ERC

A Theory of Equity, Reciprocity and Competition

Gary E Bolton
Smeal College of Business
Penn State University, USA

&

Axel Ockenfels
Faculty of Economics and Management
University of Magdeburg, Germany

September 1997

We demonstrate that a simple model, constructed on the premise that people are motivated by both their pecuniary payoff and their relative payoff standing, explains behavior in a wide variety of laboratory games. Included are games where equity is thought to be a factor, such as ultimatum, two-period alternating offer, and dictator games; games where reciprocity is thought to play a role, such as the prisoner's dilemma and the gift exchange game; and games where competitive behavior is observed, such as Bertrand and Cournot markets, and the guessing game.

Correspondence
Bolton: 309 Beam, Penn State University, University Park, PA 16802, USA; (814) 865-0611; fax (814) 863-2381; geb3@psu.edu. Ockenfels: University of Magdeburg, FWW, Postfach 4120, D-39016 Magdeburg, Germany; (+391) 67-12197, fax (+391) 67-12971; axel.ockenfels@ww.uni-magdeburg.de.

1. Introduction: The need for a unifying explanation

The various areas of inquiry that constitute experimental economics appear at times to be surveying distinct and isolated regions of behavior. What we see in experiments involving market institutions is usually consistent with standard notions of 'competitive' self-interest. But other types of experiments appear to foster sharply different conduct. 'Equity' has emerged as an important factor in bargaining games. 'Reciprocity,' of a type that differs from the standard strategic conception, is often cited to explain behavior in games such as the prisoner's dilemma. Many economists wonder what-if-anything connects these patterns of behavior. The issue goes to the heart of what it is that experimental economics can hope to accomplish, if only because economists have traditionally placed a high value on generality. If no connection can be found, we are left with a set of disjoint behavioral charts, each valid for no more than a limited domain. But to the extent a common pattern can be established, laboratory research presents a broad, and a potentially powerful map of economic behavior.

In this paper, we describe a simple model we call *ERC* to denote the three important kinds of behavior the theory captures: *e*quity, *r*eciprocity and *c*ompetition. We show that ERC is consistent with a wide variety of experimental observations gathered by many independent investigators. ERC is simple to apply – in part, because it is not a radical departure from standard modeling techniques. The major innovation is the premise that, along with the pecuniary payoff, individuals are motivated by a 'relative' payoff, a measure of how the pecuniary payoff compares to that of the other players. Different games present different sets of tradeoffs between pecuniary and relative gains. What ERC demonstrates – and the point we will stress – is that a simple model of how pecuniary and relative motives interact, organizes a large, and seemingly disparate group of experiments as one consistent pattern of behavior.

Three experiments provide a sense of the breadth of ERC. One experiment, reported by Forsythe, Horowitz, Savin and Sefton (1994), involves two elementary strategic situations: the ultimatum game and the dictator game. In the ultimatum game, the 'proposer' offers a division of $10, which the 'responder' can either accept or reject; the latter action leaves both players with a payoff of zero. The dictator game differs only in that the responder has no choice but to accept. The standard perfect equilibrium analysis of both games begins with the assumption that each player prefers more money to less. Consequently, the responder in the ultimatum game should accept all positive offers. Given this, the proposer should offer no more than the smallest monetary unit allowed. In the dictator game, the responder has no say, so the proposer should keep all the money. So, in both games, the proposer should end up with virtually the entire $10.

Figure 1 here.

Figure 1 displays the amounts proposers actually offered. While there is a great deal of heterogeneity, average offers for both games are clearly larger than minimal. Various authors have given these results an equity interpretation (Roth, 1995, provides a survey). But equity is insufficient to explain everything in Figure 1. Offers are plainly higher in the ultimatum game. This has to do with a fact well-known to those who do ultimatum experiments: Responders regularly turn down proportionally small offers. So proposers adjust their offers accordingly. Proposers may care about equity – they *do* give money in the dictator game – but it appears that it is responder concern for equity that drives the ultimatum game. Hence Figure 1 illustrates a subtle interplay between equity and strategic considerations – an interplay that ERC captures.[1]

The second experiment, performed by Roth, Prasnikar, Okuno-Fujiwara and Zamir (1991), concerns a simple auction market game. A single seller has one indivisible unit of a good to offer nine buyers. Exchange creates a fixed surplus. Buyers simultaneously submit offers. The seller is then given the opportunity to accept or reject the best offer. All subgame perfect equilibria have the seller receiving virtually the entire surplus.

Ten rounds of the auction market experiment were performed in each of four countries. In every case, by round 10, the transaction price had converged to subgame perfect equilibrium. Hence the experiment produces behavior that is remarkably consistent with standard theory. The same study examined ultimatum game play. While there were some differences across countries, the qualitative pattern was the same in all four places: offers were generally higher than subgame perfection predicts and a significant number of offers were rejected. Are the motives behind

[1] The results from dictator and ultimatum have been shown to be very stable when the experiment is performed with comparable instructions. Forsythe et al. show that dictator giving is stable with respect to time. Hoffman, McCabe, Shachat and Smith (1995) replicate the Forsythe et al. distribution. Bolton, Katok and Zwick (forthcoming) demonstrate that the amount the dictator gives is stable with respect to various game manipulations. Giving behavior is not restricted to people: capuchin monkeys give food in what is an animal version of the dictator game; see de Waal (1996, p. 148). Evidence on whether behavior is different when the experimenter can associate dictator actions with subject identities is mixed. Roth (1995) summarizes much of the research, and suggests an alternative interpretation for what positive evidence there is. The same article surveys the many ultimatum game experiments.

market behavior fundamentally different than those behind the ultimatum game? ERC answers ‘no, the same motivation suffices to explain both games.’

The third experiment, by Fehr, Kirchsteiger and Riedl (1993), involves what is sometimes referred to as the gift exchange game. Subjects assigned the role of firms offer a wage to those assigned the role of workers. The worker who accepts the wage then chooses an effort level. The higher the level chosen, the higher the firm's profit and the lower the worker's payoff. The game is essentially a sequential prisoner's dilemma, in which the worker has a dominant strategy to choose the lowest possible effort. The only subgame perfect wage offer is the reservation wage.

Figure 2 here.

Figure 2 compares the effort level actually provided with the wage offered. Behavior is clearly inconsistent with the horizontal line that indicates the equilibrium prediction. In fact, there is a strong positive correlation between wage and effort. This is sometimes taken as evidence for reciprocity (Fehr et al. suggest this interpretation). The dictionary defines reciprocity as a "return for something done." While there is surely some relationship between this concept and equity, the two are not equivalent. Dictator game giving may involve an assessment of what is equitable, but it does not involve reciprocity as defined by the dictionary. The positive correlation evident in Figure 2 suggests to some that we need more than fairness to explain behavior in the labor market game. Or do we? ERC implies that we do not.

We begin by laying out the basic ERC model (section 2). We then show that ERC can account for a variety of patterns reported for dictator, bargaining, and related games, including the Forsythe et al. experiment (section 3). Next we explain why the model predicts competitive behavior for a class of market games including the Roth et al. experiment, and the guessing game (section 4). We then describe some basic results having to do with one-shot dilemmas. We can say more with a parametric model. We fit the simplest possible version to the Fehr et al. data (section 5). We show that the fit is robust by estimating the Berg, Dickhaut and McCabe’s (1995) investment game experiment. We then make some observations concerning repeated dilemmas (section 6).

We are not alone in our pursuit. After laying out what ERC can do, we compare with other approaches found in the literature (section 7). One model, Bolton (1991), does well explaining simple bargaining games, but fails with others. It turns out that this model is ‘almost’ a special case of ERC (section 3.3).

2. The basic ERC model

Because the immediate purpose is to explain lab data, our guiding criterion in constructing the model is that the implied hypotheses be both applicable to the lab environment, and lab testable. Lab subjects can have no better than incomplete information about how their game partners trade-off pecuniary and relative payoffs, and this is what our propositions assume. On the other hand, in order to test the propositions, the investigator must be able to reliably measure the underlying trade-offs. We have found that much of what we need to know has to do with the thresholds at which behavior deviates from the standard self-interest assumption. This information is readily recovered from dictator and ultimatum game data. We demonstrate throughout the paper that knowing the distributions of these thresholds is sufficient to characterize many phenomena.

2.1 Formal statement of the model

We concern ourselves with n - player lab games, i = 1,2,…n, where players are randomly drawn from the population, and anonymously matched (face-to-face play is a known complicating factor). All game payoffs are monetary and non-negative, $y_i \geq 0$ for all i (this is relaxed in section 6). We assume that if a subject plays a game multiple times, she never plays with any particular subject more than once. We can therefore analyze each game as one-shot.

Each player i acts to maximize the expected value of his or her *motivation function*,

$$v_i = v_i(y_i, s_i) \tag{2.1}$$

where

$$s_i = s_i(c, y_i) = \begin{cases} y_i / c, \text{if } c > 0 \\ 1/n, \text{if } c = 0 \end{cases} \text{ is } i\text{'s relative share of the payoff,}$$

and

$$c = \sum_{j=1}^{n} y_j \text{ is the total pecuniary payout.}$$

Motivation functions may be thought of as a special class of expected utility functions. We prefer ‘motivation function’ because it emphasizes that (2.1) is a statement about the objectives that motivate behavior during the experiment. The weights individuals give these objectives may well change over the long-term, with changes in age, education, political or religious beliefs, etc. (Ockenfels and Weimann, 1996). It is, however, sufficient for our purposes that the trade-off be stable in the short term, for the duration of the experiment.[2]

The formulation in (2.1) is similar to that used by Bolton (1991) (the restrictions we place on (2.1) will differ). When extended to games with more than two players, (2.1) specifies preferences over how the payoff is distributed between 'self' and 'others,' but does not capture any preference there might be over how the payoff is distributed among the others. While this level of abstraction is sufficient to explain much of what we see in various games, there is evidence that (2.1) is more accurate than this defense might suggest. In section 3.5, we discuss several recent experiments that find that subjects pay little attention to how the payoff is distributed across the rest of the group.

The following assumptions characterize (2.1):

A0. v_i is continuous and twice differentiable on the domain of (y_i, s_i).

A1.1. Narrow self-interest: $v_{i1} \geq 0$, $v_{i11} \leq 0$.

A1.2. Tie breaker: Given two choices where $v_i(y_1, \sigma) = v_i(y_2, \sigma)$ and $y_1 > y_2$,
player *i* chooses (y_1, σ).

A2. Comparative effect: $v_{i2} = 0$ for $s_i = 1/n$, and $v_{i22} < 0$.

A0 is for mathematical convenience. A1.1 is a slightly weakened version of the standard assumption made about preferences for money. We do not assume that v_i is *strictly* increasing in the pecuniary argument since this would rule out players who care more about the relative argument than the pecuniary one (players who, for example, divide 50-50 in the dictator game). A1.2, however, insures that when presented with two alternative outcomes having the same relative argument, the player makes the choice with the higher pecuniary payoff. A2 states that, holding the pecuniary argument fixed, the motivation function is concave down in the relative

[2] Prasnikar (1997) examines three large ultimatum game data sets and concludes that the trade-off is stable even with repeated play. An objection sometimes raised to the motivation approach is that one "can explain anything by changing the utility functions." This objection implicitly assumes there is no way to invalidate the functional specification. In the lab, however, we can, and often do, perform these types of validation tests.

argument, with a maximum around the allocation at which ones own share is equal to the average share. This assumption implies that equal division has collective significance – hence we refer to equal division as the *social reference point*.[3]

The data for many of the games we will deal with exhibits a great deal of heterogeneity. The theory accounts for this by positing a tension, or trade-off, between adhering to the reference point (the comparative effect) and achieving personal gain (narrow self-interest[4]). Individuals are distinguished by how this tension is resolved. Much of what we need to know about this tension is captured by the thresholds at which behavior diverges from the narrow self-interest assumption. Each player has two thresholds, $r_i(c)$ and $s_i(c)$, defined as follows (note that $y_i = cs_i$):

$$r_i(c) := \arg\max_{s_i} v_i(cs_i, s_i),\ c > 0 \quad \text{and} \quad s_i(c)\text{: } v_i(cs_i, s_i) = v_i(0, 1/n),\ c > 0,\ s_i \leq 1/n$$

As we demonstrate in section 3, r_i corresponds to the division that *i* fixes in the dictator game, and s_i corresponds to *i*'s rejection threshold in the ultimatum game. Postulates A0 to A2 guarantee there is a unique $r_i \in [1/n, 1]$ and a unique $s_i \in (0, 1/n]$ for each *c*. Both r_i and s_i are, technically speaking, functions of *n*; for simplicity of exposition, we suppress this argument.

Postulate A3 provides an explicit characterization of the heterogeneity that exists among players. Let f^r and f^s be density functions.

A3. Heterogeneity: For all $c > 0$: $f^r(r|c) > 0$, $r \in [1/n, 1]$ and $f^s(s|c) > 0$, $s \in (0, 1/n]$.

Hence we assume that the full range of thresholds is represented in the player population.

