Supplement to "Magical Thinking: A Representation Result"
(For Online Publication)

Brendan Daley[†] Philipp Sadowski[‡]

April 29, 2016

This supplement contains extended formal results for Daley and Sadowski (2016) (henceforth DS16). Specifically, §S.1 establishes that in Prisoners' Dilemma (PD) games, the model of DS16 is logically distinct from three models that employ well-known forms of other-regarding preferences: altruism (Ledyard, 1995; Levine, 1998), inequity aversion (Fehr and Schmidt, 1999), and reciprocity (Rabin, 1993). §S.2 provides an axiomatic characterization of F—the perceived distribution of types in the model—being empirically valid when there are infinitely many players. §S.3 extends the axiomatic analysis to symmetric 2×2 games beyond PD games. All references to numbered sections/axioms/results/etc. are from DS16, unless otherwise indicated.

S.1 Models with Other-regarding Preferences

Consider the class of games denoted PD from Section 2, in which each game is parameterized by a pair $(x, y) \in R^2_{++}$.[1] The representation result (Theorem 1) establishes that under a condition on the slope of F, the data generated by the unique equilibrium behavior in PD of any such model satisfies four axioms, and for any data set that satisfies the axioms there exists a model, satisfying the same slope condition on F, that can explain it.[2]

Of course, there may exist other equivalent representations. As well-known models employing what are referred to as "other-regarding preferences" can sometimes accommodate cooperation by some players in some games in PD, they may seem candidates for this equivalence. In this section, we demonstrate that the models endowed with three of the most popular forms of other-regarding preferences are logically distinct from our model on PD.

Let u_i, u_j be the payoffs to players i and j as specified by the outcome of a two-player game. In each of the three models, player i seeks to maximize a different objective, which we denote v_i.

1. **Altruism**. As proposed by Ledyard (1995) and further studied by Levine (1998): $v_i = u_i + \alpha_i u_j$, where $0 \leq \alpha_i < 1$; player i may care about his opponent's payoff, but not more than his own.

2. **Inequity Aversion**. As proposed by Fehr and Schmidt (1999): $v_i = u_i - \alpha_i \max\{u_j - u_i, 0\} - \delta\alpha_i \max\{u_i - u_j, 0\}$, where $0 \leq \alpha_i$ and $0 < \delta < \min\left\{1, \frac{1}{\alpha_i}\right\}$; player i may dislike inequity, but dislikes it more if his is the smaller payoff, and is not willing to "burn" his own payoff to create equity.[3]

[†]The Fuqua School of Business, Duke University. E-mail: bd28@duke.edu

[‡]Department of Economics, Duke University. E-mail: p.sadowski@duke.edu

[1]Because the purpose of this supplement is to demonstrate that the alternative models are behaviorally distinct from that of DS16, it suffices to establish the result on the subclass of games $PD \subset PD^0$.

[2]Recall that the four axioms are Axioms 2-5, as Axiom 1 is needed only for the larger set of games PD^0.

[3]One could consider a more general version in which the $\delta\alpha_i$ term is replaced by β_i. That is, players can have two-dimensional types. This would not alter our result.

3. **Reciprocity**. As proposed by Rabin (1993), player i cares about how "fair" he and his opponent are being to one another. Fixing the action of player i, a_i, how fair player j is being to player i is captured by the "kindness" function $K_j(a_j|a_i)$. In the prisoners' dilemma, once a_i is fixed all outcomes are Pareto optimal. In this case,

$$K_j(a_j|a_i) = \frac{u_i(a_i, a_j) - \frac{1}{2}\left(u_i^h(a_i) + u_i^l(a_i)\right)}{u_i^h(a_i) - u_i^l(a_i)}$$

where $u_i^h(a_i)$ and $u_i^l(a_i)$ are, respectively, the highest and lowest possible payoffs to i given a_i. Finally, $v_i = u_i + \alpha_i K_j(1 + K_i)$, where $\alpha_i \geq 0$.

The original specifications of these models did not include heterogeneity in the degree to which players are other-regarding. To incorporate heterogeneity into these models, in each we assume there is a common prior that α-types are drawn i.i.d. from a continuous distribution with support $[\underline{\alpha}, \overline{\alpha}]$, where $\underline{\alpha} \geq 0$, and CDF F. Complete homogeneity can be thought of as a limiting case as $(\overline{\alpha} - \underline{\alpha}) \to 0$. The equilibrium notion remains as in Definition 2.1, with $V_i(\cdot)$ suitably adapted to each model.

It is not our goal here to provide a comprehensive analysis of these models (which, while doable, would require a considerably longer treatment), but to establish the following.

Proposition S.1 *Fix any model of those described above and an equilibrium, (σ_g, P_g), for each game $g \in PD$ and consider the resultant data of all collections of arbitrary size n. Either, for all collections I, $D_i = PD$ for all $i \in I$, or there is a positive measure of collections (according to the common prior, F) each of whose data violates Axioms 2-5.*[4]

The result is proved in the subsequent analysis.

S.1.1 Altruism

Fixing any $(x, y) \in PD$ and an equilibrium (σ, P),

$$V_i(c|x, y, P) = (1 - P)(1 + \alpha_i \cdot 1) + P(-y + \alpha_i(1 + x))$$
$$V_i(d|x, y, P) = (1 - P)(1 + x + \alpha_i(-y)) + P(0 + \alpha_i \cdot 0)$$

Therefore, $V_i(c|x, y, P) - V_i(d|x, y, P) = \alpha_i(1 + Px + (1 - P)y) - (1 - P)x - Py$. This expression is strictly increasing in α_i for all $(x, y), P$. Hence, all equilibria are cutoff equilibria.

For any given $(x, y) \in PD$, there exists an equilibrium with cutoff type α if and only if given $\alpha_i = \alpha$, $V_i(c|x, y, F(\alpha)) = V_i(d|x, y, F(\alpha))$. For any α, let $\tilde{M}_\alpha$ be the set of games in which there exists an equilibrium in which α is the cutoff type. Algebraically,

$$\tilde{M}_\alpha = \left\{(x, y) \in PD \middle| y = \frac{\alpha}{(1 + \alpha)F(\alpha) - \alpha} - \frac{1 - (1 + \alpha)F(\alpha)}{(1 + \alpha)F(\alpha) - \alpha} \cdot x\right\}.$$

Clearly, for all $i \in I$, $M_i \subset \tilde{M}_{\alpha_i}$.

