold in ROUG with $1-X$ is not higher than for ROUG with $1-X'$.*

Proof. See Appendix B. ■

The final result pertains to comparison of random mechanisms with deterministic offers, showing that the former generally result in lower rejection thresholds.

Statement 10 *Consider any random offers mechanism $1-X$ and some individual i. Denote the rejection threshold of i for this mechanism by $1-\bar{x}$. If probability that i accepts an offer generated by $1-X$ is strictly positive, then i's rejection threshold in a deterministic offer ultimatum game is at least $1-\bar{x}$.*

Proof. See Appendix B. ■

Indeed, as reported by Blount (1995), the rejection threshold is lower for offers generated by a fair random generator (following uniform distribution) than by human proposers (who generally offer less than half to the responder). In her Study 1, elicited rejection threshold was \$2.91 for human proposers and \$1.20 for the random mechanism. Interestingly under "third party decides" condition, rejection threshold was intermediate (\$2.08).[9] Note that according to the pure attributional view, rejection threshold should be 0 for both random and third party conditions.

In study 2, Blount used deception to compare responses to computer-made and (supposedly) human-made offers following the same distribution. Rejection threshold was significantly higher in the former, a finding believed to support the “attributional” explanation. Qualitatively identical results were obtained i.a. by Sanfey et al. (2003). Note however that this result is also predicted by the current model based on expected shares only. If offers are believed to be made by human proposers, the situation reduces to a 2-person game with certain deterministic offer, for which responses are elicited using strategy method. Chance factor is only involved at earlier stage (matching), before the actual game starts. Thus, according to Statement 10, for every subject whose rejection threshold for the computer-made offers is lower than the highest offer generated with a positive probability (i.e. for *every* subject of Blount) we shall expect a higher rejection threshold in the human-made offer version of the game. In the between-subject design used by Blount (1995) this indeed implies higher average rejection threshold for the human proposers.

Some researchers compared differences in responses to deterministic human-made first moves to random computer-made moves in other games, including "moonlighting game" (Falk et al., 2000) and gift-exchange game (Charness, 2004). These games are similar to the ultimatum game in that in both of them the first party has an opportunity to sacrifice for the benefit of the other player, who can subsequently respond with "reward" or punishment. While I do not provide detailed analysis of these games (which have richer strategy spaces than ultimatum game),[10] a short reflection shows that logic underlying Statement 10 applies also here.

[9] Given that most subjects expected offers very close to 50% from the third party, rejection threshold in this condition must be treated with caution, as subjects did not expect it to matter anyway.

[10] Responder behavior in UG can be, as we have seen, be represented in two dimensions corresponding to lower and upper bound of the acceptance range (the latte being often equal to the whole "pie"). In the "moonlighting game" and gift-exchange game response is not binary (accept/reject), thus strategy space of player 2 is substantially richer.

Consider two versions of a dynamic 2-person game, in one the first move being made deliberately by a human (Deterministic Human Made or DHM game), in the other by a computerized random generator following some known distribution (RCM game). Suppose that behavior of the second player is identical in both cases (i.e. conditional on the actual move, he chooses same action). Now, for an "unfriendly" move of player 1 in the DHM (expropriating substantial wealth from player 2), expected share (equal to the actual share, as no randomization occurs) is lower than in the RCM (in which some more "friendly" moves also occur), thus making player 2 demand a higher actual share. Thus, contrary to the assumption of identical response, player 2 has an incentive to restore (more) equality of outcomes by punishing player 1. Conversely, for a "friendly" move of player 1 in DHM, expected share (equal to the actual share) is higher than under RCM (in which some less "friendly" moves also occur), thus, contrary to the assumption of identical response, player 2 has an incentive to increase player's 1 share by rewarding him.

In general we can conclude that these experiments do not provide a satisfactory evidence that "intensions matter" because they confound two dimensions: attribution (human vs computer) and randomness (deterministic vs random). It is thus natural to consider a game in which first player chooses between probability distributions on different actions and check whether rejection threshold is still higher than when these distributions are imposed by the experimenter.

Superiority of the expected-share explanation over the attributional explanation is underscored when one considers non-fair random mechanisms. Bolton et al. (2005) investigated responses to extremely skewed (proposer-benefiting) generators and found that the difference in rejection behavior between random and human-made offers disappeared. Specifically, the authors used a strategy method to elicit responses in a random-offer ultimatum game with three possible divisions of the pie A: (0.1,0.9), B: (0.5.05) and C: (0.9,0.1). In the two "symmetric" treatments in which probabilities of A and C were equal, such that the expectation the offer equaled 0.5, C was rejected by 19% of the subjects. However, consistent with Statement 9, rejection rate in the "asymmetric" treatment with $(0.01 : A, 0.01 : B, 0.98 : C)$ was substantially higher (34%) and statistically identical to rejection rate for human-made offers (41%). Note that the positive rate of rejection in the "symmetric" treatment runs contrary to the predictions of the model proposed by the authors (see pp 1068-70) but is in line with the current model allowing for sensitivity to both procedural and outcome fairness. Interestingly, regardless of the procedure, overly-attractive offer A was sometimes

(15% of the time) rejected, but not the fair outcome B. This is in line with Statement 8.

It can be similarly shown that in ultimatum and dictator game players will display less concern about the payoffs of the others if roles have been explicitly assigned randomly.

An interesting study that yields insight into relative importance of actual and expected share is due to Kramer et al. (1995, study 3). They let subjects respond to an offer in an ultimatum game with initial offers being multiplied by a random component uniformly distributed over $[0, 2]$. Respondents were informed both about the original (intended) offer and the actual offer. The following result regarding this game can be proven.

Statement 11 *In ultimatum game with offers perturbed by uniformly distributed multiplicative noise factor, rejection threshold is negatively affected by initial (pre-perturbation) offer.*

Proof. See Appendix B. ■

The data of Kramer et al. yields support to this statement. More specifically, the authors manipulated fairness of the original offer (allegedly made by a human proposer), which was either \$12.50 or \$6.50 out of the pie of \$25. The offer was next multiplied by a noise factor (allegedly uniformly distributed over $[0, 2]$) of 0.52 or 1.92 respectively, such that the resulting offer were \$6.50 or \$12.50 respectively. Interestingly, intentionally high but effectively low offers were never rejected, whereas in the opposite case rejection rate was 23%. While this data is not rich enough to calibrate the model, it clearly suggests that the effect of expected share is stronger than of the actual share. Charness and Levine (2005) arrive at qualitatively similar results. Again, I do not provide a detailed analysis of this more complicated game.