2.2 A useful two-player game example

It will be useful to have an example motivation function to illustrate some key points as we go along. We emphasize that we will not use the example to prove any propositions.

[3] A2 runs counter to the hypothesis that people want to be first in payoff ranking (Duesenberry, 1949). By this hypothesis, we would always see dominant strategy play in prisoner's dilemma and public goods games, since this strategy is best from both a pecuniary and relative perspective. Many people, however, fail to play dominant strategy in these games (see sections 5 and 6). The equal split behavior in dictator games also contradicts the hypothesis.

[4] The reason we insist on the '*narrow*' qualifier is that we are not at all convinced that any of the behavior implied by ERC is altruistic. See remarks in section 8.

Consider the additively separable motivation function for player *i*, involved in a two-player game (we continue to write y_i as cs_i),

$$v_i(cs_i, s_i) \;=\; a_i cs_i - \frac{b_i}{2}\left(s_i - 1/2\right)^2;\; a_i \geq 0, b_i > 0 \qquad (2.2)$$

The component in front of the first minus sign is simply an expression of standard (risk neutral) preferences for the pecuniary payoff. The component after the first minus sign delineates the influence of the comparative effect. In essence, the further the allocation moves from player *i* receiving an equal share, the higher the loss from the comparative effect. Figure 3 displays a particular parameterization of (2.2).

Figure 3 here.

The functional form (2.2) allows us to express the range of heterogeneity posited by A3 in a very succinct form. A player's type is characterized by the marginal rate of substitution between pecuniary and relative argument, and is equal to the value $a/(b(s - ½))$. Strict relativism is represented by setting $a = 0$. Strict narrow self-interest is a limiting case ($b \to 0$).

2.3 ERC-equilibria

As players gain experience with game rules and the subject population, play tends to settle down to a stable pattern (see Roth and Erev, 1995). ERC makes equilibrium predictions intended to characterize the stable patterns. The basic framework is an incomplete information game in which each player's *r* and *s* are private information, but the densities f^r and f^s are common knowledge. That is, we assume that, in the stable state, players have learned the distribution from which their playing partners are drawn. But consistent with our assumption that playing partners are randomly and anonymously assigned, individual motivation functions are private information. Define an *ERC-Nash equilibrium* as a Bayesian Nash equilibrium solved with respect to player motivation functions. Define an *ERC-subgame perfect equilibrium* as a Bayesian subgame perfect equilibrium with respect to player motivation functions. (The games present no opportunities for updating strategic information, so except where noted, the ERC-subgame perfect equilibria we derive are sequential equilibria.)

ERC predictions about individual optimality that are independent of information considerations apply starting from the first round. The dictator game and second mover behavior in the gift exchange game are examples, as we show below.

ERC does not attempt to capture learning or framing effects (at least not this version). Section 7 compares what ERC explains to what present learning theories explain. ‘Framing effects’ refer to the influence on behavior that experimenters observe from changes in how the game is posed to subjects. When using ERC to make comparative predictions across games, we assume that the frame is held constant, in the sense that the directions given to subjects are parallel across games.[5]

3. Games of fairness: dictator, shrinking pie bargaining, best shot, and impunity

These games, when played in the lab, are always finite (a finite number of possible actions). For simplicity, we derive many of the results in this section assuming a continuous strategy space. Unless otherwise stated, all propositions characterize ERC-subgame perfect equilibria (recall the information conditions described in section 2.3).

3.1 Dictator and ultimatum games, and the relationship between them

First consider a dictator game in which the [D]ictator distributes a pie of maximum size $k > 0$ between self and a recipient. We represent the dictator’s division as the pair (c, s_D). So the dictator’s payoff is cs_D and the recipient payoff is $c - cs_D$.

Proposition 3.1: For all dictator allocations, $c = k$, and $s_D = r_D(c) \in$ [½, 1].

Proof: Follows directly from A1 and the definition of $r_i(c)$ given in section 2.1.

The dictator game has been the subject of several studies (e.g., Forsythe et al., 1994; Hoffman et al., 1995; Bolton et al., forthcoming). While the precise distribution of dictator giving varies with framing effects, proposition 3.1 appears equally valid for all studies: Dictators distribute all the money and (almost) always give themselves at least half. (Those taking less than half, like the one dictator in Figure 1, account for less than 1 percent of the data in the studies listed. Also see footnote 1.)

As an illustrative example, consider the additively separable motivation function given in (2.2), and suppose that $k = 1$. A straightforward calculation shows that

$$r_i = \min\left\{\frac{1}{2} + \frac{a_i}{b_i}, 1\right\}.$$

Hence the dictator's decision reflects the marginal rate of substitution between pecuniary and relative payoff.

Now consider an ultimatum game between a [P]roposer and a [R]esponder. For the moment, we assume the cake size, $k > 0$, is common knowledge. We represent the proposal by (c, s_P), interpreted analogously to the dictator notation. To keep the statements of the ultimatum game propositions as simple as possible, we assume that if a responder is indifferent between accepting and rejecting, that is, if $1 - s_P = s_R(c)$, then the responder always accepts proposal (c, s_P). We assume $s_i(c)$ is differentiable. Proposition 3.2 characterizes the responder's ERC-subgame perfect equilibrium strategy, and proposition 3.3 characterizes the proposer's.

<u>Proposition 3.2</u>: The probability a randomly selected responder will reject, $p(c, s_P)$, satisfies the following: (*i*) p has the value 0 when s_P = ½ and the value 1 when s_P = 1; (*ii*) p is strictly increasing in s_P over the interval (½, 1); (*iii*) fixing a $s_P \in$ (½, 1), p is strictly decreasing in c.

Proof: (*i*) By A1, for all responders, $v_R(c/2, 1/2) \geq v_R(0, 1/2)$. Hence, equal division is never rejected. The definition of $s_i(c)$ implies that the responder rejects the offer if 1 – $s_P < s_R(c)$, $s_i \in$ (0, 1/n]. Therefore, s_P = 1 offers are always rejected. (*ii*) This follows from integrating over the density $f^s(s|c)$. (*iii*) $s_i(c)$ is implicitly defined by $v_i(cs_i, s_i) = v_i(0, 1/2)$ for s_i ≤ ½. Differentiating yields $v_{i1}(cs_i, s_i)[s_i + cs_i'] + v_{i2}(cs_i, s_i)\, s_i' = 0$. Hence,

$$s_i'(c) = -\frac{s_i v_{i1}(cs_i, s_i)}{cv_{i1}(cs_i, s_i) + v_{i2}(cs_i, s_i)} < 0.$$

This completes the proof.

[5] The experiments of Forsythe et al. (1995) and Roth et al. (1991) are good examples of parallel presentation of different games.

Proposition 3.3: For all ultimatum proposals, $c = k$ and $s_P \geq ½$.

Proof: For any fixed $c > 0$, all proposers prefer $s_P = ½$ to any $s_P < ½$, and $s_P = ½$ is never turned down. It follows that any equilibrium proposal has $s_P \geq ½$. By proposition 3.2, $p(c, s_P)$ is strictly decreasing in c when $s_P > ½$ and constant for $s_P = ½$, so the proposer will propose dividing all of k.

Many studies, beginning with Güth et al. (1982), confirm propositions 3.2(*i*) and 3.3. The experiment of Bolton and Zwick (1995) vividly illustrates that lower offers tend to have a higher probability of rejection. Slonin and Roth (forthcoming) present evidence that the probability of rejection tends to decrease as c increases.[6]

Forsythe et al. (1994) found that, on average, offers are higher in the ultimatum game than in the dictator game. ERC predicts this relationship. By propositions 3.1 and 3.3, we may assume that all proposals divide all of k, which we normalize to size 1.

Proposition 3.4: On average, offers in the ultimatum game will be higher than offers in the dictator game. In fact, no one offers more in the dictator game, and the only players who offer the same amount are those for whom $r_i(1) = ½$.

Proof: That proposers who have $r_i(1) = ½$ offer the same in both games is obvious. Suppose instead that $r_i(1) = 1$. Since a demand of $(c, s_P) = (1,1)$ is always turned down in the ultimatum game, it is clear that the proposal will be $s_P < 1$. For all other proposers, $r_i \in (½, 1)$, we write out the first order conditions (normalize $v(0,1/2) = 0$):

FOC for the dictator game: $v_{D1}(s_D, s_D) + v_{D2}(s_D, s_D) = 0$

FOC for the ultimatum game: $v_{P1}(s_P, s_P) + v_{P2}(s_P, s_P) = \dfrac{p'(1, s_P)v_p(s_P, s_P)}{1 - p(1, s_P)} > 0$.

By inspection, $s_D > s_P$. This completes the proof.

[6] The additively separable motivation function of (2.2) implies a negative relationship between s_i and r_i; specifically, $s_i = r_i - \sqrt{r_i^2 - 1/4}$. As far as we know, there is no data on whether a relationship exists (let alone this one), although a relationship of some sort is plausible. An experiment clarifying this issue would help us towards a more precise version of the model.

3.2 Unknown pie size games

Suppose now that the responder must decide whether to accept or reject an offer of *y* monetary units without knowing the pie size, but knowing that the pie was drawn from some distribution, *f*(*k*), with support $\left[\underline{k},\bar{k}\right]$. Suppose $y < \underline{k}/2$. Mitzkewitz and Nagel (1993), Kagel et al. (1996), and Rapoport et al. (forthcoming) have all shown that responders are more likely to reject *y* under these circumstances than if they know for certain that the pie is $\underline{k}$, and less likely to reject than if they know it is $\bar{k}$. The same is true in ERC. Let $p_u(y)$ denote the probability that *y* will be rejected by a randomly selected responder. For simplicity, we assume that the size of the offer does not convey any information about the pie size (hence for this proposition, ERC-subgame perfection does not imply sequential equilibrium).

<u>*Proposition 3.5*:</u> For all $y < \underline{k}/2$, $p\left(\underline{k}, \frac{\underline{k}-y}{\underline{k}}\right) < p_u(y) < p\left(\bar{k}, \frac{\bar{k}-y}{\bar{k}}\right)$.

Proof: Suppose $y < \underline{k}/2$. Then there exists a responder *i* who, if he knew the pie size was $\bar{k}$, is just indifferent between *y* and rejecting. Then, keeping in mind postulate A2,

$$v_i(0,\frac{1}{2}) \quad = \quad v_i(y,\frac{y}{\bar{k}}) \quad < \quad \int_{\underline{k}}^{\bar{k}} v_i(y,\frac{y}{k})\,f(k)dk\,,$$

which indicates that *i* and players with similar rejection thresholds are less likely to reject when they do not know the size of the pie. A very similar argument shows that *i* is more likely to reject when he know the pie size is $\underline{k}$.

3.3 Two period alternating offer games

Each round of a two-period alternating offer game is played like an ultimatum game, with players switching roles from first to second periods. If the first period offer is rejected, the pie is discounted prior to the second period counterproposal. Bolton (1991) describes a 'comparative model' of two period alternating offer bargaining, and shows that the comparative statics fit the data well.

The comparative model is 'almost' a special case of ERC. The comparative model assumes that $v_i(cs_i, s_i)$ is strictly increasing in s_i for all *i*. The proof of proposition 3.4 implies that, for proposers with $r_P(c)$ > ½, v_P is an increasing function in some neighborhood around the proposer's demand, s_P. (This is also true for the two-period game.) So the comparative statics of the two models are, for these players, in local agreement.

ERC predicts that proposers with $r_P(c)$ = ½ will offer half the pie even if an offer of somewhat less is very unlikely to be turned down. Experimenters observe these people (e.g., Kagel, Kim and Moser, 1996). The comparative model makes no room for them, which is why it is only 'almost' a special case – ERC is a more accurate model, even for bargaining games.

3.4 Impunity and best shot

We concern ourselves with the "mini" versions of impunity and best shot games, and compare these to the mini-ultimatum game. In all three games, a proposer moves either 'left' or 'right'. The responder observes the proposer's move and then either 'accepts' or 'rejects.' The games differ only in the payoffs, which are listed in Table 1.

Table 1 here.

Note that the standard subgame perfect equilibrium is the same for all three games: the proposer plays 'right,' and the responder plays 'accept.' Applying ERC to the mini-ultimatum game is straightforward, and yields results qualitatively equivalent to those for the full version. Application of ERC to the other games leads to markedly different predictions.

<u>Proposition 3.6</u>: For the impunity game: (*i*) The only outcomes with a positive probability of occurring are (2,2) and (3,1). (*ii*) The proportion of (3,1) outcomes is equal to the proportion of the population for whom v_i (3, ¾) > v_i (2, ½). (*iii*) The probability of the (3,1) outcome is higher for impunity than for the mini-ultimatum game.

Proof: (*i*) For all responders, v_R (2, ½) $\geq$ v_R (0, ½) and v_R (1, ¼) > v_R (0,0). (*ii*) Given responders' behavior, the proposer's choice is effectively between (2,2) and (3,1). (*iii*) In ultimatum, all proposers who choose right prefer (3,1) to (2,2). But not all who choose left prefer (2,2) to (3,1). By (*ii*), impunity proposers choose right iff they prefer (3,1) to (2,2), and by (*i*), an offer of (3,1) is never rejected.