[4]Further, the proposition remains valid if collections are formed via i.i.d. draws from any distribution with support $[\underline{\alpha}, \overline{\alpha}]$, even if its CDF differs from the one perceived by the players, F.

We now argue that for any F, there exists a (generic) collection drawn from its support whose equilibrium play violates the axioms. First, let $\alpha^0 < \overline{\alpha}$ be the unique solution to $F(\alpha^0) = \frac{1}{1+\alpha^0}$. For all $\alpha \in [\alpha^0, \overline{\alpha}]$, $\tilde{M}_\alpha$ forms a line in PD that is weakly *upward* sloping. So, for any player i with $\alpha_i \in [\alpha^0, \overline{\alpha}]$ to be consistent with *Continuity* (Axiom 2) and *Monotonicity* (Axiom 3), it must be that $M_i = \emptyset$.[5] Second, fix arbitrary $\alpha \in [\alpha^0, \overline{\alpha}]$. Simple algebra shows that in the game $\left(\frac{\alpha}{1-\alpha}, \frac{\alpha}{1-\alpha}\right) \in PD$, α is the unique equilibrium cutoff, so must be in M_i for any i such that $\alpha_i = \alpha$. Hence, any player drawn from a high enough quantile of the distribution will have a violation.

The intuition for this is easy to see. Suppose that $\alpha_i = \overline{\alpha}$, so $F(\alpha_i) = 1$. Then, if in game (x, y), i is indifferent between c and d, all other players are choosing d. Therefore, i's indifference condition is $V_i(c|x, y, 1) = -y + \alpha_i(1 + x) = V_i(d|x, y, 1) = 0$. An increase in x *increases* $V_i(c)$ because it increases i's opponent's payoff, which i values altruistically. This makes player i strictly prefer c to d, and violates *Monotonicity*.

S.1.2 Inequity Aversion

Fixing any $(x, y) \in PD$ and an equilibrium (σ, P),

$$V_i(c|x, y, P) = (1 - P)(1 - \alpha_i \cdot 0) + P(-y - \alpha_i(1 + x + y))$$
$$V_i(d|x, y, P) = (1 - P)(1 + x - \delta\alpha_i(1 + x + y)) + P(0 - \alpha_i \cdot 0)$$

Therefore, $V_i(c|x, y, P) - V_i(d|x, y, P) = \alpha_i(1+x+y)(\delta - P(1+\delta)) - (1-P)x - Py$. This expression is negative for $\alpha_i = 0$, monotonic in α_i, and increasing in α_i if and only if $P < \frac{\delta}{1+\delta} \leq \frac{1}{2}$. This immediately implies that all players defecting regardless of type (i.e., $P = 1$) is an equilibrium for any $(x, y) \in PD$. It also implies that if, for a given game, there exists an equilibrium in which a type cooperates, then it is a cutoff equilibrium where the cutoff type α^* must satisfy $F(\alpha^*) < \frac{\delta}{1+\delta} \leq \frac{1}{2}$.

Fix now any player i with α_i such that $F(\alpha_i) > \frac{1}{2}$. From above, $M_i = \emptyset$. Notice, though, that in any game, in any equilibrium where any type cooperates, player i cooperates. Therefore, we have the following two cases:

Case 1: Suppose $C_i = \emptyset$. Then, by the previous paragraph, in every game players are coordinating on the "all defect" equilibrium. Therefore, $D_j = PD$ for all $j \in I$, consistent with Proposition S.1.

Case 2: Suppose $C_i \neq \emptyset$. Then, given $M_i = \emptyset$, for player i to satisfy *Continuity* (Axiom 2), it must be that $D_i = \emptyset$. We now show that this cannot hold. To see this notice that i) $V_i(c|x, y, P) - V_i(d|x, y, P)$ is monotonic (in fact, linear) in P, and ii) $V_i(c|x, y, 1) - V_i(d|x, y, 1) = -y - \alpha_i(1 + x + y) < 0$ for all α_i and $(x, y) \in PD$. Therefore, if $V_i(c|x, y, 0) - V_i(d|x, y, 0) < 0$, then there is no equilibrium for game (x, y) in which i cooperates.

$$V_i(c|x, y, 0) - V_i(d|x, y, 0) = \delta\alpha_i(1 + y) + x(-1 + \delta\alpha_i)$$

Since $\delta\alpha_i < 1$, this is negative if $x > \frac{\delta\alpha_i(1+y)}{1-\delta\alpha_i}$. For any fixed y, there exist large enough x-values to satisfy this inequality for all α_i. Hence, $D_i \neq \emptyset$, violating Axiom 2.

[5] Suppose not, and that $(x, y) \in M_i$. Then to satisfy Axiom 3, i) all other $(x', y') \in \tilde{M}_{\alpha_i}$ cannot be in M_i (so $M_i = \{(x, y)\}$), and ii) $C_i \neq \emptyset$ and $D_i \neq \emptyset$. But then Axiom 2 is clearly violated.

S.1.3 Reciprocity

It is easy to calculate that for any pair of players i, j and $(x, y) \in PD$, regardless of a_i, $K_j(a_j = d|a_i) = -\frac{1}{2}$ and $K_j(a_j = c|a_i) = \frac{1}{2}$. So, fixing any $(x, y) \in PD$ and an equilibrium (σ, P),

$$V_i(c|x, y, P) = (1 - P)(1 + \tfrac{3}{4}\alpha_i) + P(-y - \tfrac{3}{4}\alpha_i)$$
$$V_i(d|x, y, P) = (1 - P)(1 + x + \tfrac{1}{4}\alpha_i) + P(0 - \tfrac{1}{4}\alpha_i)$$

From here, the analysis is analogous to that performed for inequity-averse players. $V_i(c|x, y, P) - V_i(d|x, y, P) = \alpha_i \left(\frac{1}{2} - P\right) - (1 - P)x - Py$. This expression is negative for $\alpha_i = 0$, monotonic in α_i, and increasing in α_i if and only if $P < \frac{1}{2}$. This immediately implies that all players defecting regardless of type (i.e., $P = 1$) is an equilibrium for any $(x, y) \in PD$. It also implies that if, for a given game, there exists an equilibrium in which a type cooperates, then it is a cutoff equilibrium where the cutoff type α^* must satisfy $F(\alpha^*) < \frac{1}{2}$.

Fix now any player i with α_i such that $F(\alpha_i) > \frac{1}{2}$. From above, $M_i = \emptyset$. Notice, though, that in any game, in any equilibrium where any type cooperates, player i cooperates. Therefore, we have the following two cases:

Case 1: Suppose $C_i = \emptyset$. Then, by the previous paragraph, in every game players are coordinating on the "all defect" equilibrium. Therefore, $D_j = PD$ for all $j \in I$, consistent with Proposition S.1.