3.2 Solidarity game

Let us consider a three-person solidarity game (Selten and Ockenfels 1998) in which each player independently runs an individual gamble winning fixed amount (10DM in the original experiment) with probability 2/3 and 0 otherwise. Each player makes two decisions. Namely, he determines the value of a conditional gift offered to the loser for the case in which there is only one (s_1) and, separately, to *each* of the losers for the case when there are two (s_2).

Statement 12 *In the solidarity game, a higher expected conditional gift from fellow players leads, other things being equal, to a higher own conditional gift. More specifically, if agents i and j behave in a sequentially rational way, display identical parameters of the function v and j's beliefs over gifts of the others stochastically dominate i's beliefs, j will give more.*

Proof. See Appendix B. ■

Indeed, Selten and Ockenfels (1998) observe a positive correlation between one's conditional gift and expectation with respect to other's gifts.

The other major finding of Selten and Ockenfels, was the emergence of the "fixed total sacrifice" pattern (e.g. 17 out of 120 subjects chose $s_1 = 2DM$, $s_2 = 1DM$). We note that this type of behavior cannot be explained within the current framework. Actually, as pointed out by Selten and Ockenfels (1998), it can hardly be explained by any utility-based model and should rather be understood in terms of the way in which subjects simplified the decision problem. More specifically, it seems natural to assume that they first fixed the amount to be given and then, for the case of two losers, split it in two (cf. Bolton et al. 1998). Particularly for higher sacrifices $s_1 = 2s_2 = 5DM$, it is hard to believe that subjects fully understood the consequence of their choices. Assuming identical gifts on part of fellow players (and indeed most subjects reported expecting gifts similar to the ones they have chosen), each of two "winners", including the decision maker, would end up having two times less than the "loser". It seems difficult to reconcile this finding with any model of rational behavior.

3.3 Support for redistribution

To illustrate the power of this approach I shall also consider one of the predictions that can be derived from the model with respect to big-scale, real-life processes. I have already mentioned previously that interaction between the procedural fairness and the distributive fairness creates a link between mobility and support for redistribution. Currently I shall elaborate on this point using a simple model.

Consider a society (S1) consisting of individuals endowed with "success likelihood" parameter p – probability of obtaining wealth w of 1 rather than 0. For simplicity let $p = p_h$ for the fraction of $q_h = 1/2$ individuals and $p = p_l$, $p_l < p_h$ for the remaining $q_l = 1 - q_h = 1/2$ subjects. In the game under consideration, first everyone's wealth is determined according to his p. We can thus speak of the "privileged rich" with $p = p_h, w = 1$ (making on average $\frac{1}{2}p_h$ of the society), "privileged poor" ($\frac{1}{2}(1-p_h)$), "unprivileged rich"

($\frac{1}{2}(1-p_l)$) and "unprivileged poor" ($\frac{1}{2}p_l$). Next, amount of redistribution is decided upon.[11] This can be modeled using single parameter $r \in [0,1]$ corresponding to the resulting difference in wealth between the rich and the poor.

I shall assume that all individuals display preferences described in the current model and I allow for any form of heterogeneity of parameters within groups.

To examine the impact of inequality of initial opportunities, suppose now that these individuals face greater dispersion of chances: $p'_h > p_h$ and $p'_l < p_l$ (society S2). Assume for simplicity that magnitude of the groups and expected total wealth remain unchanged (so that we have $p_h + p_l = p'_h + p'_l$).

I will show that when inequality of opportunities increases, i.e. when we move from S1 to S2, the majority of citizens will shift their preferences toward greater redistribution, i.e. for the majority of citizens the optimal redistribution r_i^* will increase. Suppose that the actual r is not increased. To begin with, monetary incentives of each of the rich and the poor remain unchanged and on average these groups will be just as large as they were. However, as redistribution level stays constant, expected share in the game of the privileged increases relative to the first situation, while expected share of the unprivileged necessarily decreases. Now, because redistribution does not make the poor richer than the rich, by virtue of Statement 6 (point 3), the actual share of the privileged rich is now higher than the optimal one and actual share of the unprivileged poor is lower than the optimal one. Because the expected share of the former increases and of the latter – decreases, by virtue of Assumption 3 both will have additional incentive to support redistribution. While effect on the two other groups is uncertain, privileged rich and unprivileged poor comprise more than 50% of the society (which does not depend on the assumption that $q_h = q_l$), so indeed majority opts for greater redistribution. The actual shift in the policy will now depend on the assumption on the nature of the collective decision process. If we were for instance interested in the opinion of the median voter, we can note that he is likely to belong to one of these two largest groups, especially while unprivileged rich are predicted to have extreme anti-redistribution views, because not only redistribution runs contrary to their material self-interest but also their low expected share make them demand a high actual share.

In this way we are able to explain the observed stronger support for redistribution in societies in which little vertical mobility exists (i.e. everyone's

[11] We do not specify the procedure by which level redistribution is determined. Implications of various assumptions in this regard will be discussed later.

final wealth is highly predictable or, in terms of the simple model sketched above, $p_h - p_l$ is large). This is in line with findings reported in Alesina et al. (2004) and Alesina and Angeletos (2005).

4 Research into procedural justice

The most closely related model is due to Bolton et al. (2005). Basing on the observed results of a laboratory experiment (commented here on page 13), they develop a post-hoc model that incorporates procedural justice and distributive justice. They assume that fairness judgments are determined by the smaller of absolute deviations from the ideal point on either dimension. That is, an allocation is considered fair if the actual shares are equal or when they result from an unbiased procedure and the more distant from the ideal point (on this dimension on which the deviation is smaller), the less fair the allocation.

However, this model implies that a disadvantageous allocation resulting from an advantageous procedure is considered just as unfair as the same allocation resulting from a disadvantageous procedure. This is implausible. If the former situation is deemed unfair, who is privileged and who is oppressed? Further, even though the model of Bolton et al. (2005) is specifically designed to explain the results of their experiment and cannot be directly generalized to other types of games, still it cannot explain them fully, as the authors readily admit. Thus, while Bolton et al. (2005) is a major inspiration and reference point for the current work, I believe it only constitutes a first step in modeling fairness of procedures.

Another closely related approach is due to Trautmann (2006), who develops a "Process Fehr-Schmidt" model in which inequality in outcomes is replaced by inequality of expected outcomes. This model also allows explaining sensitiveness of rejection behavior to fairness of the offer-generating process in ultimatum game and has a major advantage of being quite simple.[12] Again, however, because distributive justice plays no role in the model, rejections of offers stemming from a fair mechanism (happening with substantial frequency in Bolton et al. (2005) cannot be explained.

In psychology, several theories sought to investigate the basis of importance of procedural justice and its interaction with distributive justice when assessing the outcomes (see Brockner and Wiesenfeld, 1996, for an excellent review).