Experiments by Güth and Huck (1997) and Bolton and Zwick (1995) furnish evidence for 3.6(*i*) and 3.6(*iii*). Bolton, Katok and Zwick (forthcoming) provide evidence for 3.6(*ii*).

Proposition 3.7: The probability of the (3,1) outcome is greater in best shot than in the mini-ultimatum game. The proportion of (3,1) offers rejected in best shot is the same as in mini-ultimatum.

Proof: For the proposer, the expected value of playing 'right' is the same in both games. The expected value of playing 'left' in the best shot game is strictly smaller than in the ultimatum game: Let *p* be the probability a randomly chosen best shot responder prefers (1,3) to (1,1). Then

$$p\, v_P(1, \tfrac{1}{4}) + (1 - p)\, v_P(1, \tfrac{1}{2}) < v_P(2, \tfrac{1}{2}) \text{ for all } p \in (0,1].$$

For the second half of the proposition, note that, after an offer of (3,1), responders in mini-best shot and mini-ultimatum have identical choices available to them.

Proposition 3.7 implies that, relative to the ultimatum game, best shot behavior moves towards, but is not identical to, the standard subgame perfect equilibrium. Prasnikar and Roth's (1992) best shot experiment comes close to converging to subgame perfect equilibrium.[7] Duffy and Feltovich's (forthcoming) best shot experiment clearly does not converge, even after 40 iterations, although as predicted, best shot is closer to perfect equilibrium than a corresponding ultimatum game. In sum, the experimental evidence is consistent with proposition 3.7.

3.5 Three-way ultimatum and the solidarity game

We conclude this section with a discussion of experiments that bear on the question of whether motivation is adequately captured by motivation function preferences for distribution between self and the group, or whether they are better captured by altruistic preferences, where a person cares about the distribution across all individuals.

Güth and van Damme (forthcoming) report on a three-way ultimatum game experiment in which the proposer proposes a three-way split of the pie, and one responder can accept or reject. The third player, a recipient, does nothing save collect any payoff the other two agree to give him. The experiment finds that information about the recipient's share has no direct influence on the

[7] So does Harrison and Hirschleifer (1989), but the incomplete information aspect of the game renders the result incomparable to the theory.

responder’s decision to accept or reject. ERC predicts the same, because the distribution among the other players does not enter into the motivation function (see section 2.1).

Selten and Ockenfels (forthcoming) observe a similar phenomenon studying the solidarity game. In this game, each player in a three-person group independently rolls a die to determine whether they (individually) win a fixed monetary sum. Before the die is rolled, each announces how much she wishes to compensate the losers, for both the case where there is one loser, and for the case where there are two. Most subjects give the same total amount independent of the number of losers. In addition, gifts for one loser are positively correlated with the expectation about the gifts of others. Selten and Ockenfels demonstrate that neither the behavioral pattern nor the relation between decisions and expectations are easy to justify if subjects have standard altruistic preferences. They conclude that most subjects, even though they are willing to sacrifice money for solidarity, are uninterested in the welfare of recipients, and only care about their own share of the winnings.

Bolton et al. (forthcoming) find that the total gift dictators leave multiple recipients is stable, but how dictators distribute gifts across recipients appears, in most cases, to be arbitrary. Weimann (1994) analyzes a public goods experiment directed at the question of whether individual behavior of others, or just aggregate group behavior influences the decision to contribute. He concludes that, "Whether or not the individual contributions [to a public good] are common knowledge has no impact on subject's behavior" (p.192).

4. Competitive behavior

In the last section, we showed that if a game creates a trade-off between absolute and relative motivations, we can observe behavior which sharply contradicts standard theoretical predictions. But people do not always ‘play fair.’ Many market institutions apparently induce 'competitive,' self-interested behavior. In this section we show that typical market environments interact with ERC-motivations in a way that aligns absolute and relative motives. As a consequence, traditional Nash equilibria are ERC-Nash equilibria.

Some well known experimental results come from games with symmetric equilibrium payoffs, so we begin with the symmetric case. It turns out that ERC implies an interesting difference between Bertrand and Cournot games with respect to symmetry, and we turn to this issue at the end of the section.

Bertrand and Cournot games are the standard textbook examples of (oligopolistic) markets: Suppose demand is exogenously given by $M = p + q$, where M is a constant, p denotes the price and q the quantity. Suppose $n \geq 1$ identical firms produce at constant marginal cost q ($< M$). In Cournot games, firms choose quantities $q_i \in [0, M - q]$ yielding profits given by $y_i(q) = (M - q - q_{-i})q_i - q_i^2$, where $q_i \in [0, M - q]$. In Bertrand games, firms choose prices $p_i \geq q$ yielding profits equal to $y_i(p) = (p_i - q)(M - p_i)/\tilde{n}$ if i sets the lowest price along with $\tilde{n} - 1$ other firms, or equal to zero if there exists a firm $j \neq i$ which sets a lower price. All pure strategy spaces are finite. For simplicity, we assume that the interval between admissible price offers, D, is 'small;' specifically, $(p - D - q)(M - p - D) > (1/n)(p - q)(M - p)$ for all $p > q$, and for all n (so there is a pecuniary incentive to undercut p, when all others bid p).

The informational assumptions laid out in section 2.3 continue to apply.

Proposition 4.1: For $n \geq 1$, and for either price (Bertrand) or quantity (Cournot) competition, all Nash equilibria are ERC-Nash equilibria.

Proof: For $n = 1$, $s \equiv 1$ so that the ERC-monopolist simply maximizes his profits. For $n > 1$, observe that all Nash equilibria in both the price and the quantity game, yield equal equilibrium profits for all firms (see Binmore, 1992). Hence, a firm that deviates from his Nash equilibrium strategy can neither gain with respect to absolute nor relative payoffs. This completes the proof.

The remaining propositions in this section provide a stronger characterization of ERC-equilibria. We will suppose that for some $e > 0$ proportion of the population, r is approximately 1 for all possible total payoffs c, and for all number of players, n. (How close the approximation need be will be made explicit in the relevant propositions.) These people are highly narrowly self-interested, and they will drive some, but not all, of the market results. We make two technical assumptions: First, we suppose that $v_i(0,0)$ is, for all i, the worst possible outcome. Second, we suppose that the value of $v_i(y, 1/n)$ is bounded with respect to both i and n.[8]

[8] The first technical assumption simply implies that the worst thing that can happen to i is to have to watch others receive a positive payoff while receiving none himself. The second is also mild: Bounded with respect to i (fixing n) would follow immediately if we made the realistic, but less mathematically convenient, assumption that the population were finite; we simply impose boundedness on the infinite population (see A3). With respect to n, the assumption implies that for a fixed pecuniary payoff, the value to i of achieving the social reference proportion is

We first show that the competitive outcome is the unique ERC-Nash equilibrium for the Bertrand game. The intuition is quite simple: For large *n*, there is a high probability that at least one player cares sufficiently about his pecuniary payoff to undercut high bids in pursuit of pecuniary gain. In equilibrium, everyone knows that the probability of such a person is high, and so, in equilibrium, everyone undercuts because this is what is necessary to preserve relative as well as pecuniary positions.

Proposition 4.2: For price competition and for *n* large enough, the market price in all ERC-Nash equilibria is equal to cost q or to $q + D$, the standard Nash equilibrium prices for $n > 1$ firms.

Proof: Let g be the probability that the composition of players in the game is *sufficiently narrowly self-interested* in the sense that, for all admissible *p*,

$$v_i((p-\Delta-q)(M-p-\Delta),1) > v_i((1/n)(p-q)(M-p),1/n) \text{ for at least one } i.$$

Since the $r \equiv 1$ player satisfies this condition, it follows that, as n increases, g increases monotonically to 1. Choose *n* large enough, so that g satisfies

$$\max_i \left[(1-g)v_i((1/n)(p_M-q)(M-p_M),1/n)+gv_i(0,0)-v_i(0,1/n)\right] \quad < \quad 0,$$

where p_M is the monopoly price. A maximum exists because of the boundedness assumption.

Now suppose there is an ERC-Nash equilibrium in which the maximum bid that wins with positive probability is $p_H > q + D$. Since transactions are never made at a price of greater than p_H, bidding above p_H is strictly dominated by offering a price of p_H (recall that we assume that $v_i(0,0)$ is the worst possible outcome for all *i*). Therefore, in equilibrium, all prices bid with positive probability by any player must be p_H or lower. Hence p_H wins only if *all n* firms play it. It follows that the expected value to firm *i* of bidding p_H is

$$b\ v_i\left((1/n)(p_H-q)(M-p_H),1/n\right)+(1-b)v_i(0,0) \tag{4.1}$$

where b is the probability that all firms other than *i* bid p_H. On the other hand, the expected value of firm *i* bidding $p_H - D$ is

bounded with respect to the number of players in the game. We think assuming the value of $v_i(y,1/n)$ is fixed

$$b\ v_i\left((p_H-\Delta-q)(M-p_H-\Delta),1\right)+(1-b)[\ldots] \tag{4.2}$$

For sufficiently narrowly self-interested agents, (4.2) > (4.1). Therefore, sufficiently self-interested players always bid lower than p_H. Given this, the expected value of bidding p_H for *any* player is

$$\le (1-g)v_i((1/n)(p_H-q)(M-p_H),1/n)+g\,v_i(0,0)$$

$$\le (1-g)v_i((1/n)(p_M-q)(M-p_M),1/n)+g\,v_i(0,0)<v_i(0,1/n)$$

which contradicts the assumption that p_H is a best response for at least some player (any player can guarantee himself $v_i(0,1/n)$ by playing q). Since a construction like (4.2) is always possible if $p_H > q + D$, it follows that $p_H = q$ or $q + D$ for sufficiently large *n*.

In the guessing game, *n* > 1 players simultaneously choose a number *z* from an interval [0, *k*]. For simplicity, we assume that the number of choices is finite, and that the interval between any two consecutive choices is Δ. The winner is the player whose number is closest to $g\ \bar{z}$, $g < 1$. The winner receives a fixed prize; if there is a tie, winners share the prize equally. The guessing game is very similar to a Bertrand game, save that the cake to be distributed is fixed. Nagel's (1995) experiment shows that play converges to the unique standard Nash equilibrium, $z_i \equiv 0$.

<u>*Proposition 4.2a*</u>: For *n* large enough, the unique Nash equilibrium in the guessing game is equivalent to the (unique) ERC-Nash equilibrium.

Proof: Showing that $z_i \equiv 0$ is an ERC-Nash equilibrium is straightforward. For the proof in the other direction: Note that any outcome in which *i* wins has a payoff greater than $v_i(0,1)$. Fix a strategy profile for the other *n* – 1 players, and let $\bar{x}$ be the modal average implied by the distribution. If *n* is large enough, player *i*'s influence on the average is negligible (and so we can ignore it). So when *n* is large enough, by guessing $g\ \bar{x}$, player *i* can guarantee herself greater than $\frac{\Delta}{k}v_i(0,1)+\frac{k-\Delta}{k}v_i(0,0)$. Substitute this value for $v_i(0,1/n)$, and the rest of the proof closely parallels proposition 4.2.

with respect to *n* would be reasonable, but boundedness will be sufficient.

How large must n be? By proofs of propositions 4.2 and 4.2a, the answer depends on the prevalence of 'sufficiently narrowly self-interested' subjects in the population. Hoffman et al. (1994) performed a dictator game in a buyer-seller frame similar to Bertrand games (with players being randomly assigned buyer and seller positions). The proportion giving zero was about 45 percent.[9] Then the probability of at least one subject with $r = 1$ in a group of n subjects is $1 - 0.55^n$. Assuming that r is not too sensitive to the size of the pie, c, or to the number of players, n, a lower bound on the probability of at least one sufficiently self-interested player in a group of 3 is over 83 percent. It appears then that n need not be very large for ERC-Nash equilibrium market prices to shrink to the standard Nash price. Holt (1995) reports some evidence that outcomes of oligopoly games are less competitive with two players than with three or more, but no particular effect for numbers greater than two.

Interestingly, ERC implies that the auction market game studied by Roth et al. (1991) (discussed in section 1) is sufficiently different from the Bertrand game to obtain competitive results independent of the number of buyers. Recall that, in this game, buyers simultaneously bid on an object owned by a single seller. The lowest bid is submitted to the seller who either accepts or rejects; if the latter, all players receive a zero monetary payoff.

We prove that obtaining the (competitive) subgame perfect equilibrium does not depend on the number of buyers, so long as there are at least two. We normalize the surplus that can be shared from the transaction to 1, and we represent a *bid* by the proportion of the surplus that the buyer proposes *keeping* (defined this way, the relation to proposition 4.2 will be transparent). A bid *wins* if it is both the lowest submitted and large enough to be acceptable to the seller. Analogous to the Bertrand game, we suppose that the interval between permissible bids, D, is 'small.'

Proposition 4.2b: Consider an auction market game having at least two buyers. Under the assumption that the seller accepts, all ERC-subgame perfect equilibria for the market game have a winning buyer bid of 0 or D.