Case 2: Suppose $C_i \neq \emptyset$. Then, given $M_i = \emptyset$, for player i to satisfy *Continuity* (Axiom 2), it must be that $D_i = \emptyset$. We now show that this cannot hold. To see this notice that i) $V_i(c|x, y, P) - V_i(d|x, y, P)$ is monotonic (in fact, linear) in P, and ii) $V_i(c|x, y, 1) - V_i(d|x, y, 1) = -\left(\frac{\alpha_i}{2} + y\right) < 0$ for all α_i and $(x, y) \in PD$. Therefore, if $V_i(c|x, y, 0) - V_i(d|x, y, 0) < 0$, then there is no equilibrium for game (x, y) in which i cooperates.

$$V_i(c|x, y, 0) - V_i(d|x, y, 0) = \frac{\alpha_i}{2} - x$$

This is negative if $x > \frac{\alpha_i}{2}$. Hence, $D_i \neq \emptyset$, violating Axiom 2.

S.2 Large Collections and Empirically Valid *F*

We say that F, the commonly perceived distribution of types in the model, is *empirically valid* if it agrees with empirical distribution of types in the collection. If so, magical thinking is the sole source of error in players' beliefs, and we refer to them as being *calibrated*. One issue that arises in our context, but not in axiomatic theories of individual choice, is the lack of data in the primitive itself. There, the primitive is typically assumed to be the agent's preference relation over all possible acts/choices. While our primitive includes each player's preferences over actions in all games in the domain, the collection of players is assumed to be finite.[6] It is easy to see that this precludes the observation of almost all α-types in $[0, 1]$ and therefore the recovery of a unique F from the primitive. In addition, even if adhering to the population/sample interpretation discussed

[6]There are common experimental techniques to circumvent the requirement of collecting infinite data on individual choice. In particular, infinite data can be approximated by finite data, indifference points can be elicited directly, or the individual can be asked to specify a decision rule. In contrast, the concern about the number of players in the sample is novel.

in Section 2.3, it is difficult to give behavioral meaning to the empirical validity of F when the analyst's data is generated by a finite collection.

To address both of these issues, in this supplement we let the collection of players be the interval $I = [0, 1]$, endowed with the Lebesgue measure. This can be thought of as an approximation of an arbitrarily large collection or of drawing an arbitrarily large (and therefore completely representative) random sample in the population/sample interpretation, or as simply satisfying a theoretical curiosity. For simplicity, we consider the domain to be PD, and primitive $(D_i, C_i)_{i\in I}$.[7] In order for analysis to be tractable, we assume that the following are Lebesgue measurable: for all $(x, y) \in PD$, the sets $\{i \in I|(x, y) \in D_i\}$ and $\{i \in I|(x, y) \in C_i\}$, and for any arbitrary individual behavior (D, C), the set $\{i \in I|(D_i, C_i) = (D, C)\}$.

Axioms 2-5 immediately apply to the larger set of behavioral data, but they are more restrictive in the following sense.

Definition S.2 *Let $\mathcal{M}$ be the set of behavioral models, $\left[F, (\alpha_i)_{i\in I}\right]$, for which* (i) *$F$ is continuous on $[0, 1)$,* (ii) *if $\alpha < \alpha'$, then $F(\alpha') \leq F(\alpha)\frac{\alpha'(1-\alpha)}{\alpha(1-\alpha')}$,* (iii) *if, for $i \in I$, $\alpha_i \in (0, 1)$, then $F(\alpha_i) \in (0, 1)$, and* (iv) *if, for $\{i, j\} \subset I$, $\alpha_i < \alpha_j$, then $F(\alpha_i) < F(\alpha_j)$.*

Proposition S.2 *The primitive $(D_i, C_i)_{i\in I}$ satisfies Axioms 2-5 if and only if it can be explained by a behavioral model $\left[F, (\alpha_i)_{i\in I}\right] \in \mathcal{M}$. Furthermore, for all $i \in I$, α_i is unique, and if $\alpha_i > 0$, then $F(\alpha_i)$ is unique.*

First, the convenient assumption that F is differentiable has no behavioral content in the case of finite I, but is no longer without loss of generality when I is a continuum. Consequently, the class of behavioral models $\mathcal{M}$ does not require differentiability. Further, (ii) is the meaningful content of Condition S without differentiability.[8] We show that it is both necessary and sufficient for uniqueness of the equilibrium cutoff in all games. Second, while full support is not implied by the axioms when I is finite, it does encompass (iii) and (iv) (which are now joint restrictions on F and $(\alpha_i)_{i\in I}$). Finally, notice that atoms at $\alpha = 0, 1$ are permitted.

Definition S.3 *Given any $(\alpha_i)_{i\in I}$, let $\widehat{F}$ be the CDF of types in I.*

If the analyst views I as a perfectly representative sample of a grand population, then it is easy to evaluate whether or not $F(\alpha_i)$ is empirically valid for any $\alpha_i > 0$: simply compare the uniquely recovered value $F(\alpha_i)$ to $\widehat{F}(\alpha_i)$, which is identical to the population CDF by hypotheses. Any disagreement between the two represents miscalibration of the players.

There are two concerns with this evaluation method. First, it is *ad hoc* in that the analyst compares objects derived from the representation, instead of testing properties of the primitive directly. Second, the analyst cannot be sure that players are correctly calibrated regarding $F(\alpha)$ for $\alpha \notin (\alpha_i)_{i\in I}$. We now establish the behavioral content of the empirical validity of F (i.e., $F = \widehat{F}$), thereby eliminating both concerns.

Our first additional axiom rules out atoms of players with identical, nonextreme behavioral data. That is, there may be positive masses of players who strictly prefer to defect in all games, or strictly prefer to cooperate in all games. But, of all the players who exhibit both weak preference for defection and weak preference for cooperation somewhere within PD, it would seem nongeneric for a mass of them to cluster on any given (D, C) pair. Formally, for arbitrary (D, C), let $\mathcal{L}(D, C)$ be the Lebesgue measure of the set $\{i \in I|(D_i, C_i) = (D, C)\}$.

[7] Extending results to PD^0 is trivial via Axiom 1.

[8] If F is differentiable, then Conditions $S \iff$ (ii).

Axiom 6 (Smooth Data)
For all (D, C) such that $D \neq PD$ and $C \neq PD$, $\mathcal{L}(D, C) = 0$.