[12] It is not entirely clear from the paper however, how the expected values are to be computed in a more complex game.

First, we can interpret the taste for fair procedures as a long-run strategy securing acceptable outcomes in an uncertain environment (Thibaut and Walker 1975,1978). It has been suggested that low current outcome makes procedural justice more important – it is precisely when we do not immediately benefit from the interaction *and* cannot see long-term mechanisms supporting our case that we feel dissatisfied with it and may consider deviation.

Within the framework of Referent Cognitions Theory (RCT, see Folger 1986a, 1986b), agents who obtain unfavorable outcomes find it easy to imagine a more positive alternative when they perceive the procedure employed as unfair (cf. "simulation heuristic" in Kahneman and Tversky, 1982). Contrary, when they accept the procedure (e.g. because they have had a say on or input into the allocation decision), the more attractive alternative is less salient. As a result, interaction between procedural and distributive justice emerges that makes a combination of unfair procedure and bad outcome particularly unattractive.

Under Group Value Theory (Lind and Tyler, 1988, pp. 230-239), the wish to be treated fairly is driven by the desire to be considered an important member of the group which shapes self-identity and contributes to self-esteem. The less these needs are satisfied (because of lacking fair procedure), the more the interaction might be perceived as transactional rather than personal. This amplifies the need for immediate material benefits. Tyler and Lind (1992) further develop a related model (termed Relational Model of Authority) that explains how procedural justice of an authority determines the extent to which citizens feel treated as they "deserve", thus shaping legitimization of power.

Finally, in view of the attribution theorists, interaction between procedural and distributive justice may be explained in terms of the causal attributions performed on outcomes. Basically, if poor outcomes result from an unfair distribution, an individual attributes them to it. If however, an unfair procedure generates outcome which is considered good, this is attributed to individual's skills, luck etc. Thus the unfairness of the procedure is not salient any more. Additionally, from the viewpoint of self-perception theory, fair procedure may justify personal commitment. When, however, procedure is considered biased, an individual may question why he ever participates in the interaction. He or she then seeks to explain it in terms of immediate benefits, making actual outcomes more important.

Yet another way to obtain insight into importance of procedural justice is via analysis of emotional responses to interactions, while systematically varying favorability of procedures and outcomes. Two recent papers that investigate this issue are Weiss et al. (1999) and Krehbiel and Cropanzano (2000). Both studies used a similar technique of procedural justice manipulation, namely they made use of a confederate that declared having "insider" knowledge lending an unfair advantage to his/her team in a competitive task (and, needless to say, disadvantage to the other team). Not surprisingly, happiness/joy and sadness result from positive and negative outcomes respectively. It is however also apparent from this research that emotional reaction is strongly influenced by interaction between procedure and outcome. As a result a fourfold response pattern emerges: if procedure is disadvantageously biased (e.g. the informed confederate is the member of the other team) and outcome is bad (competition is lost), anger-related emotions obtain. If outcome is positive, despite unfavorable conditions, pride is the prevalent emotion. Pride is also found to be present when both procedure and outcome are favorable. However, it is accompanied by guilt and anxiety. In the case of a negative outcome that obtained despite advantageously biased procedure, disappointment results (although the latter is most strongly influenced by the negative outcome itself, not necessarily an interaction). Boiling these complex emotional responses to sheer valence, unfavorable bias of the procedure hurts most if the outcome is bad, whereas a favorable one – if the outcome is good.

5 Conclusion and extensions

Basing on psychologically sound assumptions I have derived a simple model with considerable predictive power. I have shown that incorporating rough measures of perceived distributive and procedural justice makes it possible to explain a wide range of experimental results observed in games of divisions involving risk. To be sure, for sake of simplicity, I neglect several issues that seem to be of importance in social interactions, both in the lab and elsewhere. In particular, the model may fail to generate correct predictions in some games where my crude measures of procedural and distributive justice cannot provide a proxy for the intensions of the other agents. One such example is menu dependence. Several studies have shown that manipulating the set of claims allowed by rules of a (Mini) Ultimatum Game can affect the rejection behavior. This reflects the fact that even a low offer does not seem so low if it is actually the highest offer allowed by the rules of the game. It is also apparent that framing of the game can affect what players perceive as fair shares. For example, in the power-to-take game (Bosman and van Winden, 2002) players representing tax authority tend to claim substantial amounts, thus opting for allocation far from equality, apparently exhibiting more greediness than they would in the ultimatum game. A relatively simple[13] extension of my model that would admit capturing such regularities is a modification of the assumption regarding reference point, e.g. in line with the idea developed by Bolton et al. (2005). To keep the current paper relatively short and the model parsimonious, this important line is left for future research.

I have to stress that it was not my goal to argue here that perceived intentions do not matter. I certainly believe they do, even in relatively impersonal environments like the market or experimental laboratory. What requires further investigation is however when and how strongly intentions actually matter and whether it is an advisable research strategy to incorporate them directly in the formal models.

While the findings investigated in detail in this paper support the model, further empirical verification is surely needed. One can hope that the renewed interest in procedures within the economic science will inspire experimental and field research fostering theoretical modeling.

References

[1] Alesina, A. and G-M. Angeletos (2005). Fairness and Redistribution, U.S. versus Europe, American Economics Review, 95(4), 960-980.

[2] Alesina, A., R. Di Tella and R. MacCulloch (2004). Inequality and Happiness: Are Europeans and Americans Different? *Journal of Public Economics, 88*, 2009– 2042.

[3] Babcock, L. and G. Loewenstein (1997). Explaining Bargaining Impasse: The Role of Self-Serving Bias, *Journal of Economic Perspectives*, *11*(1), 109-126.

[4] Battigalli, P. and M. Dufwenberg (2005). Dynamic Psychological Games, *Econometrica*, under revision.

[5] Blount, S. (1995). When Social Outcomes Aren't Fair: The Effect of Casual Attributions on Preferences,*Organizational Behavior and Human Decision Processes*, *63*, 131-144.

[13] "Simple" in terms of the formal structure of the model. Finding actual reference points in particular games can be quite demanding.

[6] Bolton G.E., J. Brandts and A. Ockenfels (2005). Fair Procedures: Evidence from Games Involving Lotteries,*Economic Journal*, *115*(506), 1054-1076.

[7] Bolton, G. E. and A. Ockenfels (2000). ERC: A theory of Equity, Reciprocity and Competition. *American Economic Review*, *90*(1), 166-193.

[8] Bosman, R. and F. van Winden (2002). Emotional Hazard in a Power-to-Take Game. *Economic Journal, 112*(1), 147-169.

[9] Brockner, J. and B.M. Wiesenfeld (1996). An Integrative Framework for Explaining Reactions to Decisions: Interactive Effects of Outcomes and Procedures. *Psychological Bulletin*, *120*(2), 189-208.