[9] We refer to Hoffman et al.'s buyer-seller dictator game with contest selection of roles. They also ran a buyer-seller dictator game with random selection of roles. The proportion giving zero was lower, but the proportion almost giving zero (10 percent or less) was about 40 percent, and in this sense our calculation is appropriate for both treatments. We refer to these particular dictator treatments because they are *roughly* framed (buyer-seller) in the same way as Bertrand experiments. We nevertheless think of the resulting calculations as illustrations. A

Proof: Suppose, contrary to proposition, that there is an equilibrium in which $z_H > \Delta$ is the highest bid that wins with positive probability. The proof that, in equilibrium, no one ever bids higher is analogous to the proof of proposition 4.2 if one substitutes "price (*p*)" for "bid (*z*)" and "firm" for "buyer". However, in contrast to the Bertrand game, in this market one buyer with the smallest bid is chosen randomly, and divides the surplus with the seller, who is an actual subject in the experiment. Consequently, equations (4.1) and (4.2) of proposition 4.2 become

$$b\ \left[\ (1/n)\ v_i(z_H, z_H) + (1 - 1/n)\ v_i(0,0)\ \right]\ +\ (1 - b)\ v_i(0,0) \qquad (4.1a)$$

$$b\ \ v_i(z_H - D,\ z_H - D)\ +\ (1 - \ b)\ [\ \ldots]. \qquad (4.2a)$$

The inequality (4.2a) > (4.1a) holds for *all* players, regardless of type. This contradicts the assumption that bidding z_H is a best response. This completes the proof.

About the assumption concerning seller behavior: From the point of view of ERC, its validity is an empirical question. In fact, Roth et al. report that no best bid was ever rejected in a non-practice round (p. 1075). The assumption is basically equivalent to positing that $v_i(s_i, s_i) > v_i(0,1/n)\ \forall s_i \in (1/n,1]$,[10] which implies an asymmetry with respect to fairness: 'I reject offers that are very unfair to me but accept offers that are very unfair to you.' Asymmetry of this sort is suggested by Loewenstein et al. (1989), and by Fehr and Schmidt (1997). While ERC has no problem accommodating this assumption, we have avoided it to highlight the fact that it is not relevant to any proof in this paper save proposition 4.2b – where it has but a very minor role. In particular, the assumption is not necessary to explain the competitive behavior of buyers in the Roth et al. game. Is there a restriction we could place on the motivation function to guarantee the competitive results in Propositions 4.2 and 4.2a for any sized group (greater than 1, of course)? The only one we can think of is a stronger asymmetry assumption: $v_i(c,1) > v_i(c/n,1/n)$ for all *i*, *c* and *n*. But this is falsified by dictator game experiments.

careful, meaningful calculation requires running dictator and Bertrand games in closely parallel frames (parallel directions).

[10] Strictly speaking, proposition 4.2b requires that the seller accepts *all* bids, not just those greater than 1/*n*. The proof, however, is easily extended: Suppose that the z_H in the proof gives the seller less than 1/*n*. Revise both (4.1a) and (4.2a) to reflect the fact that undercutting increases the probability the seller will accept.

Proposition 4.1 shows that the standard Cournot-Nash equilibrium is an ERC-Nash equilibrium. We do not know if this is the unique ERC-Nash equilibrium for the type of incomplete information game played in the lab (the type we have been studying). If we assume complete information, however, and restrict to pure strategies, we can prove uniqueness. Proposition 4.3 extends the classic textbook graph proof of duopoly Cournot equilibrium in pure strategies (e.g. Binmore, 1992, p. 290) to ERC motivations.

Proposition 4.3: Consider a Cournot duopoly in which both players know one another's motivation function. If $r_i(c)$ > 1/2 for all *c*, for at least one player *i*, then the unique ERC-Nash equilibria in pure strategies is the standard Nash equilibrium.

Proof:

Figures 4 (a) and 4 (b) here.

In figures 4 (a) and (b), the *x* axis shows the quantity of firm *j* and the *y* axis the quantity of firm *i*. The thick lines show the *standard* Nash-reaction curves of player *i* (BE) and player *j* (CF). Two things need to be proved. First, observe that for all quantity combinations lying on the diagonal AD, the marginal utility with respect to relative payoffs is zero, because payoffs are equal (assumption A2). Since the marginal utility with respect to absolute payoffs is strictly increasing for at least one player (note that $r_i(c)$ > 1/2 for all *c* implies $v_{i1} > 0$ for $s_i = 1/2$), the only location *on* AD which can be an ERC-Nash equilibrium is point X, the Cournot equilibrium. Second, note that: (1) on the Nash-reaction curves, $y_i'(q_i) = 0$ and $y_j'(q_j) = 0$, respectively; (2) $y_i'(q_i) > 0$ iff (q_i, q_j) is within ABE, $y_j'(q_j) > 0$ iff (q_i, q_j) is within ACF; (3) $y_i < y_j$ iff (q_i, q_j) is within ADE, $y_i > y_j$ iff (q_i, q_j) is within ACD; and (4) $s_k'(q_k) > 0$, *k = i, j*, everywhere in the interior of ACE. With these properties, it is easy to see that ERC-reaction curves are bounded by the Nash-reaction curves and the diagonal: *j*'s ERC-reaction curve must lie somewhere in the darkly shaded areas and *i*'s ERC reaction curve must lie somewhere in the brightly shaded areas. (The areas include the Nash-reaction curves for both players and exclude AD with the exception of point X for at least one player.) The only possible point of intersection of ERC reaction curves is X. This completes the proof.

The proof requires a sufficiently self-interested player in a weaker sense than do the Bertrand propositions, specifically $r_i(c)$ > 1/2 for one player. From dictator games, we estimate the proportion of $r(c) > 1/2$-players to be 80 percent. This is a conservative estimate – most dictator studies find a higher proportion than this. Then we estimate the probability of a standard Nash equilibrium (under complete information) to be at least 96 percent. The calculation ignores the 'pure strategy' requirement.

Evidence for the standard Cournot-Nash equilibrium is less than conclusive. Holt (1985) conducted single-period duopoly experiments of the type we study here. While in the beginning some subjects try to cooperate, quantity choices tend ultimately to Cournot level. Holt (1995) surveys a number of studies, and reports some support for Nash equilibrium, but also expresses reservations. Huck, Normann and Oechssler (1997) report rough convergence to Nash equilibrium in the four-person case.

Finally, ERC implies symmetric payoffs are important to Cournot outcomes in a way that they are not to Bertrand games. Consider a Cournot duopoly in which firm *i* has a cost advantage: $q_i < q_j$. The standard Nash equilibrium profit of firm *i* is greater than the profit of firm *j*. But this may not be an ERC-equilibrium because firm *i* may choose a smaller quantity in order to boost the relative payoff. On the other hand, consider cost heterogeneity in Bertrand games; i.e., each firm *i* is randomly assigned to costs $q_i \in \{q^1, q^2, \ldots, q^k\}, k < \infty$. Then, the competitive price is the lowest q in the market, and it is also a standard Nash equilibrium.[11] It continues to be an ERC equilibrium if the market is large enough; the proof is analogous to that of proposition 4.2.[12]

5. Dilemma games: a simple quantitative ERC model

All dilemma games share two defining characteristics. First, if players are purely narrowly self-interested, then their set of choices includes a dominant strategy that yields the highest payoff regardless of what others do. Second, deviation from the dominant strategy contributes to a higher joint payoff for the group, and enough contributions produce an outcome Pareto superior to the dominant strategy outcome. Dominant strategy is not a good description of the behavior

[11] This holds if there is more than one firm with minimum cost. If there is only one firm with minimum cost, there is a Nash equilibrium in which the price is the second lowest cost and the firm with minimum cost gets all the surplus.

[12] Roughly speaking, for *n* large enough there is one firm among the firms with minimum cost which is sufficiently self-interested so that it undercuts any price greater than minimum cost.

we typically see. In this section, we show that ERC is consistent with many of the patterns we do observe in the prisoner's dilemma (PD) and associated one-shot dilemma games (we discuss repeated dilemma games in section 6). We can say more with a fitted parametric model. In sections 5.2 and 5.3 we fit the gift exchange and investment games, both essentially (sequential) PDs.[13]

5.1 What's necessary to induce cooperation in simultaneous and sequential PD's?

We will demonstrate that, in ERC, the extent of cooperation can depend on the interaction between (*i*) heterogeneity with respect to how players trade-off pecuniary for relative gains, and (*ii*) the size of payoffs, especially the size of the efficiency gains that can be achieved through cooperation. These factors are important in both simultaneous and sequential PDs, although the factors interact in somewhat different ways across the two games.

Table 2 here.

Consider the PD payoff matrix in Table 2. To illustrate how trade-offs between pecuniary and relative payoffs matter to ERC predictions, we will suppose that individuals can be described by the motivation function given in (2.2), $v_i(cs_i, s_i) = a_i cs_i - \frac{b_i}{2}\left(s_i - \frac{1}{2}\right)^2$. Then the marginal rate of substitution between pecuniary and relative payoffs, *a*/*b*, fully characterizes a subject's type. The population distribution of types will be denoted by *F*(*a/b*).[14]

To see what influences cooperation in a one-shot *simultaneous PD*, examine the optimal decision rule for a subject with type *a*/*b*:

$$C \succ D \Leftrightarrow \quad \frac{a}{b} < \frac{p-1/2}{4(1-m)(1+2m)^2} =: g(m,p)\ .$$

Here, *p* is the probability that the opponent cooperates. Thus, cooperation is influenced by the extent to which subjects are motivated by relative payoffs, the magnitude of the mpcr, *m*, and the proportion of cooperating subjects in the population. There is always an equilibrium in which

[13] Not every player in either the gift exchange or the investment game has a dominant strategy, so technically speaking neither game is a dilemma game. But, as will become clear, they are both very close off-shoots.

nobody cooperates, but depending on the shape of *F*(*a/b*), there may also be equilibria in which a proportion of subjects cooperate, while others defect.

We next consider the *sequential PD*, in which the second mover decides after being informed of the first mover's action. We obtain an interesting result: Cooperation requires both subjects who are willing to sacrifice pecuniary for relative gains, *and* subjects who are mostly interested in absolute payoffs. To see this, examine the optimal decision rules (the information assumptions laid out in section 2.3 continue to apply):

second mover: $C \succ D \Leftrightarrow$ 1. first mover plays "C"

2. $\frac{a}{b} < g(m,1)$

first mover: $C \succ D \Leftrightarrow$ 1. $-1 + m(1+\hat{p}) > 0$

2.

$$\frac{a}{b} > \frac{1-\hat{p}}{8(m\hat{p}+m-1)(1+2m)^2}$$

Here, $\hat{p} = \hat{p}(m) = F(g(m,1))$ is the probability that the second mover responds cooperatively if the first mover cooperates. The second mover's optimal decision rule corresponds to the one applied in the simultaneous PD with $p = 0$ or 1 respectively. The second mover cooperates if and only if she is sufficiently motivated by the relative payoff, and the first mover cooperated. The first mover cooperates iff she is sufficiently motivated by *pecuniary* payoffs, and the expected monetary net return of cooperation (= $-1+m(1+\hat{p})$) is positive. The reasoning behind the required first mover motivation is simple: A first mover who is interested in relative payoff can guarantee equal payoffs by defecting, since in this case, the second mover defects for sure. Only if a first mover is sufficiently interested in his absolute payoff, will he take the chance of being exploited in an attempt to 'trigger' second mover cooperation.

[14] Of course, the results we derive will be special to this class of motivation functions. But keep in mind that our goal here is to demonstrate that particular factors *can* play an important role in what ERC predicts.

Heterogeneity guarantees that the proportion of both first and second movers who cooperate increases with the mpcr ($\hat{p}'(m) > 0$ and $\partial \dfrac{1-\hat{p}(m)}{8(m\hat{p}+m-1)(1+2m)^2} / \partial m < 0$). Even if $\hat{p}\,(m)$ is very small, a sufficiently high mpcr may induce the first mover to cooperate.[15]

Several studies support the view that potential efficiency gains and the propensity of others to cooperate (measured in ERC by the marginal rate of substitution between absolute and relative payoffs) are major determinants of cooperation in both simultaneous and sequential PDs. In a well-known survey, Rapoport and Chammah (1965) demonstrate that cooperation rates in PDs increase when the gains from cooperation increase, or when the 'sucker' payoff decreases.[16] Ledyard (1995) surveys the literature on public good games, and concludes that, besides communication, the mpcr is the only control variable that has a strong positive effect on cooperation rates. Many experiments show a strong relation between own and opponent decisions. Cooper, DeJong, Forsythe, and Ross (1996) found two behavioral types in one-shot PDs, which are perfectly in line with the ERC-decision rules in PDs derived above: "egoists", who always defect, and "best response altruists", for whom *C* (*D*) is a best response to *C* (*D*).[17] Similarly, Pruitt (1970) and Rapoport and Chammah (1965, p. 56-66) found strong positive interactions between cooperative choices of players. Several studies have manipulated the expectation about the cooperation behavior of the opponent and found a positive correlation of own defective choices and the probability that the opponent defects (ex., Bixenstine and Wilson, 1963, and Lave, 1965). Bolton, Brandts and Katok (1996) and Fehr, Gaechter and Kirchsteiger (1997) provide demonstrations that cooperation is sensitive to other player strategy choice in sequential dilemma games. While some (not all) of these studies involve repeated play, ERC implies that the particular behavior is not due to repetition.