Next, in our behavioral model, player i compares the perceived benefit of cooperation, α_i, with the perceived cost of cooperation, $(1-\alpha_i)\left(x\left(1-P\left(x,y\right)\right)+yP(x,y)\right)$, where $P(x,y)$ is the perceived probability that a random opponent in I will defect contingent on not being influenced by i. If $(x,y) \in M_i$, then i is indifferent between c and d, so $x\left(1-P\left(x,y\right)\right)+yP(x,y) = \frac{\alpha_i}{1-\alpha_i}$. That is, $x\left(1-P\left(x,y\right)\right)+yP(x,y)$ is constant on M_i. If player i is correctly calibrated, then the perceived probability $P\left(x,y\right)$ should coincide with the empirical frequency of defection in the population.

Definition S.4 *Given $(D_i, C_i)_{i\in I}$, for each $(x,y) \in PD$, define $\widehat{P}(x,y)$ as the Lebesgue measure of the set $\{i \in I | (x,y) \in D_i\}$, and let $Q(x,y) := x\left(1-\widehat{P}\left(x,y\right)\right)+y\widehat{P}(x,y)$.*[9]

Because i cannot, in fact, directly influence his opponent's action choice, for each $(x,y) \in PD$, $Q(x,y)$ represents the true expected (opportunity) cost of cooperating in (x,y) against a random opponent in I. Our final axiom captures correct calibration by requiring this true cost of cooperation to be constant on M_i.

Axiom 7 (Willingness to Pay for Own Cooperation)
For all $i \in I$, if $\{(x,y),(x',y')\} \subset M_i$, then $Q\left(x,y\right) = Q\left(x',y'\right)$.

To motivate the axiom without invoking the representation, imagine that there is a grand population, and that over time i plays various games in PD against random opponents from the population. In addition, I is a perfectly representative sample from this population. If player i cooperates in a given game, he does so at a cost to his own game-payoff due to some nonstandard feature affecting his choice behavior, commonly referred to as a *bias* (not necessarily magical thinking). The axiom states that there is a single level for this true cost such that i is equally drawn to playing optimally (defecting) or being overcome by his bias to play suboptimally (cooperating). That is, $Q(x,y)$ for arbitrary $(x,y) \in M_i$, is the maximum cost associated with cooperation that i can endure.[10]

Proposition S.3 *The primitive $(D_i, C_i)_{i\in I}$ satisfies Axioms 2-7 if and only if there exists $(\alpha_i)_{i\in I}$ such that* (i) $\left[\widehat{F}, (\alpha_i)_{i\in I}\right]$ *explains $(D_i, C_i)_{i\in I}$ and* (ii) $\left[\widehat{F}, (\alpha_i)_{i\in I}\right] \in \mathcal{M}$. *Furthermore, for all $i \in I$, α_i is unique.*

Given Proposition S.2, this shows that Axioms 6 and 7 are the behavioral content of empirical validity. In fact, the role of each of the two can be isolated. Axiom 7 is the content of players being

[9]Notice that we are interpreting $\widehat{P}(x,y)$ as the empirical analog of $P(x,y)$. Within the context of our axioms this is valid. However, in general, the empirical frequency of defection in game (x,y) may depend on the implementation of actions by players for which $(x,y) \in M_i$. This can be accommodated in a straightforward manner (see footnote 15).

[10]Recall that PD normalizes the payoff from (c,c) and (d,d). If considering all of PD^0 and applying Axiom 1, this maximum cost would be interpreted on a relative scale: if the stakes are higher all-around, then this maximum cost is likewise higher. This is consistent with an interpretation that i perceives gaining something from cooperating that scales with the game's payoffs. It is inconsistent with a bias such as *inattention* or *cognitive costs*, where i chooses cooperation only when the stakes are too small to bother figuring out the correct choice.

correctly calibrated about the types in the collection: $F(\alpha_i) = \widehat{F}(\alpha_i)$ for all $i \in I$. It is slightly more subtle to see that Axiom 6 is needed to ensure they are also correctly calibrated in their beliefs about those types *not* in the collection (i.e., they do not assign them positive probability). This is because the original axioms (2-5) require continuity of F on $[0, 1)$. If Axiom 6 fails, then $\widehat{F}$ will not be continuous on $[0, 1)$—there still exist behavioral models in $\mathcal{M}$ that can explain the data, but none with $F = \widehat{F}$.

S.2.1 Proofs

Proof of Proposition S.2.
Representation $\Longrightarrow$ Axioms: Consider a collection I that satisfies the representation. Our first step is to establish the analogs of Propositions 1-2 in this setting generated by replacing every appearance of "$F \in \mathcal{F}$" with "(i) of Definition S.2," and "Condition S" with "(ii) of Definition S.2." The proof of the modified version of Proposition 1 follows easily. To prove the modified version of Proposition 2, let F satisfy (i), and first suppose that condition (ii) is also satisfied. For the purpose of contradiction, suppose there exists $(x, y) \in PD$ that has two equilibrium cutoffs $\alpha_1^* < \alpha_2^*$, each of which satisfy (2). Writing out these two linear equations we can attempt to solve for x and y. Notice that when $F(\alpha_1^*) = F(\alpha_2^*)$, the two equations are inconsistent, and there is no solution, contradicting the hypotheses. If $F(\alpha_1^*) < F(\alpha_2^*)$, then solving for x and y yields unique values:

$$x = \frac{\alpha_1^*(1-\alpha_2^*)F(\alpha_2^*) - \alpha_2^*(1-\alpha_1^*)F(\alpha_1^*)}{(1-\alpha_1^*)(1-\alpha_2^*)(F(\alpha_2^*) - F(\alpha_1^*))}, \quad y = x + \frac{\alpha_2^* - \alpha_1^*}{(1-\alpha_1^*)(1-\alpha_2^*)(F(\alpha_2^*) - F(\alpha_1^*))} \tag{1}$$

The denominator of x is positive. However, the numerator of x is weakly negative by condition (ii) of Definition S.2. Therefore, $(x, y) \notin PD$, contradicting the hypothesis. Hence, condition (ii) is sufficient for uniqueness of the cutoff. Second, to see that it is necessary, suppose that it is not satisfied, so there exists $\alpha < \alpha'$ such that $F(\alpha') > F(\alpha)\frac{\alpha'(1-\alpha)}{\alpha(1-\alpha')}$, which implies $F(\alpha') > F(\alpha)$. Then, setting $\alpha_1^* = \alpha$ and $\alpha_2^* = \alpha'$, (x, y) as given by (1) is in PD and simultaneously satisfies (2) for both types. Hence, there exists a game in which the equilibrium cutoff is not unique.