[10] Charness, G. (2004). Attribution and Reciprocity in an Experimental Labor Market. *Journal of Labor Economics*, *22*(3).

[11] Charness, G. and D. I. Levine (2005). Intention and Stochastic Outcomes: An Experimental Study. Working paper.

[12] Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly journal of economics, 117*(3), 817-869.

[13] Cox, J. C. and C. A. Deck (2005). On the Nature of Reciprocal Motives. *Economic Inquiry, 43*(3), 623–635.

[14] Cropanzano, R., and Greenberg, J. (1997). Progress in organizational justice: Tunneling through the maze. In C. L. Cooper and I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (317-372). New York: John Wiley.

[15] Cropanzano, R.,Prehar, C. A. and Chen, P. Y. (2002). Using Social Exchange Theory to Distinguish Procedural From Interactional Justice. *Group and Organization Management, 27*(324-351).

[16] Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47: 268–298.

[17] Falk, A., E. Fehr and U. Fischbacher (2000). Testing Theories of Fairness – Intentions Matter. Working Paper, University of Zurich.

[18] Fehr, E., and S. Gächter. (2000). Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, *14*(3), 159–81.

[19] Fehr, E. and K. Schmidt (1999), A theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, *114*, 817-868.

[20] Folger, R. (1986a). Rethinking equity theory: A referent cognitions model. In: H. W. Bierhoff, R. L. Cohen, andJ. Greenberg (Eds.). *Justice in social relations* (145-162). New York: Plenum Press.

[21] Folger R. (1986b). A Referent Cognitions Theory of Relative Deprivation. In: J. M. Olson, C. P. Herman and M. P. Zanna (eds.). *Relative Deprivation and Social Comparison.* London: Lawrence Erlbaum Publishers.

[22] Geanakoplos, J., J. D. Pearce and E. Stachetti (1989): Psychological Games and Sequential Rationality. *Games and Economic Behavior, 1*, 60-79.

[23] Hoffman, E., and M.L.Spitzer (1985), Entitlements, rights, and fairness: an experimental examination of subjects' concepts of distributive justice,.*Journal of Legal Studies* 14 (2), 259–297.

[24] Karni, E., T. Salmon and B. Sopher (2001). Individual Sense of Fairness: An Experimental Study. Mimeo.

[25] Kramer, R. M., P. P. Shah, P. P., and S. Woerner (1995). When ultimatums fail: Social identity and moralistic aggression in coercive bargaining. In: Kramer, R. M. and D. M. Messick (Eds.). *Negotiation as a social process* (285–308). Thousand Oaks, CA: Sage.

[26] Kroll Y. and L. Davidovitz (2003). Inequality Aversion vs Risk Aversion, *Economica, 70*, 19-29.

[27] Krehbiel, P. J. and R. Cropanzano (2000). Procedural Justice, Outcome Favorability and Emotion, *Social Justice Research, 13.*

[28] Lind, E.A and T. R. Tyler (1988). The Social Psychology of Procedural Justice, *Plenum Press*, New York.

[29] Loewenstein, G. F., E. U. Weber, C. K. Hsee and N. Welch (2001). Risk as Feelings. *Psychological Bulletin, 12*7(2), 267-286.

[30] Rabin, Matthew (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review. 83*(5), 1281-1302.

[31] Rawls, J. (1971). Theory of Justice, Harvard University Press.

[32] Reuben, E. and F. van Winden (2005). Negative Reciprocity and the Interaction of Emotions and Fairness Norms. Discussion Paper 05-014/1. Tinbergen Institute.

[33] Sanfey A.G., J.K. Rilling, J.A. Aronson, L.E. Nystrom, and J. D. Cohen (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 13*, 300(5626). 1673-5.

[34] Schmitt, P. (2004). On perceptions of fairness: The role of valuations, outside options, and information in ultimatum bargaining games. *Experimental Economics* 7: 49-73.

[35] Selten, R. and A. Ockenfels (1998). An Experimental Solidarity Game,*Journal of Economic Behavior and Organization*, *34*(1998), 517-539

[36] Sen, A. K. (1979). Personal Utilities and Public Judgement. *Economic Journal* 89, 537-558.

[37] Sonnegard J. (1996). Determination of First Movers in Sequential Bargaining Games: An Experimental Study. *Journal of Economic Psychology, 17*, 359-386

[38] Strotz, R H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. Review of Economic Studies*, 23*(3), 165-180.

[39] Thibaut, J. and L. Walker (1975). Procedural Justice: A psychological analysis. Hillsdale, NJ: Erlbaum.

[40] Thibaut, J. and L. Walker (1978), A theory of procedure,*California Law Review*, 66, 541-566.

[41] Trautmann, S. (2006). A Procedural Extension of the Fehr-Schmidt Model of Inequality Aversion, MPhil Thesis, Tinbergen Institute.

[42] Tyler, T. R, (2000). Social Justice. Outcome and Procedure. *International Journal of Psychology*, *35*(117-125).

[43] Tyler, T. R and E. A. Lind (1992). A relational Model of Authority in Groups,in: M. P. Zanna (Ed.). *Advances in Experimental Social Psychology*, *25*(115-191) San Diego: CA Academic Press.

[44] Weiss, H. M., Suckow, K., and Cropanzano, R. (1999). Effects of justice conditions on discrete emotions. Journal of Applied Psychology, *84*, 786–794.

A Psychological games of fairness

A major difficulty in modeling expected-share-dependent utilities is that the effect of strategies on the utilities is not limited to the fact that they determine the actual outcome – non-played paths also affect the expected share of any of the players, thus shaping their satisfaction with the outcome. Such effects cannot be accounted for within the standard game theory. Therefore I have to make use of the concept of dynamic psychological games (see Battigalli and Dufwenberg (2005), thereafter BD; in this Appendix I reproduce the concepts introduced there) that allows utilities depend directly on beliefs and strategies.[14]

I begin by considering a n-person *material payoff game* in extensive form $\langle N \cup \{0\}, H, (y_i)_{i \in N} \rangle$, where $N \cup \{0\} = \{1, 2...n\} \cup \{0\}$ is the *player* set, whereby player 0 is interpreted as "nature", H is the set of feasible *histories* of the game and $(y_i : Z \to R^+ \cup \{0\})_{i \in N}$ is a *payoff function*, where $Z \subset H$ is a set of *terminal histories* (*end nodes*). A history of length l is a sequence $h = (a^1, ..., a^l)$ where each $a^t = (a^t_0, a^t_1, ..., a^t_n)$ represents the profile of actions chosen at stage t $(1 \leq t \leq l)$. Note that every player makes a move at every stage of the game which however causes no loss of generality, as set of available actions can be a singleton. I also assume that a history h becomes public information as soon as it occurs. For notational convenience, I let H contain the empty history, denoted by h^0 (the history of length 0). The set of feasible actions for player i at history h is denoted by $A_i(h)$ and it may be a singleton, meaning that i is not active at h. $A_i(h)$ is empty if and only if h is a terminal history.