5.2 A parametric analysis of the gift exchange game: the a-model

In a well known paper, Fehr, Kirchsteiger and Riedl (1993) investigated wage and effort decisions in an experimental labor market. In the first stage of this gift exchange game, a firm

[15] The specific class of motivation functions also implies an income effect. Suppose we multiply all payoffs in Table 2 by a fixed positive number. Then there is a stronger tendency for all players to behave according to standard game theoretic predictions. Rabin's (1993) model makes a similar prediction.

[16] Rapoport and Chammah (1965), p. 39, Figure 1. Lave (1965) includes similar results.

[17] The hypothesis that altruistic subjects cooperate unconditionally ("dominant strategy altruism") is clearly rejected in their study.

offers a wage *w*; and in the second stage, a worker who accepts chooses an effort level *e*. Since efforts are costly and the game is one-shot, the standard subgame perfect equilibrium has workers providing the minimal effort possible regardless of the wage, and the firm should therefore provide the minimal wage. This is not what Fehr et al. observed (see Figure 2).

What can ERC say about this game? First, since gift exchange is essentially the sequential dilemma game analyzed in section 5.1, the qualitative type of cooperative outcome Fehr et al. did observe – an above minimal wage followed by an above minimal effort level – can be sustained in ERC-equilibrium. Somewhat more substantively, ERC's most basic prediction is that all workers will try to give themselves at least half the pie (proposition 3.1). Workers in three cases had no option that gave them half or more. Consistent with ERC, all three chose the minimum effort. In 96 percent of the other 273 cases, the worker gave himself at least the same payoff as the firm. In four of the 11 anomalous cases, the worker chose to keep 22 and gave 22.8. If they had chosen the next smallest effort level, the payoff distribution would have been (15.2, 23). Clearly, these 4 equated payoffs up to rounding. In sum, 97.5 percent of worker responses are in, or nearly in, the range predicted. Hence, the very basic facts of the game are in line with the ERC model.

But we would like to say more about this experiment. To do so, we need a parametric model. We use the Fehr et al. data to construct a very simple, parameterized ERC. Quantitatively fitting firms comes down to the rather shallow claim that we can find a set of expectations and risk postures to justify their actions. We therefore confine ourselves to fitting a model of optimal worker responses. (One of the things we will find is that observed firm behavior is quite sensible given worker behavior.) Fehr et al. report that they found no learning effect among workers – evidence that motivation functions are in fact stable. We therefore fit the model to all 276 wage-effort pairs collected over the 12 rounds of play.

For reasons of tractability, we fit the simplest possible model – one that uses a single parameter, a, to express the shape of worker heterogeneity. Specifically, we represent the range with the end points: Suppose then there are only two types of workers, a proportion α of [R]elativists and a proportion $(1-\alpha)$ of [E]goists. The goal of the egoist is to maximize pecuniary payoff. The goal of the relativist is to "mitigate" payoffs; that is, the relativist minimizes $\left| u(e) - p(w) \right|$, where $u(e)$ and $p(w)$ are respectively worker and firm payoffs.

Both the data and the general ERC model (see A3) imply that many people are somewhere in between these two categories. Think of a as a summary measure of the propensity to reciprocate found in the heterogenous population. We will show that the value of α obtained from the gift exchange game closely approximates values obtained from the investment game experiment of Berg, Dickhaut and McCabe's (1995), and from several dictator game experiments. Hence, even though the a-model does not attempt to characterize the precise behavior of most individuals, its estimate of the propensity to reciprocate is very robust. Just as importantly, this very sparse model explains the Fehr et al. experiment in substantial detail.

For the Fehr et al. experiment, payoffs for the firm and the worker were respectively $\pi(w) = (v - w)e$ and $u(e) = w - c(e) - c_0$, with v = 126, c_0 = 26, $e \in \{0.1, 0.2, \ldots, 1\}$, and $w \in \{30, 35, 40, \ldots, 125\}$. To keep the exposition simple, we assume continuous strategy spaces $e \in [0.1, 1]$ and $w \in [c_0, v]$, and a continuous cost function $c(e)$ with $c(0.1)$ = 0, $c(1)$ = 18 (see Appendix for Fehr et al.'s whole cost function), $c'(e) \geq 0, c''(e) \geq 0$. The data analysis, however, accounts for the discontinuities in the experiment's strategy spaces. The Fehr et al. design involved an excess supply curve, but Charness (1996) replicated the experiment without one, and so we will not consider supply conditions here.[18]

Define $\underline{w}$ and $\overline{w}$ by

$$w \leq \underline{w} \Leftrightarrow e = e^{\min} = 0.1 \text{ minimizes } \left| u(e) - p(w) \right|$$

$$w \geq \overline{w} \Leftrightarrow e = e^{\max} = 1 \text{ minimizes } \left| u(e) - p(w) \right|$$

Then the best response-functions for the workers are

$$e^{E}(w) \equiv e^{\min} = 0.1 \quad \text{for egoists;}$$

$$e^{R}(w) = \begin{cases} 0.1, & w \leq \underline{w} \\ e^{*}(w), & \underline{w} < w < \overline{w} \\ 1, & w \geq \overline{w} \end{cases} \quad \text{for relativists.}$$

Here, $e^{*}(w)$ is implicitly defined by equating $u(e)$ and $\pi(w)$: $(v - w)e^{*}(w) = w - c\left(e^{*}(w)\right) - c_0$.

[18] Charness set v = 120 and c_0 = 20 and $w \in [c_0, v]$ with all other variables the same. The results of our analysis are valid for both designs.

We state three hypotheses concerning efforts and payoffs, and provide a rough sketch of the proofs. The formal derivations – all straightforward calculations – are in the Appendix. We compare each hypothesis with the experimental data collected by Fehr et al.

Proposition 5.1 (effort hypothesis): A higher wage induces a higher average effort level; specifically, $\bar{e}'(w) := \partial\left[(1-a)e^E + a\, e^R(w)\right]/\partial w \geq 0$ with strict inequality for $w \in (\underline{w}, \overline{w})$.

Sketch of proof: For $w \in (\underline{w}, \overline{w})$, an increase in the wage leads to an increase in workers' payoff, which relativists mitigate through higher effort. Since egoists' effort levels are constant, average effort levels increase. For $w \notin (\underline{w}, \overline{w})$, the model predicts constant average effort levels for both egoists and relativists (see the best response-functions and Appendix).

The effort hypothesis is clearly confirmed by the data. Fehr et al. report strongly significant correlation measures for highly aggregated data (p. 447-448). On a somewhat less aggregated level, the Spearman rank correlation coefficient between wages and average effort levels calculated over all 17 values of wages actually chosen ($w \in [30{,}110]$) shows a clear correlation ($\rho(\bar{e}, w)$ = 0.965, two-tail *p*-value < 0.00012). The α-model predicts that the wage-effort correlation is less prominent on the individual level since the egoists do not respond at all to different wage offers. The Spearman rank correlation coefficient between efforts and wages on the disaggregated data is $\rho(e, w)$ = 0.495, a lower value than what is observed on the aggregate level, but nevertheless one that is very significant (two-tail $p < 10^{-14}$).

Proposition 5.2 (worker payoff hypothesis): Higher wages increase the worker payoff; $u'(w) := \partial u(e(w))/\partial w > 0$.

Sketch of proof: This is obviously true for egoists since $u^E(w) = w - c_0$. The payoff for relativist workers is

$$u^R(e^R(w)) = w - c(e^R(w)) - c_o = \begin{cases} w - c_0, & w \leq \underline{w} \\ w - c\left(e^*(w)\right) - c_0, & \underline{w} < w < \overline{w} \\ w - 18 - c_0, & w \geq \overline{w} \end{cases}$$

u^R is increasing for small and large wages ($w \notin (\underline{w}, \overline{w})$). In the middle interval, a relativist's payoff is increasing because increases in wage and effort lead to efficiency gains, as measured by total payoffs (see the Appendix).

The Spearman coefficient between wages and the worker payoff using individual data is $\rho(u,w)$ = 0.94 (two-tail $p < 10^{-52}$), consistent with the worker payoff hypothesis.

Proposition 5.3 (firm's payoff hypothesis): The average profit, $\overline{\pi}(w)$, is decreasing in $\left[c_0, \underline{w}\right]$, then, for a not too small (> 12%), increasing up to a maximum $w^* \le \overline{w}$, and finally decreasing for $w > w^*$.

Sketch of proof: The average payoff to a firm within the α-model is given by

$$\overline{\pi}(w) = (v-w)\left[(1-\alpha)e^E + \alpha\ e^R(w)\right] = \begin{cases} (v-w)0.1, \quad w \le \underline{w} \\ (v-w)\left((1-\alpha)\,0.1 + \alpha\, e^*(w)\right), \ \underline{w} < w < \overline{w} \\ (v-w)\left((1-\alpha)0.1 + \alpha\right), \quad w \ge \overline{w} \end{cases}$$

Since effort levels are constant for very small and very high wages, our model predicts a negative relationship between $\overline{\pi}(w)$ and w for $w \notin (\underline{w}, \overline{w})$. For $w \in \left(\underline{w}, \overline{w}\right)$, $\overline{p}(w)$ is strictly concave, because marginal total payoffs are decreasing in *w* (see Appendix). Relativists are willing to share total payoffs equally so that the marginal expected profit $\overline{p}'(w)$ is decreasing.

Of course, the exact shape of $\overline{\pi}(w)$, and whether it pays for firms to deviate from the minimum wage, depends on a. Let $\underline{a}$ be implicitly defined by $\overline{p}'(\underline{w};\underline{a}) = 0$. Since $\overline{p}(w)$ is strictly concave for $w \in$ $(\underline{w}, \overline{w})$, the profit of a firm is decreasing for all *w*, iff $a \le \underline{a}$. However, if a > $\underline{a}$ ≈ 12%,[19] which is very reasonable in view of other experimental results (see section 5.3), $\overline{\pi}(w)$ is increasing for $w > \underline{w}$ up to a maximum w^* $\le \overline{w}$ and decreasing beyond w^* (see Appendix).

In order to compare the firm's payoff hypothesis to the data, we need an estimate of α. We obtain an estimate in the most straightforward manner possible. We calculate the average

[19] 12% is the value calculated with the discrete strategy spaces and cost function used in the experiment. With the continuous strategy spaces and cost function, the corresponding value is 10% (see Appendix).

effort level for each wage level actually offered, and then calculate $a(w)$ by solving $\bar{e}(w) = (1-a)0.1 + ae^{R}(w)$. Then $a = \sum_{w>\underline{w}} \frac{\#w}{273} a(w)$ (for $w \le \underline{w}$, all subjects chose minimum effort – as predicted). Calculating a in this way yields (exactly) a = 0.5.[20]

Figures 5(a) to 5(c) demonstrate that the estimated model fits the data very closely. Note that the a-model (and ERC generally) predicts no variance in the effort levels and payoffs for $w \le 35 = \underline{w}$, and in fact all three observations in this range are at the minimum effort level. There is only one observation for w = 110, and the corresponding actual effort and actual payoffs are within the range permitted by the model. Finally, Figure 5(c) shows how actual wage offers cluster around the optimal wage offer. It appears that firms learned to accurately anticipate effort response during the course of the experiment.

Figure 5 here.

Finally, the ‘fair wage-effort hypothesis’ that Fehr et al. studied posits a correlation between wages and effort. This is, as we have indicated, confirmed in the data. But Figure 5(c) shows that higher wages are not always met by higher *profits*. If we think that higher than minimal effort indicates *reciprocal* behavior, in the dictionary sense, we might have expected a strictly positive relation; that is, we might have expected higher worker payoffs to always be rewarded by higher profits. In fact, there is no correlation that is both clearly significant *and* positive, no matter how we calculate it: $r(u,p) = –0.16$ (two-tailed p = 0.0065), $\rho(\bar{\pi}, w) \approx \rho(\bar{u}, \bar{\pi}) = 0.304$ (two-tailed p > 0.22), $r(p,w) = .109$ (two-tailed p > 0.07), and $r(\bar{p}, w) = .316$ (two-tailed p > 0.20). We cite additional evidence of the same flavor in the next subsection, and return to comment in section 7.