With this established, the remainder of the proof is analogous to the one used for Theorem 1.

Axioms $\Longrightarrow$ Representation: The proof follows the same steps as for Theorem 1. Lemma A.1 remains valid. Lemma A.2 must be modified as follows:

Lemma S.2 *Fix any player i. If (D_i, C_i) satisfies Axioms 2-4, then there exists a pair $(\alpha_i, F_i) \in [0, 1]^2$ such that (D_i, C_i) can be explained by any behavioral model $[F, (\alpha_i, \alpha_{-i})]$ such that F is continuous on $[0, 1)$ and $F(\alpha_i) = F_i$. Further, α_i is unique, and F_i is unique if and only if $C_i \neq \emptyset$, as follows:*

$$(\alpha_i, F_i) = \begin{cases} \left(\frac{int_i}{1+int_i+slp_i}, \frac{1}{1+slp_i}\right) & \text{if } D_i, C_i \neq \emptyset \\ (1, 1) & \text{if } D_i = \emptyset \\ (0, K_i), K_i \in [0, 1] & \text{if } C_i = \emptyset \end{cases} \tag{2}$$

Proof. The proof is completely analogous to the proof of Lemma A.2 except in the following places: Case 1, parts (ii) and (iii); Case 3.

Case 1:

ii) Suppose that $(x, y) \in C_i$. By Lemma A.1, this implies that $y < int_i - slp_i \cdot x$. Let $d(\alpha) := V(c|\alpha = \alpha^*) - V(d|\alpha = \alpha^*) = \alpha[1 + (1 - F(\alpha))x + F(\alpha)y] - [(1 - F(\alpha))x + F(\alpha)y]$. Using the assignments of $(\alpha_i, F(\alpha_i) = F_i)$ from (2), it follows that $d(\alpha_i) > 0$. Notice that $d(0) = x(F(0) - 1) - yF(0) < 0$ for any $F(0) \in [0, 1)$. F continuous on $[0, 1)$ implies that d is continuous $[0, \alpha_i]$. Hence, there exists $\alpha \in (0, \alpha_i)$ that achieves $d(\alpha) = 0$ and is therefore an equilibrium cutoff in the game $(x, y) \in C_i$ (by the analog of Proposition 1).

iii) Suppose that $(x, y) \in D_i$. By Lemma A.1, this implies that $y > int_i - slp_i \cdot x$. Using the assignments of $(\alpha_i, F(\alpha_i) = F_i)$ from (2), it follows that $d(\alpha_i) < 0$. Further, $\lim_{\alpha \uparrow 1} d(\alpha) = 1$. F continuous on $[0, 1)$ implies that d is continuous on $[\alpha_i, 1)$. Hence, there exists $\alpha \in (\alpha_i, 1)$ that achieves $d(\alpha_i) = 0$ and is therefore an equilibrium cutoff in the game $(x, y) \in D_i$ (by the analog of Proposition 1).

Case 3: In the behavioral model, for any F continuous on $[0, 1)$, a player i strictly prefers d in every $(x, y) \in PD$ if and only if his type is $\alpha_i = 0$. Given $\alpha_i = 0$, the value of $F(0)$ is irrelevant for i's behavior, so it cannot be determined. ■

Lemmas A.3 and A.4, with references to Lemma A.2 now made to Lemma S.2, also remain valid. Hence, for any collection whose data satisfy Axioms 2-5, using (2), each player i can be assigned a unique α_i and corresponding quantile F_i, that is also unique if $\alpha_i > 0$, and i's behavior can be explained by any model $[F, (\alpha_i, \alpha_{-i})]$ such that F satisfies (i) of Definition S.2 and $F(\alpha_i) = F_i$.

We now show that there exists a model in $\mathcal{M}$ that simultaneously explains the behavior of all $i \in I$. Let $A^+ := \{\alpha_i > 0 | i \in I\}$. The four lemmas (A.1, S.2, A.3, A.4) imply that any behavioral model that satisfies $F(\alpha_i) = F_i$ for all $i \in I$ also satisfies (iii) and (iv) of Definition S.2 as well as (i) and (ii) *restricted to the domain* A^+, where continuous on A^+ means: for every $\alpha^0 \in A^+$, every sequence $(\alpha^m)_{m \in \mathbb{N}}$, $\alpha^m \in A^+$ for all m, such that $\lim_{m \to \infty} \alpha^m = \alpha^0$ also satisfies $\lim_{m \to \infty} F(\alpha^m) = F(\alpha^0)$. All that remains is to establish existence by extending F from A^+ to $[0, 1)$ preserving continuity, (weak) monotonicity, and condition (ii) of Definition S.2. From the proof of the opposite direction above, these properties imply that the behavioral model emits a unique equilibrium cutoff in all games $(x, y) \in PD$. This ensures that in each game there is an equilibrium consistent with the behavior of all players; hence, the behavioral model using this assignment of F and $(\alpha_i)_{i \in I}$ can explain $(D_i, C_i)_{i \in I}$.

To extend F from A^+ to $[0, 1)$, consider arbitrary $\alpha^0 \in [0, 1) \backslash A^+$. There are three exhaustive cases. First, if there exists a sequence $(\alpha^m)_{m \in \mathbb{N}}$, $\alpha^m \in A^+$ for all m, such that $\lim_{m \to \infty} \alpha^m = \alpha^0$, simply assign $F(\alpha^0) = \lim_{m \to \infty} F(\alpha^m)$. Second, let $\underline{\alpha} := \inf(A^+)$ and $\overline{\alpha} := \sup(A^+)$. If $\alpha^0 < \underline{\alpha}$, assign $F(\alpha^0) = F(\underline{\alpha})$, and if $\alpha^0 > \overline{\alpha}$, assign $F(\alpha^0) = F(\overline{\alpha})$—notice that even if $\underline{\alpha}, \overline{\alpha} \notin A^+$, $F(\underline{\alpha}), F(\overline{\alpha})$ are assigned in the previous case. Third, and finally, if A^+ does not contain a sequence converging to $\alpha^0 \in [\underline{\alpha}, \overline{\alpha}]$, then $\underline{\alpha}^0 := \sup\{\alpha \in A^+ | \alpha < \alpha^0\} < \alpha^0 < \overline{\alpha}^0 := \inf\{\alpha \in A^+ | \alpha > \alpha^0\}$. Notice that even if $\underline{\alpha}^0, \overline{\alpha}^0 \notin A^+$, $F(\underline{\alpha}^0), F(\overline{\alpha}^0)$ are assigned in the first case. Let L^0 be the line that passes through both $(\underline{\alpha}^0, F(\underline{\alpha}^0))$ and $(\overline{\alpha}^0, F(\overline{\alpha}^0))$. For all $\alpha \in (\underline{\alpha}^0, \overline{\alpha}^0)$, assign $F(\alpha) = L^0(\alpha)$. It is immediate that these assignments preserve continuity and monotonicity in each case and also condition (ii) in the first and second cases. For the assignments in the third case, it is trivial to verify that the linearity of F between $[\underline{\alpha}^0, \overline{\alpha}^0]$ preserves condition (ii) on this interval, and then in general since condition (ii) is transitive. ■

Proof of Proposition S.3.

