It hurts more to lose an unfair game. On dynamic psychological games of fairness.

Michal Wiktor Krawczyk
Center for Research in Experimental Economics
and Political Decision-Making (CREED).
Universiteit van Amsterdam. Roetersstraat 11, Amsterdam 1018 WB NL.
e-mail: m.w.krawczyk@uva.nl

Keywords: procedural justice, distributive justice, dynamic psychological games.

March 31, 2006

In this paper I present a new model aimed at predicting behavior in games involving risk. The model is designed to capture the relative importance and interaction between procedural justice and distributive justice. Departing from the standard consequentialist perspective, I look beyond sheer outcomes of interactions by incorporating also expected payoffs, given strategies. While keeping the model parsimonious and avoiding explicit reference to players' intensions, I am able to account for several regularities observed both in the lab and in the field that jointly pose a challenge to classic models, social-utility models and intentions-based models alike.

Dependence of the motivation function on the expected payoffs (which in turn depend on the strategies) cannot be accounted for by classical game theory (in which truncations of strategies to non-played paths are payoff-irrelevant). Therefore I make use of dynamic psychological game theory (Battigalli and Dufwenberg 2005). I begin by considering an 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 $h = (a^1, ..., a^l) \in H$ where each $a^t = (a_0^t, a_1^t, ..., a_n^t)$ represents the profile of actions chosen at stage t $(1 \leq t \leq l)$. Null history (before any actions are made) is dented by h^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.

We let $S_{i \in N}$ denote the set of (pure) strategies of player i. Individual strategy is denoted by $s_i = (s_{i,h})_{h \in H \setminus Z}$, where $s_{i,h} \in A_i(h)$ 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, we 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.

We 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 Battigalli and Dufwenberg (2005) pages 16-19.

We assume that material interest aside, individuals care both for the equality of actual payoffs (i.e. distributive justice) and the equality of expected payoffs, where expectation is taken over the strategies of other players (using correctly updated beliefs about those) and a (mixed) strategy of "nature". This comparison of expected payoffs serves as a proxy for procedural justice. While a symmetric lottery is a prototypical fair procedure (see Rawls 1971, page 86) as it is transparent, impartial and gives everyone an equal share in the long run, there are of course aspects of procedural fairness that cannot be captured in terms taken from probability theory. Extensions of the model to other operationalizations of procedural justice are discussed in the paper.

For every terminal history each player's share in the game can be defined:

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

where $c = \sum_{j=1}^{n} y_j(z)$ is the total material payoff at z. Further note that strategy and beliefs (at z) of player i jointly determine his expected vector of payoffs of all players $(y_1^{E,i}, y_2^{E,i}, ...y_n^{E,i})$ in the game, where expectation is taken over all strategies of others that allow z and all possible choices of nature:

$$y_j^{E,i}(z) = \sum_{s_0 \in S_0} \sum_{s_{-i} \in S_{-i}(z)} \mu_i(s_0|h^0)\mu_i(s_{-i}|z) y_j(\zeta(s, s_0)),$$

where s is a strategy profile combining s_i and s_{-i}. Thus $y_j^{E,i}(z)$ is what player i thinks player's j expected payoff at the onset of the game was, given what he (i) has learned about strategies of other players (including j but not the nature) observing actually realized history z. Note also that, by definition of belief consistency, $\mu_i(s_0|h^0)$ are identical for all players. Similarly $\mu_i(s_k|z) = \mu_j(s_k|z)$ for $i \neq j \neq k$.

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

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

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

Players are assumed to maximize expected value of the motivation function that can now be written as:

$$v_i(y_i(z), \sigma_i(z), \sigma_i^E(z)),$$

The model employs standard assumptions regarding utility of money: $v_{i1} > 0, v_{i11} \leq 0$. I also assume that, holding other things constant, motivation function is concave in expected share and in actual share: $v_{i22} \leq 0, v_{i33} \leq 0$. Absolute marginal valuation of expected share is also assumed to be non-decreasing in the size of the "pie".

Perhaps most important and non-trivial assumption is that individuals display sensitivity not only to procedural and distributive justice per se, but also to the interaction between the two, namely that the marginal utility of actual (expected) share decreases in expected (actual) share, $v_{i23} \leq 0$.

This assumption is supported by substantial psychological evidence that deviations from standards of procedural and distributive fairness are particularly unattractive if they go in the same direction, i.e. procedures and outcomes are simultaneously unfavorable or overly favorable (see Brockner & Wiesenfeld 1996 for an overview).

In the further part of the paper some of the model's predictions are examined. Several laboratory games that allow for fairness considerations and employ risk factor are found to lend full support to the model. Among other examples, it correctly predicts giving behavior in the "solidarity game" by Selten&Ockenfels (1998) as well as the link between the magnitude of gifts and expectation on other's gifts. It also accounts for the importance of intended offer in randomly perturbed ultimatum game (Kramer et al. 1995) and responses to randomly generated offers (Blount 1995, Bolton et al. 2005, Cox&Deck 2005), depending on (the direction of) the bias of the randomization procedure. Essentially, players must set higher rejection threshold for strongly unfair random mechanisms, to reduce procedural injustice. In similar vein, the model also offers explanation alternative to the standard, "attributional" one, of the fact that randomized computer-made offers in ultimatum-like games are less often rejected than human-made offers. Further, the model correctly predicts positive offers in a version of dictator game involving division of lottery tickets (Karni et al. 2001). No outcome-based models can explain these results and most of them are also difficult to account for in intention-based models.

In the domain of public economics, my model suggests i.a. that support for redistribution policies (aimed at fostering distributive justice) be stronger in economies where little vertical mobility is believed to exist (low procedural justice). This is in line with findings e.g. in a study by Alesina et al. (2001) which compares Europe to the United States.

I conclude with suggestions for further verification of the model and possible extensions.

References

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

[2] Battigalli, P. and M. Dufwenberg (2005), "Dynamic Psychological Games," IGIER Working Paper No. 287.

[3] 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.

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

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

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

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

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

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

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