5.3 Checking the robustness of the a-model

One quick way to check the robustness of our estimate of a is to determine whether it is consistent with the values obtained from dictator game experiments. While the rates reported

[20] The described estimation technique is somewhat crude, but has the advantage of being transparent. A somewhat more sophisticated method is minimizing the weighted deviations from actual and predicted payoffs:

$$\sum_{w} \frac{\#(w)}{276} \left(\left| \bar{p}^{actual}(w) - \bar{p}^{a-\text{model}}(w,a) \right| + \left| \bar{u}^{actual}(w) - \bar{u}^{a-\text{model}}(w,a) \right| \right).$$

Doing so, we obtain the value α = 0.46, very close to the value from the simpler estimation method.

vary somewhat due to framing effects and other design differences, the Forsythe et al. (1994) experiment, discussed in section 1, has an average rate of giving of 0.23, one of the Hoffman et al.'s (1994) dictator games (buyer-seller frame, random selection of roles) has 0.27, and a recent study by Andreoni and Miller (1996) obtained a value of 0.25. Within the context of a dictator game, a strict egoist gives 0, and a strict relativist gives half. Hence all of the average rates of return mentioned imply $\alpha \approx 0.5$, very similar to the estimate from the Fehr et al. data.

Another way to check the robustness of our result is to see if we obtain a similar estimate from a different dilemma experiment. Berg, Dickhaut and McCabe (1995) report on an investment game in which an investor and the responder each begin with an endowment of $10. The investor may then send some of his endowment to the responder. Whatever is sent, immediately triples in value. The responder then decides how much, if any, of the money to return to the investor. We denote the investment by x and the return by z. Keep in mind that *both* players start with a $10 endowment. From the general ERC model, we would expect $z \leq 2x$. In fact, the inequality holds for 30 out of 32 cases (94 percent).

In estimating a, we confine ourselves to the 32 independent observations in the "no history" treatment (history complicates the analysis). We compute a for each game in which the investment was greater than 0, precisely in the same manner as for the Fehr et al. data. The resulting value of a is 0.42, very close to the value we obtained in the gift exchange game.[21]

Looking for evidence of reciprocity, Berg et al. hypothesized that, on average, z/x and x are positively correlated (p. 127). On the other hand, since $z = a\,2x \Rightarrow z/x = 2a$, the a-model implies that z/x is constant for all x. The data confirms the a-model: the Spearman rank correlation coefficient is 0.01 (p. 131).

The a-model is the simplest model rich enough to capture the interplay between heterogeneity and efficiency gains necessary to achieve cooperation. In fact, the predicted interplay is different across gift exchange and investment games, and we can use this to explain why counting on reciprocity paid in one game, but not in the other. In the gift-exchange game, marginal efficiency gains are extremely high for small wages so that self-interested firms should

[21] The more sophisticated estimation technique mentioned in the last footnote is, for the investment game, equivalent to the simpler technique.

cooperate even if a is much lower than the 0.5 we calculated.[22] Marginal efficiency gains eventually decrease, and so increasing the wage is less attractive for a firm that already pays a high wage. Nevertheless, in an expected value sense, it pays to offer a higher than minimal wage (see Figure 5 (c)). In the investment game, investments of any size are multiplied by a fixed factor of 3. So efficiency gains are fixed – and at a rather low level – the factor 3 has to be matched with an a of at least 0.5 just for an investment to break even. Given our estimate of a, we are not surprised that investments in the investment game failed by just a bit to generate a positive net return: the average net return was – $0.50.

6. Some observations on the finitely repeated prisoner's dilemma

Defection in all rounds is the unique standard subgame perfect equilibrium for the finitely repeated prisoner's dilemma (PD). Subjects in experiments, however, systematically cooperate, although they typically fail to reach full efficiency. In a famous paper, Kreps et al. (1982) present two models of the finitely repeated PD. One model demonstrates that if each player assesses a (small) positive probability that his partner is 'cooperative' (i.e., he prefers to cooperate (defect) if the other cooperates (defects)), then sequential equilibria exist wherein purely money-motivated and perfectly rational players cooperate until the last few stages.

Note that, by the ERC heterogeneity assumption, cooperative subjects exist in reality, not just in people's minds.[23] This is not to say that the two models are the same. They differ on two important points. First, ERC predicts that cooperation rates may be positive even in the last round of a repeated PD (consider two players who are mostly interested in the relative payoff and believe with a high degree of certainty that there partner is too), and even among experienced players (experience teaches that some people *are* willing to cooperate until somebody defects on them). Second, in ERC, the proportion of cooperative subjects is not exogenous, but depends on the stage game payoff matrix (see section 5.1). We discuss evidence in favor of ERC taken from different sets of studies:

[22] If one of the players sacrifices one payoff unit in the subgame perfect equilibrium, total payoffs are increased by about ten payoff units. This is a much higher efficiency gain than in most experimental dilemma games.

[23] For a subject with r = ½, $u(\frac{c}{2},\frac{1}{2}) > u(\frac{1+m}{4m}c,\frac{1+m}{4m})$. Since $\frac{1+m}{1+2m} > \frac{1+m}{4m} > \frac{1}{2}$, $u(\frac{c}{2},\frac{1}{2}) > u(\frac{1+m}{4m}c,\frac{1+m}{1+2m})$. Hence, a subject who prefers the equal split in a dictator game also prefers to cooperate if the PD partner cooperates. The same conclusion holds for a '½ - *e*' -type for sufficiently small *e*. This, together with the heterogeneity assumption (A3) and the fact that all subjects prefer to defect if the opponent defects yields the result.

Andreoni and Miller (1993) conducted a series of experiments to test the performance of the sequential equilibrium prediction of Kreps et al. They ran several experimental conditions including *partners* (each subject partners with another subject for a 10-period game, repeated 20 times, each time rotating partners) and *strangers* (each subject plays 200 iterations of the PD with a new partner every iteration). Andreoni and Miller conclude (p. 582):

> “Subjects in a finitely repeated prisoner’s dilemma were significantly more cooperative than subjects in a repeated single-shot game. Moreover, by increasing subjects’ beliefs about the probability that their opponent is altruistic, we can further increase reputation building. Several findings in the experiment suggest that, rather than simply believing that some subjects may be altruistic, many subjects actually are altruistic.”

Among the findings which make the authors think there is a stable number of cooperative subjects is that the mean round of first defection in the partners-treatment is *increasing* across the 10-period games, whereas strangers quickly develop a stable pattern of cooperation. Under the assumption that subjects update their belief about the proportion of cooperative subjects, this clearly contradicts the rationality hypothesis, but is consistent with ERC.

Likewise, Cooper et al. (1996) conclude that the reputation model fails to explain positive cooperation rates observed in their one-shot PDs, whereas altruism alone (without reputation building) cannot explain the significantly higher cooperation rates and the path of play in their finitely repeated PDs. Apparently isolated models of reputation or altruism fall short of explaining typical behavior patterns. ERC, however, suggests that it is the *interplay* of strategic triggering behavior of egoists and altruistic responses of cooperative subjects which drives the results in repeated and sequential (cf. section 5) dilemma games. Additional evidence comes from two other studies dealing with repeated dilemma games:

Camerer and Weigelt (1988) conducted an experimental test of a one-sided reputation model in a supergame. Subjects played 8 periods of a stage game with the same partner. Table 3 shows the payoff matrix for each stage. Player 1 chooses first and player 2 chooses knowing the first mover’s choice. Payoffs when the first mover cooperates and the second mover defects varied across sessions as indicated in Table 3. The 'proportions' column indicates that the authors induced 2/3 of the second movers to prefer to defect and 1/3 to prefer to cooperate, at least when one applies the standard analysis to the stage game. First movers were not told the type of their partner. From their data, Camerer and Weigelt conclude that the sequential equilibrium is a good approximation for aggregate behavior, save that subjects cooperate longer and more often than predicted. They explain the discrepancy with evidence indicating that the actual proportion of cooperative subjects exceeded the proportion induced.[24]

Table 3 here.

Cell 1 of a follow-up study by Neral and Ochs (1992) replicated Camerer and Weigelt's result (see Table 3). Cell 2 modified second mover payoffs, but left the induced probability of cooperative subjects constant. Sequential equilibrium predicts that (*SE1*) a cell 2 first mover is on average less willing to cooperate in each round once mixed strategy play begins (mixed strategy should begin in the same round in both cells); and (*SE2*) there is no systematic influence on the second mover. But Neral and Ochs found a systematic influence on both movers.

To analyze the game using ERC, we need a method for calculating relative payoffs for outcomes with negative pecuniary payoffs. The most straightforward way to do so is to normalize each outcome involving a negative payoff by adding the absolute value of the smallest payoff at the outcome. For the games represented in Table 3, this means adding 100 to each payoff in an outcome with a –100 payoff. (We emphasize that this normalization is solely for the purpose of calculating relative payoffs. Absolute payoffs are as originally stated.) So for outcomes where one player gets a negative pecuniary payoff and the other gets a positive pecuniary payoff, the relative payoffs are normalized to 0 and 1 respectively, which is consistent with the types of outcomes we calculate with solely positive pecuniary payoffs.[25]

ERC predicts that for both movers, mixed strategy play begins later in cell 2 compared to cell 1, but once mixed strategy play begins, the probability shifts are as predicted by *SE1* and *SE2*. The reason for the delay in defecting is that ERC anticipates a greater proportion of cooperative subjects in cell 2 than in cell 1. To see this, first observe that the relative payoffs for each outcome are unaffected by the modification of the payoff structure. Since there is no change in the absolute payoffs of the first mover, his preference over the outcomes remain constant. On the other hand, the absolute payoff of the second mover is smaller when he defects, while the value from cooperation remains constant. Therefore, ERC predicts that more second movers choose to

[24] A further test, in which no cooperative players were induced, provided evidence for the sequential equilibrium hypothesis under the assumption that first movers have "homemade" priors regarding the proportion of cooperative subjects.

[25] For example, pecuniary outcomes of (0, 100) and (0, 150) both produce the relative weights (0, 1).

cooperate in cell 2 than in cell 1. This implies that mixed strategy play begins later for both the first and the second mover than predicted by SE1 and SE2.

The following figures show the observed frequency of defection in cell 1 and 2 of the Neral and Ochs experiment for first movers (figure 6(a)) and for second movers (figure 6(b)), respectively. A cross symbolizes a significant difference on the 5% level.[26]

Figures 6(a) and 6(b) here.

The results support the ERC hypothesis. First, as predicted by ERC both first and second movers start to defect significantly later (round 3 in cell 1 and round 4 in cell 2). Furthermore, as predicted by both models, there is no significant difference with respect to the second mover behavior once mixed strategy is played *in both cells* (from round 4 on). For the first mover, both ERC and sequential equilibrium predict a 'jump' from full cooperation to a constant fixed probability until the last round. Also, both theories predict that this fixed probability is smaller in cell 2. The data shows that, although the first movers do not jump in the lower cooperation rate in cell 2 there is a clear downward trend and finally, in round 6, first movers in cell 2 cooperate significantly less. Jung et al. (1994) obtains a similar result.

We do not want to overstate the case: There are some clear limits to what a static equilibrium model like ERC can explain. For example, Camerer and Weigelt observe that some subjects seem to apply simple cutoff strategies, and Selten and Stoecker (1986) identify some simple adaptive learning rules. That said, ERC captures some important behavioral regularities that are not captured by the rationality hypothesis alone.[27]

7. Other theories

Here we provide a brief comparison of ERC with two other approaches in the literature.

[26] The data and statistics are from Neral and Ochs, table V, p. 1163. Values are averages for experienced subjects.

[27] Estimates of those willing to cooperate are remarkably stable across investigations (particularly given differences in frames and payoffs). We explained above that all dictators who share the cake equally are willing to cooperate in a prisoner's dilemma. In section 4, we mentioned evidence that the proportion of 'strictly relativistic' dictators is about 20%. Andreoni and Miller reported corresponding values in dilemma games (1993, p. 581): "... the behaviour in the stranger condition is consistent with an imperfect-information equilibrium in which individuals share a common prior on the probability of experiencing cooperation, p^*, of about 0.20. Two previous studies have also estimated subjects' priors on cooperation. Camerer and Weigelt (1988) estimated 'homemade priors' of 0.17 that an opponent would play cooperatively, and McKelvey and Palfrey (1992) estimated the proportion of altruists to be 0.05 and 0.10." Furthermore, Cooper et al. (1996) estimated the proportion of cooperative subjects ("best-response altruists") in their one-shot PDs to be 12.5-15 percent.

Reinforcement learning theory, describes behavior as a learning-through-adaptation process (e.g., Roth and Erev, 1995; also see Gale, Binmore and Samuelson, 1995). Differences in behavior across games are attributed to the differential reinforcement delivered by differing payoff structures. Roth and Erev have shown that simulations based on this sort of learning generate paths like those observed for the ultimatum, best shot, and auction market games.

Learning theory and ERC are complimentary along certain lines. For example, learning theory makes predictions about the dynamic path of play, taking the initial conditions as given. ERC predicts the stable outcome at the end of the learning path, and also characterizes some initial conditions (e.g., second movers will have a propensity to reject in the ultimatum game but not in the impunity game; see Abbink et al., 1997). ERC can explain dilemma games, whereas adaptive learning does not easily explain the failure of dominant strategy. A recent paper by Erev and Roth (1997) extends learning to constant sum games, which ERC has not yet tackled. When the theories make contrary predictions, it usually has to do with whether there is any learning to be observed. For example, ERC characterizes second mover behavior in the ultimatum and best shot games as stable, whereas learning implies change (we mentioned evidence on this question in section 2).