Payoff function $(y_i)_{i \in N}$ determines for each terminal history $z \in Z$ a vector of length N representing non-negative material payoffs obtained by each agent. Allowing for negative payoffs would make defining shares cumbersome. Furthermore, I shall mostly confine my attention to situations in which this assumption does not seem overly restrictive.

I let $S_{i \in N}$ denote the finite set of (pure) strategies of player i. Individual strategy is denoted by $s_i = (s_{i,h})_{h \in H \setminus Z}$, where $s_{i,h}$ is the action that would be selected by strategy s_i if history h occurred. Define $S = \Pi_{i \in N} S_i$ and $S_{-i} = \Pi_{j \neq i} Sj$. The set of strategies of player i that allow history h is denoted $S_i(h)$. A similar notation is used for strategy profiles: $S(h) = \Pi_{i \in N} S_i(h)$ and $S_{-i}(h) = \Pi_{j \in N} S_j(h)$. Finally, I let $\zeta(s, s_0) \in Z$ denote the terminal history induced by strategy profile $s = (s_i)_{i \in N}$ and strategy of nature s_0.

[14] Earlier notion of psychological games introduced by Geanakoplos et al. (1989) does not allow utilities depend on strategies in other ways than through terminal histories obtained.

I can finally define motivation function of player i, $u_i(z, s) : Z \times S \to R$. Note that I simplify framework of BD in that I do not allow utilities to depend explicitly on (higher order) beliefs, but only on actual outcome z and behavior strategies δ (through agent's expected share in the game). Still, conjectures regarding other players' strategies are crucial for maximization of motivation function.

It turns out to be necessary to prove existence of equilibrium to let each player be endowed with beliefs regarding other players' strategies (first-order beliefs) and beliefs (higher-order beliefs) $\mu_i(\cdot|h)$ satisfying collective coherence and consistency as defined by BD pp 16, 18n. Roughly, the assumption is that different-order beliefs of a single agent are not contradictory and that individuals correctly update their beliefs upon observed histories. Further, any two subjects have identical beliefs about a third subject. Admittedly, identical beliefs regarding third-party is a rather strong assumption. Of course, it is harmless in the case that I deal with throughout most of this paper, namely in the domain of 2-person games. For multiple-person games, one can drop this assumption and perform non-equilibrium analysis based on the notion of sequential rationality, that is, assuming that subjects form beliefs and act optimally according to them, but those need not fulfil the above-mentioned criteria.

Note that non-pure strategies can be modeled here in terms of other players' beliefs regarding given strategy. In particular, randomness is a key feature of nature's actions. More specifically, I shall define a profile of behavior strategies $(\delta_1, \delta_2, ...\delta_n)$ with a typical element

$$\delta_i = (\delta_i(\cdot|h))_{h \in H \setminus Z} \in \Pi_{h \in H \setminus Z} \Delta(A_i(h)), \tag{17}$$

where $\Delta(A_i(h))$ denotes set of probability measures on set of actions available to i at h. I will interpret δ_i as representing (history-dependant) first-order beliefs regarding i and held by his fellow players. Similarly δ_0 denotes behavioral strategy of nature.

Our main object of interest can be defined as follows:

Definition 13 *A psychological game is a triple*

$$\Gamma = \langle N \cup \{0\}, H, (u_i : Z \times S \to R)_{i \in N} \rangle. \tag{18}$$

Because of dependence on strategies, expected payoff of player i conditional on belief profile μ and non-terminal history h has to be defined in a rather cumbersome way. Denoting beliefs of an arbitrary other player j about i's strategy by μ^1_{ji}, we have that expected payoff is given:

$$E[v_i|h,a_i] = \sum_{s_{-i}\in S_{-i}(h)} \mu^1_i(s_{-i}|h) \sum_{s_i\in S_i(h)} \mu^1_{ji}(s_i|(h,a_i,s_{-i,h}))u_i(\zeta(s),\mu,s), \tag{19}$$

Thus "agent of player i at h" is inherently uncertain about the choices of future agents of player i – he holds believes just as any other player does. This stems from the fact that dependence on strategies implies dynamic inconsistency – a player may find himself unable (i.e. unwilling) to carry out what seemed ex ante to be the optimal plan. Because in the equilibrium other players' beliefs μ^1_{ji} are correct, formulation of expected payoff as in 19 assures that no player can perform effective self-deception – just as he cannot affect other players' behavior using threats or promises which (as is readily seen), he will not be willing to execute[15] (this idea of "consistent planning" can be traced back to Strotz 1955-56).[16]

Definition 14 *Equilibrium concept. Pair* $(\delta_i,\mu_i)_{i\in N}$ *is an equilibrium in procedural and distributional fairness motivation functions (a PDF-equilibrium) if* μ *is consistent,* δ *coincides with first-order believes in* μ *and for all* $i\in N$, $h\in H$,

$$supp(\delta_i(\cdot|h)) \subseteq \arg\max_{a_i\in A_i(h)} E[v_i|h,a_i], \tag{20}$$

where $supp(\delta_i(\cdot|h))$ denotes support of the behavior strategy δ_i and $v_i = u_i(z,s)$ is a PDF motivation function satisfying Assumptions listed in Subsection 2.2.

[15] An illustrative example may help understand why strategy-dependence leads to dynamic inconsistency. Suppose your friend wants to talk to you. While you're uncertain about the topic, you think it's quite likely he will ask you to lend him some money. On your way there, you need to plan what to do if he indeed does. Now, it plausible, that this plan will also affect your well-being *if he doesn't.* More specifically, you might enjoy the thought of being eager to help out your friends in need. Given that he might indeed not ask, this makes the plan (give if he asks) relatively more attractive. Suppose that, incorporating this motivation, you decided to follow this strategy. You meet your friend and he asks for money. Now, you do not care any more about the way you would have felt had he not asked, because the course of events turned out to be different. Thus, giving becomes less attractive and your preference may actually reverse, thus making you abandon your previously designed plan. Current formulation assures that players disregard strategies like this one, which they know they will not be willing to follow.

[16] See also BD for a discussion of this formulation.

This definition assures that for every history players correctly update their beliefs and choose (and are believed to choose) non-optimal actions with probability 0.