Representation $\Longrightarrow$ Axioms: Consider a collection I that satisfies the representation. The fact that $F = \widehat{F}$ is irrelevant for the proof that $(D_i, C_i)_{i\in I}$ satisfies Axioms 2-5, so this is established by Proposition S.2. Next, $F = \widehat{F}$, which is continuous on $[0,1)$, immediately implies Axiom 6. Finally, verifying Axiom 7 is a straightforward calculation: fix any player i such that $M_i \neq \emptyset$ and recall that $M_i = \left\{(x,y) \in PD \middle| y = \frac{\alpha_i}{(1-\alpha_i)F(\alpha_i)} - x\left(\frac{1-F(\alpha_i)}{F(\alpha_i)}\right)\right\}$. Therefore, for any $(x,y) \in M_i$, we can substitute the expression for y into $Q(x,y)$ to get,

$$Q(x,y) = x\big(1 - \widehat{P}\,(x,y)\,\big) + \left(\frac{\alpha_i}{(1-\alpha_i)F(\alpha_i)} - x\left(\frac{1-F(\alpha_i)}{F(\alpha_i)}\right)\right)\widehat{P}(x,y) \tag{3}$$

Given that $F = \widehat{F}$ and that α_i is the cutoff type for $(x,y) \in M_i$, $\widehat{P}(x,y) = F(\alpha_i)$; so (3) simplifies to $Q(x,y) = \frac{\alpha}{1-\alpha}$, which does not vary with (x,y).

Axioms $\Longrightarrow$ Representation: The proof of Proposition S.2, establishes that if $(D_i, C_i)_{i\in I}$ satisfies Axioms 2-5, then it can be explained by any model $[F, (\alpha_i)_{i\in I}] \in \mathcal{M}$, where α_i and $F(\alpha_i) = F_i$ are given by (2) (and therefore α_i is unique and, if $\alpha_i > 0$, so is $F(\alpha_i)$). Therefore, let $(\alpha_i)_{i\in I}$ be as given by (2), and $\widehat{F}$ be the resultant CDF. It is sufficient to show that 1) for all i such that $\alpha_i > 0$, $\widehat{F}(\alpha_i) = F_i$, and 2) $\left[\widehat{F}, (\alpha_i)_{i\in I}\right] \in \mathcal{M}$.

To see the first, notice that the structure of $(D_i, C_i)_{i\in I}$ characterized by Lemmas A.1, S.2, A.3, and A.4 implies that for any i such that $M_i \neq \emptyset$, $\widehat{P}(x,y)$ is constant and equal to $\lim_{\alpha\uparrow\alpha_i}\widehat{F}(\alpha)$ along M_i. By Axiom 6, $\lim_{\alpha\uparrow\alpha_i}\widehat{F}(\alpha) = \widehat{F}(\alpha_i)$. Consider i such that $\alpha_i \in (0,1)$, so $M_i \neq \emptyset$. For $(x,y) \in M_i$,

$$\begin{aligned} Q(x,y) &= x(1 - \widehat{P}(x,y)) + y\widehat{P}(x,y) \\ &= x(1 - \widehat{F}(\alpha_i)) + y\widehat{F}(\alpha_i) \\ &= x(1 - \widehat{F}(\alpha_i)) + (int_i - x \cdot slp_i)\widehat{F}(\alpha_i) \end{aligned}$$

By Axiom 7, Q is constant along M_i, so $\widehat{F}(\alpha_i) = \frac{1}{1+slp_i} = F_i$. If instead, $\alpha_i = 1$, then because $\widehat{F}$ is a CDF on $[0,1]$, $\widehat{F}(\alpha_i) = 1 = F_i$.

To see the second, we need to show that $\left[\widehat{F}, (\alpha_i)_{i\in I}\right]$ satisfies the four requirements of Definition S.2. Axiom 6 implies (i), and Lemmas S.2 and A.4 imply (iii) and (iv). For (ii), notice that if α and α' are elements of $(\alpha_i)_{i\in I}$, then the property holds due to Lemma A.4 and if $\alpha = 0$ or $\alpha' = 1$, the property is trivial. Consider now an arbitrary pair $0 < \alpha < \alpha' < 1$, and for the purpose of contradiction suppose that $\frac{\widehat{F}(\alpha')}{\widehat{F}(\alpha)} > \frac{\alpha'(1-\alpha)}{\alpha(1-\alpha')}$. Since $\widehat{F}$ is the CDF of $(\alpha_i)_{i\in I}$, and is continuous on $[\alpha, \alpha']$, for any $\varepsilon > 0$, there must exist $\{i,j\} \subset I$ such that $\alpha \leq \alpha_i < \alpha_j \leq \alpha'$, $\widehat{F}(\alpha_i) - \widehat{F}(\alpha) < \varepsilon$, and $\widehat{F}(\alpha') - \widehat{F}(\alpha_j) < \varepsilon$. Hence, by our supposition that $\frac{\widehat{F}(\alpha')}{\widehat{F}(\alpha)} > \frac{\alpha'(1-\alpha)}{\alpha(1-\alpha')}$, for ε small enough,

$$\frac{\widehat{F}(\alpha_j)}{\widehat{F}(\alpha_i)} > \frac{\alpha'\,(1-\alpha)}{\alpha\,(1-\alpha')} \geq \frac{\alpha_j\,(1-\alpha_i)}{\alpha_i\,(1-\alpha_j)}$$

As we just discussed, Lemma A.4 implies that $\frac{\widehat{F}(\alpha_j)}{\widehat{F}(\alpha_i)} \leq \frac{\alpha_j(1-\alpha_i)}{\alpha_i(1-\alpha_j)}$, producing a contradiction. ■