Rabin (1993) exemplifies an equilibrium approach based on the idea that people help those who help them, and hurt those who hurt them. Note the emphasis on *intentionality*; that is, a player is conjectured to care whether another player's actions were intended to help or hurt. The model is limited to two-person normal form games, but it successfully accounts for behavior in games such as the simultaneous prisoner’s dilemma. Levine (1997) presents a related theory for extensive form games. An inherent limitation of this approach is that it cannot explain the dictator game. The recipient does not have a chance to act either kindly or unkindly to the dictator, so there is nothing to reciprocate. This limitation is important for two reasons. First, the dictator game is arguably the simplest possible dominant strategy game (it is not even a game – it is a one person decision problem). It seems to us that a satisfactory explanation for why people violate dominant strategy should explain the simplest violation.

Second, the fact that dictators *do* give suggests that a substantial portion of the other-regarding behavior we see in these games is based in something other than intentions. In fact, the dictator game is not the only evidence for this. Charness (1996) ran a gift exchange experiment with three treatments. The first essentially replicated the experiment of Fehr et al. (1993). In the second treatment, wages were determined by an unpaid third party. In the third, wages were randomly drawn from a bingo cage. There was no difference between outside party and random treatments, and only a mild difference between these and the standard game. Perhaps most importantly, strong evidence for what is usually thought to be the tell-tale sign of reciprocity in gift exchange games, positive correlation between offers and second mover actions, is found in all three treatments.[28] This result is consistent with ERC but not with intentionality models.[29]

8. Summary

ERC demonstrates that much of what we want to understand about behavior over a wide class of strategic situations can be deduced from two of the most elementary games: ultimatum and dictator. Taken together, these games expose the thresholds – the flash points – at which the pull of narrow self-interest is subjugated to concern for relative standing. These flash points, when combined with the structure of a specific game, determine the strategic opportunities open to players. The success of the standard equilibrium concepts employed by ERC implies that players do indeed behave in a bona fide strategic manner.

But what is this concern for relative standing? Is it altruism, equity, or reciprocity? Regardless of the label we choose, there is a second, deeper question: *Why* should people care about relative standing? We speculate that the answer to the first question is ‘reciprocity’ – of a non-standard type – and that the answer to the second question has to do with biology. As we explained in section 3.5, several experimental studies cast doubt on the proposition that people care about distribution in a way that we would expect an altruist to care. The same evidence suggests that people are willing to sacrifice little to defend equity as a principle. People appear self-centered, albeit in a way that differs from received theory.

While the dictionary definition of reciprocity will not work (see section 1), we think there is a sense in which ‘reciprocity’ can be defended. People have always lived in groups, and so we

[28] The correlation is not only positive but also very similar in all three treatments. The range of the (highly significant) Spearman rank correlation coefficient between wages and effort is 0.404 (random) to 0.491 (standard game), and between wages and *average* effort is 0.905 (random) to 1 (third party).

[29] Blount (1995) found no difference in minimum acceptable offers in ultimatum games whether the proposals comes from the first mover, or an unpaid third party. The minimum acceptable offer dropped substantially when proposals were drawn randomly, but responders still rejected an average of 12 percent. So not all of the motive for rejection can be attributed to intentions, and section 3 shows that much of what we observe in ultimatum games can be explained without considering intentions. Bolton, Brandts and Katok (1996) and Bolton, Brandts and

expect that evolution has molded them for successful group living. People may then have a propensity to contribute to the group, because a successful group contributes to their own individual biological success. A propensity to punish the non-contributors could be the way evolution (partially) solves the free riding problem inherent in such an arrangement. The reward or punishment need not come directly from the benefactor or the injured; hence we refer to our reciprocity conjecture as the *indirect* reciprocity hypothesis. Güth (1995), Huck and Oechssler, (1995), and Kockesen, and Ok and Sethi (1997) study evolutionary models that produce conclusions along these lines. Ellingsen (1997) studies a bargaining game in which evolutionary forces are allowed to shape behavior. The model suggests that a concern for fairness persists because it averts exploitation and reduces the probability of conflict.

In its present form, ERC has some clear limitations. Most have to do with the fact that ERC is a theory of “local behavior.” ERC explains *stable patterns* for relatively *simple games*, played over a *short time span* in a *constant frame*. The most important challenges for extending ERC have to do with the italicized phrases. Incorporating learning requires a dynamic theory (although the present version of ERC helps us to understand *what* people learn). We suspect that an analysis of more complicated games will require us to deal with cognitive limitations – bounded rationality. Consideration of longer time spans will force us to deal with changes in motivation; that is, changes in how people weight their goals as their age and outlook change. There is room to extend ERC to the framing issue: A more sophisticated definition of the social reference point, for example, may be a function of the frame.

All of these limitations can, in principle, be straddled. At the same time, if only because of the sheer volume of data it organizes, we doubt that the necessary extensions will supplant the basic message of the present work: The interaction between pecuniary and relative motives drives behavior in many games. For this reason, we think that ERC provides a promising base for a much larger theory of economic behavior.

Ockenfels (1997) describe evidence that dictator gifts accurately predict dilemma game contributions (as ERC would predict).

References

Abbink, Klaus, Gary E Bolton, Abdolkarim Sadrieh and Fang-Fang Tang (1996), "Adaptive Learning versus Punishment in Ultimatum Bargaining," working paper, University of Bonn.

Andreoni, James (1989), "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, 97, 1447-1458.

Andreoni, James, and John H. Miller (1996), "Giving According to GARP: An Experimental Study of Rationality and Altruism," working paper, University of Wisconsin.

Andreoni, James, and John H. Miller (1993), "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *Economic Journal*, 103, 570-585.

Berg, Joyce, John Dickhaut, and Kevin McCabe (1995), "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10, 122-142.

Binmore, Ken (1992), *Fun and Games: A Text on Game Theory*, Lexington, MA: D.C. Heath.

Bixenstine, V. Edwin, and Kellog V. Wilson (1963), "Effects of Level of Cooperative Choice by the Other Player on Choices in a Prisoner's Dilemma Game. Part II," *Journal of Abnormal and Social Psychology*, 67 (2), 139-147.

Blount, Sally (1995), "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior and Human Decision Processes*, 63, 131-144.

Bolton, Gary E (1997), "Strong and Weak Equity Effects: Evidence, Significance and Origins," in *Games and Human Behavior* (D. Budescu, I. Erev, and R. Zwick, eds.), Kluwer Academic Publishers.

Bolton, Gary E (1991), "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, 81, 1096-1136.

Bolton, Gary E, Jordi Brandts and Elena Katok (1996), "A Simple Test of Explanations for Contributions in Dilemma Games," working paper, Institut d'Analisi Economica.

Bolton, Gary E, Jordi Brandts and Axel Ockenfels (1997), "Measuring Motivation in the Reciprocal Responses Observed in a Dilemma Game," working paper, Institut d'Analisi Economica.

Bolton, Gary E, Elena Katok and Rami Zwick (forthcoming), "Dictator Game Giving: Rules of Fairness versus Acts of Kindness," *International Journal of Game Theory*.

Bolton, Gary E, and Rami Zwick (1995), "Anonymity versus Punishment in Ultimatum Bargaining," *Games and Economic Behavior*, 10, 95-121.

Camerer, Colin, and Keith Weigelt (1988), "Experimental Tests of a Sequential Equilibrium Reputation Model," *Econometrica*, 56 (1), 1-36.

Charness, Gary (1996), "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation," working paper, UC Berkeley.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross (1996), "Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games," *Games and Economic Behavior*, 12, 187-218.

Duffy, John, and Nick Feltovich (forthcoming), "Does Observation of Others Affect Learning in Strategic Environments? An Experimental Study," *International Journal of Game Theory.*

Duesenberry, James S. (1949), "Income, Savings, and the Theory of Consumer Behaviour," Cambridge, Mass.: Harvard University Press.

Ellingsen, Tore (1997), "The Evolution of Bargaining Behavior," *Quarterly Journal of Economics*, 112 (2), 581-602.

Erev, Ido, and Alvin E. Roth (1997), "Modeling How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Equilibria," working paper, University of Pittsburgh.

Fehr, Ernst, Simon Gaechter and Georg Kirchsteiger (1997), "Reciprocity as a Contract Enforcement Device, Experimental Evidence", *Econometrica*, 65 (4), 833-860.

Fehr, Ernst, Georg Kirchsteiger and Arno Riedl (1993), "Does Fairness Prevent Market Clearing: An Experimental Investigation," *Quarterly Journal of Economics*, 108, 437-459.

Fehr, Ernst, and Klaus Schmidt (1997), "How to Account for Fair and Unfair Outcomes – A Model of Biased Inequality Aversion," July, Gerzensee Symposium on Economic Theory.

Forsythe, Robert, Joel Horowitz, N. E. Savin and Martin Sefton (1994), "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, 6, 347-369.

Fouraker, Lawrence E. and Sidney Siegel (1963), *Bargaining Behavior*, New York: McGraw Hill.

Gale, John, Kenneth G. Binmore and Larry Samuelson (1995), "Learning to be Imperfect: The Ultimatum Game," *Games and Economic Behavior,* 8, 56-90.

Güth, Werner (1995), "An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives," *International Journal of Game Theory*, 24, 323-344.

Güth, Werner, and Steffen Huck (forthcoming), "From Ultimatum Bargaining to Dictatorship: An Experimental Study of Four Games Varying in Veto Power," *Metroeconomica*.

Güth, Werner, and Eric van Damme (forthcoming), "Information, Strategic Behavior and Fairness in Ultimatum Bargaining: An Experimental Study," *Journal of Mathematical Psychology*.

Güth, Werner, R. Schmittberger and B. Schwarze (1982), "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, 3, 367-388.

Harrison, Glenn W., and Jack Hirshleifer (1989), An Experimental Evaluation of Weakest Link/Best Shot Models of Public Goods, *Journal of Political Economy*, 97 (1), 201-225.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat and Vernon Smith (1994), "Preferences, Property Rights and Anonymity in Bargaining Games," *Games and Economic Behavior*, 7, 346-380.

Holt, Charles A. (1985): "An Experimental Test of the Consistent-Conjecture Hypothesis," *American Economic Review*, 75, 314-325.

Holt, Charles A. (1995), "Industrial Organization: A Survey of Laboratory Research," *Handbook of Experimental Economics* (John H. Kagel and Alvin E. Roth, eds.), Princeton: Princeton University Press, 349-443.

Huck, Steffen, and Joerg Oechssler (1995), "The Indirect Evolutionary Approach to Explaining Fair Allocations," working paper, Humboldt University.

Huck, Steffen, Hans-Theo Normann and Joerg Oechssler (1997), "Stability of the Cournot Process – Experimental Evidence," working paper, Humboldt University.

Jung, Yun Joo, John Kagel and Dan Levin (1994), "On the Existence of Predatory Pricing: An Experimental Study of Reputation and Entry Deterrence in the Chain-Store Game," *RAND Journal of Economics*, 25 (1), 72-93.

Kagel, John, Chung Kim and Donald Moser (1996), "Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs," *Games and Economic Behavior*, 13, 100-110.

Kockesen Levent, Efe A. Ok, and Rajiv Sethi (1997), "Interdependent Preference Formation," working paper, Barnard College.

Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982), "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27, 245-252.

Lave, Lester B. (1965), "Factors Affecting Co-operation in the Prisoner's Dilemma," *Behavioral Science*, 10, 26-38.

Ledyard, John (1995), "Public Goods: A Survey of Experimental Research," in *Handbook of Experimental Economics* (John H. Kagel and Alvin E. Roth, eds.), Princeton: Princeton University Press, 111-194.

Levine, David K. (1995), "Modeling Altruism and Spitefulness in Experiments," working paper, UCLA.

Loewenstein, George F., Leigh Thompson and Max H. Bazerman (1989), "Social Utility and Decision Making in Interpersonal Contexts," *Journal of Personality and Social Psychology*, 57 (3), 426-441.

McKelvey, Richard D., and Thomas R. Palfrey (1992), "An Experimental Stud of the Centipede Game," *Econometrica*, 60 (4), 803-836.

Mitzkewitz, Michael, and Rosemarie Nagel (1993), "Envy, Greed and Anticipation in Ultimatum Games with Incomplete Information," *International Journal of Game Theory*, 22, 171-198.

Nagel, Rosemarie (1995), "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85, 1313-1326.

Neral, John, and Jack Ochs (1992), "The Sequential Equilibrium Theory of Reputation Building: A Further Test," *Econometrica*, 60 (5), 1151-1169.

Ockenfels, Axel, and Joachim Weimann (1996), "Types and Patterns – An Experimental East-West Comparison of Cooperation and Solidarity," working paper, University of Magdeburg.

Prasnikar, Vesna (1997), "Learning the Decision Rules in Ultimatum Games," working paper, University of Pittsburgh.