B Proofs

Proof of Lemma 6. Parts (1) and (2) of Lemma 6 follow immediately from Assumption 2 and the fact that $\sigma_i, \sigma_i^E \in [0,1]$. To prove Part (3) consider some y_i and y_i'. Without loss of generality we can assume that $y_i < y_i'$. We know that

$$v_{i3}(y_i', \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) = 0. \tag{21}$$

I will show that also

$$v_{i3}(y_i, \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) = 0. \tag{22}$$

Suppose that

$$v_{i3}(y_i, \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) < 0.. \tag{23}$$

By 13, $v_{i13}(y_i', \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) \leq 0$. It can be easily seen that $v_{i3}(z, \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) < 0$ and so $v_{i13}(z, \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) \leq 0$ for all $z \in [y_i, y_i')$. Therefore $0 > v_{i3}(y_i, \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) > v_{i3}(y_i', \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) = 0$, a contradiction. Identical reasoning shows that $v_{i3}(y_i, \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) > 0$ cannot hold. Thus $v_{i3}(y_i, \sigma_i, \sigma_i^{E*}(\sigma_i, y_i')) = 0$. But it must also be that $v_{i3}(y_i, \sigma_i, \sigma_i^{E*}(\sigma_i, y_i)) = 0$. Thus, by 8, $\sigma_i^{E*}(\sigma_i, y_i') = \sigma_i^{E*}(\sigma_i, y_i)$. Part (4) follows directly from Assumption 3 and proof of Parts (1) and (2) – consider any actual share $\sigma_i' \in [0,1]$ and corresponding optimal expected share $\sigma_i^{E*}(\sigma_i')$. Due to Part (1), we have that

$$\frac{\partial v_i(y_i, \sigma_i, \sigma_i^E)}{\partial \sigma_i^E}|_{\sigma_i = \sigma_i'} < 0 \; \forall \sigma_i^E \in [\sigma_i^{E*}, 1]. \tag{24}$$

Now consider some $\sigma_i'' \in [\sigma_i', 1]$. By virtue of Assumption 3,

$$\frac{\partial v_i(y_i, \sigma_i, \sigma_i^E)}{\partial \sigma_i^E}|_{\sigma_i = \sigma_i''} < 0 \; \forall \sigma_i^E \in [\sigma_i^{E*}, 1], \tag{25}$$

so $\sigma_i^{E*}(\sigma_i'') \leq \sigma_i^{E*}(\sigma_i')$.

Proof of Statement 8. For any fixed σ_i^E resulting from the distribution of offers and rejection behavior, responder's motivation value when he rejects an offer, no matter of what magnitude, is given as:

$$u_R \equiv v_i(0, 1/2, \sigma_i^E). \tag{26}$$

If responder accepts offer of $1-x$, his motivation value is equal to

$$u_A(1-x) \equiv v_i(1-x, 1-x, \sigma_i^E). \tag{27}$$

In equilibrium player accepts whenever $u_A(1-x) > u_R$ and is indifferent between accepting and rejection if $u_A(1-x) = u_R$. By virtue of Assumption 1 and 2, u_A is a concave function of $1-x$. It is thus also quasi-concave and the set $A'(\sigma_i^E)$ of values of $1-x$ for which $u_A(1-x) \geq u_R$ is convex, i.e. it is an interval. Now,

$$u_A(0.5) = v_i(1/2, 1/2, \sigma_i^E) > v_i(0, 1/2, \sigma_i^E) = u_R \tag{28}$$

due to Assumption 1 so $1/2 \in A'(\sigma_i^E)$.

$$u_A(0) = v_i(0, 0, \sigma_i^E) < v_i(0, 1/2, \sigma_i^E) = u_R \tag{29}$$

due to Assumption 2, so $0 \notin A'(\sigma_i^E)$. I define $A(\sigma_i^E) = A'(\sigma_i^E) - \{\sup(A'(\sigma_i^E))\} - \{\inf(A'(\sigma_i^E))\}$ (where '$-$' denotes set subtraction). and $1-\overline{x} = \inf(A'(\sigma_i^E))$, $1-\widetilde{x} = \sup(A'(\sigma_i^E))$. For each σ_i^E, $A(\sigma_i^E)$ satisfies conditions given in Statement 8. I now have to prove that for some σ_i^E and some p, q, expected share corresponding to $A(\sigma_i^E)$, p and q will indeed be σ_i^E. Parameter p and q can be any numbers in $[0,1]$ except for the case when $v_i(1, 1, \sigma_i^E) > u_R$ which implies that $1-\widetilde{x} = 1$ and $q = 1$. Consider a correspondence between intervals of $[0,1]$ and admissible expected shares $f(A)$ containing all σ_i^E that correspond to expected shares resulting from acceptance on A, rejection on $B = [0,1] - A - \{1-\overline{x}\} - \{1-\widetilde{x}\}$ and acceptance with probabilities p, q on $1-\overline{x}$ and $1-\widetilde{x}$ respectively:

$$\sigma_i^E \in f(A) \Leftrightarrow \exists_{p,q} \text{ s. t.}$$

$$\frac{P(1-x \in A)E(1-x|1-x \in A) + pP(1-x = 1-\overline{x})(1-\overline{x}) + qP(1-x = 1-\widetilde{x})(1-\widetilde{x})}{P(1-x \in A) + pP(1-x = 1-\overline{x}) + qP(1-x = 1-\widetilde{x})} = \sigma_i^E. \tag{30}$$

A correspondence can now be defined $r : [0,1] \rightrightarrows [0,1]$, $r(\sigma_i^E) = f(A(\sigma_i^E))$. It can be easily verified that $r(\sigma_i^E)$ is nonempty and convex and have closed graph. Thus, by Kakutani's fixed point theorem, there are such σ_i^E, A that $\sigma_i^E \in f(A), A = A(\sigma_i^E)$. It also follows from the general Theorem 7. Further, such A is unique, because $1-\overline{x}$ and $1-\widetilde{x}$ are non-decreasing in σ_i^E (compare proof of Lemma 9). Rejection rates p, q must be chosen in a specific way to assure that $A = A(\sigma_i^E)$. This problem generally, has one degree of freedom, thus choosing some particular q leaves (at most). one appropriate p. In particular, when $v_i(1, 1, \sigma_i^E) > u_R$ such that $q = 1$, p is determined uniquely.