S.3 Axiomatic Analysis Beyond the PD

The defining feature of the Prisoners' Dilemma is that there are strict gains to a player for selecting d whether his opponent is playing c or d (i.e., $x, y > 0$). We first enlarge our domain by relaxing the latter. That is, we consider games in which there are strict gains from unilaterally deviating away from the better symmetric outcome. To do so, let $G^0 = \{(r, p, x, y) | r > p, x > 0\}$, with labels as in Figure 1, and let our primitive, $(D_i^0, C_i^0)_{i \in I}$, as well as $(M_i^0, \overline{D}_i^0, \overline{C}_i^0)_{i \in I}$, be extended to this larger class of games in the obvious way. Finally, define $G = \{(r, p, x, y) | r = 1, p = 0, x > 0\} \subset G^0$, with arbitrary element (x, y), and, as before, $D_i = D_i^0 \cap G$ and analogously for $C_i, M_i, \overline{D}_i$, and $\overline{C}_i$. Notice that G is the union of the games in quadrants I and IV of Figure 4.

Each of the Axioms 1-5 can be applied verbatim on this larger class of games (simply replace each PD^0 and PD with G^0 and G, respectively). In addition, with the caveat of changing all instances of "cooperate" and "defect" to "play c" and "play d," respectively, the interpretations of each of the axioms are also unchanged.

We introduce an additional axiom. Fixing all other payoff parameters, the societal benefit from (either or both) players selecting c, the action corresponding to the better symmetric outcome, is increasing in r. The following axiom requires that increases in r should increase the propensity to select c.

Axiom 8 (Sensitivity to Benefits from Action c)
For all $i \in I$, if $(r, p, x, y) \in \overline{C}_i^0$ and $r' > r$, then $(r', p, x, y) \in C_i^0$.

It is not difficult to show that the representation in Theorem 1 satisfies Axiom 8 on PD^0, meaning Axiom 8 is implied by Axioms 1-5 on this domain. On G^0, this is no longer the case.

Fact S.3 *Axioms 1-5 $\Longrightarrow$ Axiom 8 on PD^0. Axioms 1-5 $\not\Longrightarrow$ Axiom 8 on G^0.*

Notice that the axiom is consistent with the experimental evidence discussed in Section 5.1.[11] Further, in line with the axiom, Rapoport and Chammah (1965) and Minas et al. (1960) compare behavior across different Prisoners' Dilemma games and provide evidence that the fraction of players selecting c indeed increases with r.[12]

By adding Axiom 8, the representation result of Theorem 1 extends to G^0.

Theorem S.1 *The primitive $\left(D_i^0, C_i^0\right)_{i \in I}$, on G^0, satisfies Axioms 1-5 and 8 if and only if it can be explained by a behavioral model $\left[F, (\alpha_i)_{i \in I}\right]$, where $F \in \mathcal{F}$ satisfies Condition S. Furthermore, for all $i \in I$, α_i and $F(\alpha_i)$ are unique.*

The extended representation also satisfies the more stringent definition of *can explain* attained if the requirements of Definition 2.4 must instead hold in *all* equilibria (see Section 2.3).

[11] It is easy to derive that for any Hawk-Dove game $(r, 0, x, y)$, $x \neq -y$, the game $(0, 0, x, y)$ is a Battle of the Sexes game with the same symmetric Nash equilibrium. Hence, insofar as subjects adhere to the symmetric Nash equilibrium in Battle of the Sexes games, but play c more frequently than in the symmetric Nash equilibrium in Hawk-Dove games (see Section 5.1), their play is consistent with Axiom 8.

[12] Up to adding constants (as permitted once we assume Axiom 1), see games labeled G4 and G5 in Minas et al. (1960) and games numbered 1 and 4 in Rapoport and Chammah (1965). This evidence is also summarized in Table 1 of Steele and Tedeschi (1967).

Next, one can extend the domain to include games in which $x \leq 0$ and $y \leq 0$ (i.e., quadrant III of Figure 4 when r and p are normalized). In these games c is *both* the action leading to the better symmetric outcome and a dominant strategy (even without magical thinking), with the dominance being strict on the interior of the quadrant. It seems natural that all players should choose c then, as they do in the our behavioral model (Section 5.1.1). In addition, for each player i such that $M_i \cap G \neq \emptyset$, this behavior is a consequence of Axioms 1-5 and 8 when the primitive is likewise extended. Under the (seemingly mild) additional requirement that in the extended domain $\overline{C}_i \neq \emptyset$ for all $i \in I$, the representation result extends with only minor alteration.[13]

How can our axioms be extended to the games with $x \leq 0$ and $y > 0$ (i.e., quadrant II of Figure 4 when r and p are normalized)? We suggest three possible ways. First, and most immediately, one can add an axiom that specifies c as the preferred action for all players whenever $x \leq 0$ and restrict our other axioms to games with $x > 0$. Second, one can extend our theory as discussed in the context of quadrant-III games, but additionally weaken Axiom 5 to allow the extended M_i-lines to intersect when $x \leq 0$. It can then be shown that the resulting representation in terms of our behavioral model would entail that, in each game, each player selects his action in accordance with an equilibrium, implying his choice is rationalizable (but not all players will play in accordance with the same equilibrium when there are multiple).

Third, one could try to really capture if/when there is multiplicity. For instance, suppose players would be willing to participate in different profiles of play (as would be the case if they actually conceived of multiple equilibria). How could this manifest itself in behavior? Since our primitive requires each player to rank d and c for every possible game, one would need to consider a richer primitive. One possibility mirrors the menu-choice approach in theories of individual choice. The analyst could instruct players that they will face an anonymous opponent in a game in period 2. In period 1, the analyst could ask players to specify for each game whether they are willing to commit to d, to c, or whether they have a preference for flexibility in the sense that they do not want to precommit to an action choice for period 2. Such preference for flexibility could be interpreted as the anticipation of coordination on an equilibrium based on some state of the world that is unobserved (or indecipherable) by the analyst and that realizes between periods 1 and 2. One could try to formulate axioms that restrict period-1 preferences over menus of actions across games and players to ensure that multiplicity is consistent with our model. In particular, the axioms should correspond to Axioms 1-5 and 8 on quadrants I and IV.

S.3.1 Proofs

The representation proof uses the following preliminary lemma.