Proof of Statements 9 and **11.** Essentially both Statements assert that rejection threshold be higher for offers with higher expected value. Denote the lower rejection thresholds for offers given by random variables $1 - X_{(1)}$ and $1 - X_{(2)}$ (such that the former stochastically dominates the latter) by $1 - \overline{x}_{(1)}$ and $1 - \overline{x}_{(2)}$ respectively.[17] Motivation value experienced by respondent who accepts (A). an offer of $1 - x$ resulting from a random number generator $1 - X_{(1)}$ is given by:

$$u_{(1)}(A, 1 - x) = v_i(1 - x, 1 - x, \sigma_{i(1)}^E), \text{ where} \tag{31}$$

$$\sigma_{i(1)}^E = E((1 - x)|\text{accepted}). \tag{32}$$

On the other hand, motivation value resulting from rejecting (R) an offer is equal to

$$u_{(1)}(R, 1 - x) = v_i(0, \frac{1}{2}, \sigma_{i(1)}^E). \tag{33}$$

In identical fashion I define $u_{(2)}(A', 1 - x)$,

$$u_{(2)}(A', 1 - x) = v_i(1 - x, 1 - x, \sigma_{i(2)}^E), \tag{34}$$

$$u_{(2)}(R', 1 - x) = v_i(0, \frac{1}{2}, \sigma_{i(2)}^E). \tag{35}$$

and $\sigma_{i(2)}^E$ for the random offer generator $1 - X_{(2)}$.

Suppose that (1) $1 - \overline{x}_{(1)}$ is higher than $1 - \overline{x}_{(2)}$ or that (2) $1 - \overline{x}_{(1)} = 1 - \overline{x}_{(2)}$ and probability of rejection at the threshold is at least as high in the former case as it is in the latter Then, by stochastic dominance, $\sigma_{i(1)}^E > \sigma_{i(2)}^E$. From Lemma 8 we know that $1 - \overline{x}_{(1)} \leq \frac{1}{2}, 1 - \overline{x}_{(2)} \leq \frac{1}{2}$. Now we have that

[17] For simplicity we only consider the case that the upper rejection thresholds $1 - \widetilde{x}_1$ and $1 - \widetilde{x}_2$ are equal to 1. This is indeed the case for most of the subjects and wide range of random offer generators. Without this assumption proof becomes somewhat more complicated. The same applies to the proof of Statement 10.

$$u_{(1)}(A, 1-x) - u_{(2)}(A, 1-x) = \int_{\sigma^E_{i(2)}}^{\sigma^E_{i(1)}} \frac{\partial v_i}{\partial \sigma^E_i}|_{(y,1-x,\sigma^E_i)} d\sigma^E_i. \quad (36)$$

Similarly

$$u_{(1)}(R, 1-x) - u_{(2)}(R, 1-x) = \int_{\sigma^E_{i(2)}}^{\sigma^E_{i(1)}} \frac{\partial v_i}{\partial \sigma^E_i}|_{(0,1/2,\sigma^E_i)} d\sigma^E_i. \quad (37)$$

The function being integrated is, by virtue of Assumptions 3 and 5 greater in 36 than in 37. Thus,

$$u_{(1)}(A, 1-x) - u_{(2)}(A, 1-x) > u_{(1)}(R, 1-x) - u_{(2)}(R, 1-x), \quad (38)$$

$$0 = u_{(1)}(A, 1-x) - u_{(1)}(R, 1-x) > u_{(2)}(A, 1-x) - u_{(2)}(R, 1-x) = 0, \quad (39)$$

where equalities result from definition of rejection threshold. This is a contradiction, so $1-\overline{x}_{(1)}$ must be lower than $1-\overline{x}_{(2)}$ (or $1-\overline{x}_{(1)} = 1-\overline{x}_{(2)}$ and probability of rejection at the threshold is at lower level in the former case as it is in the latter).

Proof of Statement 10. Consider a deterministic offer lower than $1-\overline{x}$, the rejection threshold for the random mechanism. Obviously, if it is accepted, then $\sigma^E_{i(DETERMINISTIC)} = 1-\overline{x} < \sigma^E_{i(RANDOM)}$, thus reasoning from the proof of Statements **9** and **11** yields contradiction, so any offer below $1-\overline{x}$ will be rejected as claimed.

Proof of Statement 12. Recall that every player faces independent probability of winning $p = 2/3$. I will use the notation $s_i = (s_i^1, s_i^2)$ for player's i strategy – his conditional gifts if there is one loser and when there are two respectively. Clearly, for player 1 there is *a priori* no reason to expect different level of support from player 2 than from player 3. Supposing player 1 knows the value of $s_2 = (s_2^1, s_2^2) = s_3$, his expected utility is given by:

$$\begin{aligned} E(v_i) \quad = \quad & p^3 v_i(1, \frac{1}{3}, \sigma^E_i) + 2p^2(1-p) v_i(1-s_1^1, \frac{1-s_1^1}{2}, \sigma^E_i) + p(1-p)^2 v_i(1-2s_1^2, 1-2s_1^2, \sigma^E_i) + \\ & p^2(1-p) v_i(0+2s_2^1, s_2^1, \sigma^E_i) + 2p(1-p)^2 v_i(s_2^2, s_2^2, \sigma^E_i) + (1-p)^3 v_i(0, \frac{1}{3}, \sigma^E_i) \end{aligned} \quad (40)$$

with

$$\sigma_i^E = p^3 1 + 2p^2(1-p)(1-s_1^1)-p)+p(1-p)^2(1-2s_1^2)+p^2(1-p)0+2p(1-p)^2 s_2^2+(1-p)^3 0. \tag{41}$$

The first three terms on the RHS of 40 refer to the situation when player 1 actually wins the gamble (and, respectively, two, one or no other player does), the three last terms – to the situation when player 1 loses and, respectively, two, one or no other player wins. As will be seen later, we can conclude the proof without having to simplify the expression for σ_i^E and plug it into 40.

To examine the impact of expected others' gifts on own optimal gifts I shall compute the cross derivative $\frac{d^2E(u_i)}{ds_1^2 ds_2^2}$. Analysis of analogous effects of s_2^2 on s_1^1 and of s_2^1 on s_1^1 and s_1^2 is left for the reader.

Note that increasing s_1^2, player 1 decreases his payoff and actual share iff he happens to be the only winner, which happens with probability $p(1-p)^2$. Further, it decreases his expected payoff. Differentiation with respect to s_1^2 thus yields:

$$\begin{aligned}
\frac{dE(v_i)}{ds_1^2} \quad = \quad & -2p(1-p)^2[v_{i1}(1-2s_1^2, 1-2s_1^2, \sigma_i^E) + v_{i2}(1-2s_1^2, 1-2s_1^2, \sigma_i^E)] + \\
& + \frac{\partial \sigma_i^E}{\partial s_1^2}[p^3 v_{i3}(1, \frac{1}{3}, \sigma_i^E) + 2p^2(1-p)v_{i3}(1-s_1^1, \frac{1-s_1^1}{2}, \sigma_i^E) + \\
& + p(1-p)^2 v_{i3}(1-2s_1^2, 1-2s_1^2, \sigma_i^E) + p^2(1-p)v_{i3}(2s_2^1, s_2^1, \sigma_i^E) + \\
& + 2p(1-p)^2 v_{i3}(s_2^2, s_2^2, \sigma_i^E) + (1-p)^3 v_{i3}(0, \frac{1}{3}, \sigma_i^E)].
\end{aligned} \tag{42}$$

Given that increasing s_2^2 increases expected share of player 1 and, if he happens to be one of the two losers, increases his payoff and actual share, we can compute:

$$\begin{aligned}
\frac{d^2E(u_i)}{ds_1^2 ds_2^2} \quad = \quad & -2p(1-p)^2 \frac{\partial \sigma_i^E}{\partial s_2^2}[v_{i13}(1-2s_1^2, 1-2s_1^2, \sigma_i^E) + v_{i23}(1-2s_1^2, 1-2s_1^2, \sigma_i^E)] + \\
& \frac{\partial \sigma_i^E}{\partial s_1^2}\frac{\partial \sigma_i^E}{\partial s_2^2}[p^3 v_{i33}(1, \frac{1}{3}, \sigma_i^E) + 2p^2(1-p)v_{i33}(1-s_1^1, \frac{1-s_1^1}{2}, \sigma_i^E) + \\
& + p(1-p)^2 v_{i33}(1-2s_1^2, 1-2s_1^2, \sigma_i^E) + p^2(1-p)v_{i33}(2s_2^1, s_2^1, \sigma_i^E) + \\
& + 2p(1-p)^2 v_{i33}(s_2^2, s_2^2, \sigma_i^E) + (1-p)^3 v_{i33}(0, \frac{1}{3}, \sigma_i^E)] + \\
& + \frac{\partial \sigma_i^E}{\partial s_1^2}[2p(1-p)^2 v_{i31}(s_2^2, s_2^2, \sigma_i^E) + 2p(1-p)^2 v_{i32}(s_2^2, s_2^2, \sigma_i^E)]
\end{aligned} \tag{43}$$

First note that, given that $c = 1$ when there is only one winner, $|v_{i13}(1-2s_1^2, 1-2s_1^2, \sigma_i^E)| < |v_{i23}(1-2s_1^2, 1-2s_1^2, \sigma_i^E)|$ holds by virtue of Assumption 5.Further, $v_{i22} \leq 0$ and

$$\frac{\partial \sigma_i^E}{\partial s_1^2} = -\frac{P(\text{win, 1 loser})}{TMP^E} = -\frac{2p^2(1-p)}{2} = -p^2(1-p) < 0, \tag{44}$$

$$\frac{\partial \sigma_i^E}{\partial s_2^2} = \frac{P(\text{lose, 1 other loser})}{TMP^E} = \frac{2(1-p)^2 p}{2} = (1-p)^2 p > 0, \tag{45}$$

I can now rewrite the expression above as:

$$\begin{aligned}
\frac{d^2E(u_i)}{ds_1^2 ds_2^2} &= POS - 2p(1-p)^2(1-p)^2p[v_{i13}(1-2s_1^2, 1-2s_1^2, \sigma_i^E) + \\
&\quad v_{i23}(1-2s_1^2, 1-2s_1^2, \sigma_i^E) + \frac{1}{2}v_{i33}(1-2s_1^2, 1-2s_1^2, \sigma_i^E)] + \\
&\quad -p^2(1-p)2p(1-p)^2[v_{i31}(s_2^2, s_2^2, \sigma_i^E) + v_{i32}(s_2^2, s_2^2, \sigma_i^E) + v_{i33}(s_2^2, s_2^2, \sigma_i^E)] \\
&> P - 2p^2(1-p)^4[v_{i13}(\cdot) + \frac{1}{TMP}v_{i23}(\cdot) + \frac{1}{TMP}v_{i33}] - p^3(1-p)^3[v_{i13}(\circ) + \\
&\quad \frac{1}{TMP}v_{i23}(\circ) + \frac{1}{TMP}v_{i33}(\circ)] \\
&> 0
\end{aligned} \tag{46}$$

where

$$\begin{aligned}
POS &= \frac{\partial \sigma_i^E}{\partial s_1^2}\frac{\partial \sigma_i^E}{\partial s_2^2}[p^3 v_{i33}(1, \frac{1}{3}, \sigma_i^E) + 2p^2(1-p)v_{i33}(1-s_1^1, \frac{1-s_1^1}{2}, \sigma_i^E) + \\
&\quad p^2(1-p)v_{i33}(2s_2^1, s_2^1, \sigma_i^E) + (1-p)^3 v_{i33}(0, \frac{1}{3}, \sigma_i^E)],
\end{aligned} \tag{47}$$

is a positive number

and the last inequality follows because expressions in brackets have to be negative by virtue of Assumption 5.

I have thus proven that increasing sacrifice of the other (s_2^2). increases marginal gain from own sacrifice. Thus, for internal solutions, optimal own gift is increased. Similar reasoning applies to effects of changing any of s_2^1

and s_2^2 on marginal effects of changes in s_1^1 and s_1^2. Thus players expecting more help from the others (if this belief is given by a single point predictor), are inclined to help more themselves.

If beliefs are expressed as a distribution with joint density $f(s_2^1, s_2^2)$, expected utility is given by:

$$E(v_i) = \iint\limits_{s_2^1=0, s_2^2=0}^{s_2^1=1, s_2^2=0.5} E(v_i|s_2^1, s_2^2) f(s_2^1, s_2^2) ds_2^2 ds_2^1 \tag{48}$$

Clearly, because $E(v_i|s_2^1, s_2^2)$. is concave in s_1^2 for all values of s_2^1, s_2^2, I can solve the FOC for a unique maximum

$$\frac{dE(v_i)}{ds_1^2} = \iint\limits_{s_2^1=0, s_2^2=0}^{s_2^1=1, s_2^2=0.5} \frac{dE(v_i|s_2^1, s_2^2)}{ds_1^1} f(s_2^1, s_2^2) ds_2^2 ds_2^1 \tag{49}$$

Given the fact that $\frac{dE(v_i|s_2^1, s_2^2)}{ds_1^2}$ is, as shown before, increasing in s_2^1 and s_2^2, we conclude that if distribution with density $f(s_2^1, s_2^2)$. is replaced by another distribution with density $g(s_2^1, s_2^2)$. such that new marginal distribution of s_2^2 stochastically dominates the old one, $\frac{dE(v_i)}{ds_1^2}$ increases. Thus, maximum of the function $E(v_i)$. is now taken for greater s_1^2, as claimed. Similarly for s_1^1 and for effect of increasing s_2^1 (again, in terms of first-order stochastic dominance).