Axiom 8′ *For all $i \in I$, if $(x, y) \in \overline{C}_i$ and $\kappa \in (0, 1)$, then $\kappa\,(x, y) \in C_i$.*

Lemma S.3 *Under Axiom 1, Axioms 8 and 8′ are equivalent.*

[13] If extending the axioms verbatim, the representation will require that $\alpha_i \neq 0$ for all $i \in I$. Since this event already has probability one according to any $F \in \mathcal{F}$, no other change to the corresponding behavioral model is required. Alternatively, one could slightly relax the extensions of Axioms 3 and 8 and maintain the original class of behavioral models.

Proof. Suppose that Axioms 1 and 8 hold and that $\kappa \in (0,1)$. Then,

$$(x,y) \in \overline{C}_i \implies (1,0,x,y) \in \overline{C}_i^0 \underset{\text{Axiom 8}}{\implies} \big(\frac{1}{\kappa},0,x,y\big) \in C_i^0 \underset{\text{Axiom 1}}{\implies} \kappa\big(\frac{1}{\kappa},0,x,y\big) \in C_i^0$$
$$\implies (1,0,\kappa x,\kappa y) \in C_i^0 \implies (\kappa x,\kappa y) \in C_i \implies \kappa(x,y) \in C_i.$$

Hence, Axiom 8′ is implied. Now, suppose that Axioms 1 and 8′ hold and that $r' > r$. Then,

$$(r,p,x,y) \in \overline{C}_i^0 \underset{\text{Axiom 1}}{\implies} \big(1,0,\frac{x}{r-p},\frac{y}{r-p}\big) \in \overline{C}_i^0 \implies \big(\frac{x}{r-p},\frac{y}{r-p}\big) \in \overline{C}_i$$
$$\implies \frac{1}{r-p}(x,y) \in \overline{C}_i \underset{\text{Axiom 8}'}{\implies} \frac{1}{r'-p}(x,y) \in \overline{C}_i \implies \big(\frac{x}{r'-p},\frac{y}{r'-p}\big) \in C_i$$
$$\implies \big(1,0,\frac{x}{r'-p},\frac{y}{r'-p}\big) \in C_i^0 \underset{\text{Axiom 1}}{\implies} (r',p,x,y) \in C_i^0.$$

Hence, Axiom 8 is implied. ■

Proof of Fact S.3. Relying on Lemma S.3, we consider whether or not Axioms 2-5 imply Axiom 8′ on PD and G for the first and second claims respectively. For the first claim, fix player i, for whom $(D_i \cap PD, C_i \cap PD)$ satisfies Axioms 2-4, with $\overline{C_i} \cap PD \neq \emptyset$. Then, from the proof of Theorem 1, we have that either $C_i \cap PD = PD$ or $M_i \cap PD = \{(x,y) \in PD | y = int_i - slp_i \cdot x\}$ and $C_i \cap PD = \{(x,y) \in PD | y < int_i - slp_i \cdot x\}$, where int_i, slp_i are positive constants. In either case, Axiom 8′ follows immediately. For the second claim, consider a player i with $M_i = \{(x,y) \in G | y = -1-x\}$, and C_i and D_i being the strict-lower and strict-upper contour sets of M_i respectively. It is immediate that (D_i, C_i) satisfies Axioms 2-4. However, (D_i, C_i) fails Axiom 8′: for any $(x,y) \in M_i$, $\frac{1}{2}(x,y) \in D_i$. The fact that $(D_j, C_j)_{j \in I}$ satisfies Axiom 5 does not rule out the existence of such a player, meaning the result is established. ■

Proof of Theorem S.1. First, note that Lemma 1 and Propositions 1-2 (and their proofs) remain valid when each PD^0 and PD are replaced by G^0 and G respectively.

Representation $\implies$ Axioms: Given that Lemma 1 and Propositions 1-2 extend to the larger domain, the proof that the representation satisfies Axioms 1-5 is completely analogous to that provided for Theorem 1. Using Lemma S.3, we are left to verify that Axiom 8′ is satisfied. First, if $\alpha_i = 0$, then $D_i = G$ so the axiom is vacuous; and if $\alpha_i = 1$, then $C_i = G$ so the axiom is trivial. Second, if $\alpha_i \in (0,1)$ and $(x,y) \in \overline{C}_i$, then $y \leq int_i - x \cdot slp_i$, where $int_i, slp_i > 0$. It follows that, for any $\kappa \in (0,1)$, $\kappa y \leq \kappa\big(int_i - x \cdot slp_i\big) < int_i - (\kappa x) slp_i$. Hence, $\kappa(x,y) \in ML_i = C_i$, verifying the axiom.

Axioms $\implies$ Representation: The only aspect of the proof that is not completely analogous to that given for Theorem 1 is in extending the following aspect of Lemma A.1. Consider a player i for which $D_i \neq \emptyset$ and $C_i \neq \emptyset$. Such a player can be characterized by a pair (int_i, slp_i), where $slp_i > 0$. When the domain was PD, $int_i > 0$ immediately. This is no longer immediate when the domain is G. However, it is ensured by Axiom 8. Suppose to the contrary that $int_i \leq 0$. Take now $(x,y) \in M_i \subset \overline{C}_i$, which must then satisfy $y = int_i - x \cdot slp_i < 0$. But then, for any $\kappa \in (0,1)$, $\kappa y = \kappa(int_i - x \cdot slp_i) \geq int_i - (kx) slp_i$. Hence, $\kappa(x,y) \notin ML_i = C_i$, violating Axiom 8′ (and therefore also Axiom 8 by Lemma S.3). With this established, the remainder of the proofs follows identical steps to those in the proof of Theorem 1. ■

References

Daley, B. and P. Sadowski (2016): "Magical Thinking: A Representation Result," *Theoretical Economics*, Forthcoming.

Fehr, E. and K. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 114, 817–868.

Ledyard, J. (1995): "Public Goods: A Survey of Experimental Research," in *Handbook of Experimental Economics*, ed. by J. H. Kagel and A. E. Roth, Princeton University Press, 111–194.

Levine, D. K. (1998): "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, 1(3), 593–622.

Minas, J. S., A. Scodel, D. Marlowe, and H. Rawson (1960): "Some Descriptive Aspects of Two-Person Non-Zero-Sum Games. II," *J. of Conflict Resolution*, 4(2), 193–197.

Rabin, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83(5), 1281–1302.

Rapoport, A. and A. M. Chammah (1965): *Prisoner's Dilemma*, Univ. of Michigan Press.

Steele, M. W. and J. T. Tedeschi (1967): "Matrix indices and strategy choices in mixed-motive games," *J. of Conflict Resolution*, 11, 198–205.