Prasnikar, Vesna, and Alvin E. Roth (1992), "Considerations of Fairness and Strategy: Experimental Data From Sequential Games," *Quarterly Journal of Economics*, 107, 865-888.

Pruitt, Dean G. (1970), "Motivational Processes in the Decomposed Prisoner's Dilemma Game," *Journal of Personality and Social Psychology*, 14 (3), 227-238.

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

Rapoport, Amnon, James A. Sundali and Richard E. Potter (1992), "Ultimatum Games with Incomplete Information: Effects of the Variability of the Pie Size," University of Arizona, mimeo.

Rapoport, Anatol, and Albert M. Chammah (1965), "*Prisoner's Dilemma: A Study in Conflict and Cooperation*", Ann Arbor: University of Michigan Press.

Roth, Alvin E. (1995), "Bargaining Experiments," in *Handbook of Experimental Economics* (J. Kagel and A. E. Roth, eds.), Princeton: Princeton University Press.

Roth, Alvin E., and Ido Erev (1995), "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 8, 164-212.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991), "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo," *American Economic Review*, 81, 1068-1095.

Selten, Reinhard, and Axel Ockenfels (forthcoming), "An Experimental Solidarity Game," *Journal of Economic Behavior and Organization*.

Selten, Reinhard, and Rolf Stoecker (1986), "End Behaviour in Sequences of Finite Prisoner's Dilemma Supergames: A Learning Theory Approach," *Journal of Economic Behavior and Organization*, 7, 47-70.

Slonin, Robert, and Alvin E. Roth (forthcoming), "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic," *Econometrica*.

de Waal, Frans (1996), *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*, Cambridge, MA: Harvard University Press.

Weimann, Joachim (1994), "Individual Behavior in a Free Riding Experiment," *Journal of Public Economics*, 54, 185-200.

Appendix

Actual costs of effort in Fehr et al. (1993):

e	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
c(*e*)	0	1	2	4	6	8	10	12	15	18

Proof of the Effort Hypothesis: $\bar{e}'(w) := \partial\left[(1-\alpha)e^E + \alpha e^R(w)\right] / \partial w \geq 0$

From the best response functions the following is true: $e^{*\prime}(w) > 0 \Rightarrow \bar{e}'(w) \geq 0$.

From the implicit definition of *e**(*w*), we have:

$$ve^{*\prime}(w) - e^*(w) - we^{*\prime}(w) = 1 - c'(e^*(w))e^{*\prime}(w) \Leftrightarrow$$

$$e^{*\prime}(w) = \frac{1+e^*(w)}{v-w+c'(e^*(w))} > 0$$

q.e.d.

Proof of the Worker Payoff Hypothesis: $\bar{u}'(w) := \partial\bar{u}(e(w)) / \partial w \geq 0,\ \forall\alpha \in [0,1]$

From $\bar{u}(w) = \ a u^R(e^R(w)) + (1-a)u^E(e^E(w)) \ = \ a(w - c(e^*(w)) - c_0) + (1-a)(w - c_0)$

the following is true: $\bar{u}'(w) \geq 0$ for $\underline{w} < w < \overline{w} \Rightarrow \bar{u}'(w) \geq 0 \forall w$.

For $\underline{w} < w < \overline{w}$, we have:

$$\bar{u}'(w) = 1 - ac'(e^*(w))e^{*\prime}(w) \geq 0 \Leftrightarrow$$

$$1 \geq ac'(e^*)\frac{1+e^*(w)}{v-w+c'(e^*(w))}$$

Since $\alpha \leq 1, c'(e) \leq 30, e^*(w) \leq 1,$ and $v - w \geq v - \overline{w} = 41$ in Fehr et al. (v = 126 and $\overline{w}$ = 85), we have:

$$\alpha c'(e^*)\frac{1+e^*(w)}{v-w+c'(e^*)} < 0.85 < 1$$

q.e.d.

Proof of the concavity of the profit function: $\overline{\pi}''(w) < 0$ for $\underline{w} < w < \overline{w}$

$$\overline{\pi}'(w) = -\frac{1-\alpha}{10} + \alpha e^{*\prime}(w)(v-w) - \alpha e^{*}(w)$$

$$\overline{\pi}''(w) = \alpha e^{*\prime\prime}(w)(v-w) - 2\alpha e^{*\prime}(w)$$

From the implicit definition of $e^*(w)$ we have:

$$ve^{*\prime\prime}(w) - 2e^{*\prime}(w) - we^{*\prime\prime}(w) = -c''(e^*(w))\left(e^{*\prime}(w)\right)^2 - c'(e^*(w))e^{*\prime\prime}(w) \Leftrightarrow$$

$$e^{*\prime\prime} = -\frac{e^{*\prime}(w)\left(c''(e^*)e^{*\prime} - 2\right)}{v - w + c'(e^*)}$$

This yields:

$$\overline{p}''(w) = -a(v-w)\frac{e^{*\prime}(w)\left(c''(e^*(w))e^{*\prime}(w) - 2\right)}{v - w + c'(e^*(w))} - 2ae^{*\prime}(w) < 0 \Leftrightarrow$$

$$-c''(e^*(w))e^{*\prime}(w)(v-w) < 2c'(e^*(w))$$

which is true by the convexity of $c(e)$ and the proof of the effort hypothesis.

q.e.d.

Calculation of $\underline{a}$: $\overline{\pi}'(\underline{w};\underline{\alpha}) = 0 \Rightarrow \underline{\alpha} \approx 10\%$

$$\overline{\pi}'(\underline{w};\underline{\alpha}) = -\frac{1-\alpha}{10} + \underline{\alpha}e^{*\prime}(\underline{w})(v-\underline{w}) - \underline{\alpha}e^{*}(\underline{w}) = 0 \Leftrightarrow$$

$$(v-\underline{w})\frac{1+e^*(\underline{w})}{v-\underline{w}+c'(e^*(\underline{w}))} - e^*(\underline{w}) = \frac{1-\alpha}{10\alpha}$$

Since $e^*(\underline{w}) = 0.1$ we have:

$$\underline{\alpha} = \frac{v - \underline{w} + c'(0.1)}{11(v-\underline{w})}$$

In Fehr et al. (Charness) we have $c'(0.1) = 10$ (10) and $\underline{w} = 40\,(29)$, so that $\underline{\alpha} = 10.1\%$ (rounded) for both papers.

q.e.d.

Figure 1. Amounts offered to the recipient in dictator and ultimatum games (Forsythe et al., 1994)

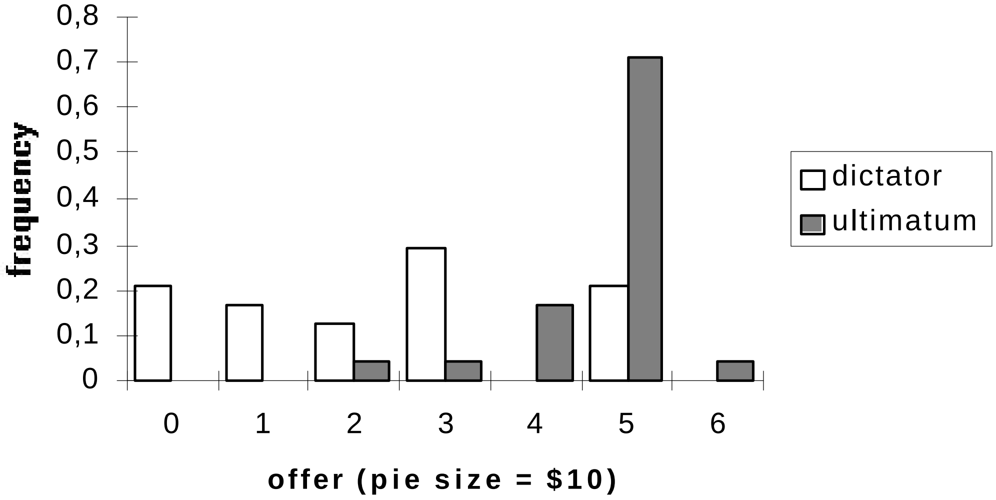


Figure 2. Average Effort in Response to Wage (Fehr et al., 1993)

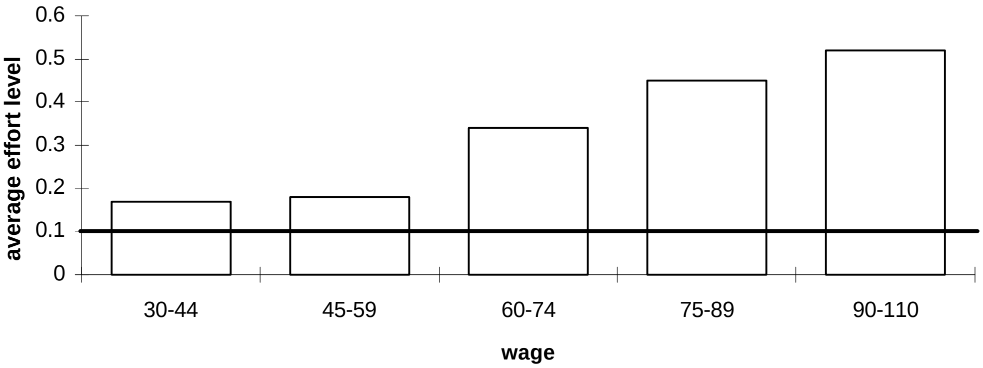
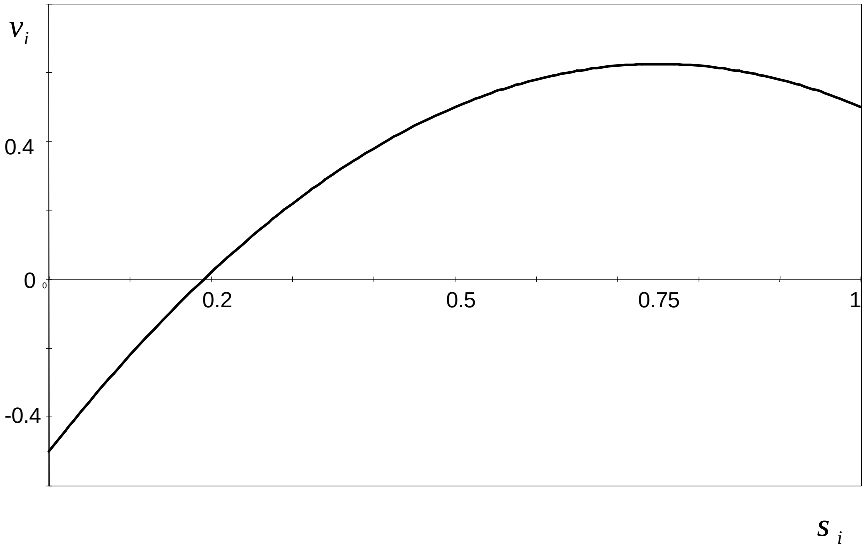


Figure 3. Additively separable motivation function, with c = 1, and a/b = ¼ (r_i = 3/4, s_i = 1/5)



Figures 4 (a) and 4(b). ERC-reaction curves in a Cournot duopoly

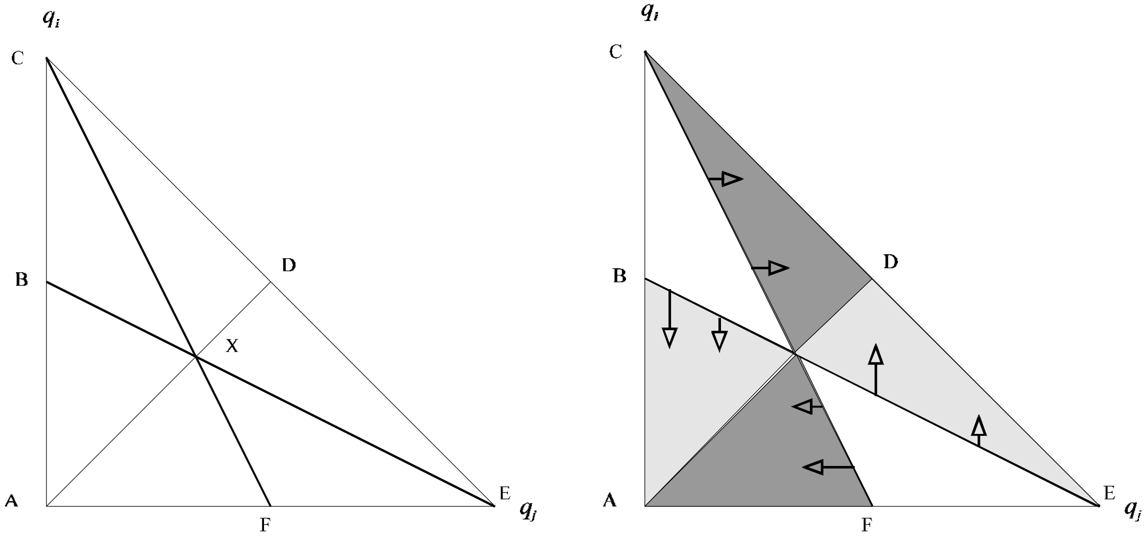


Figure 5(a). Actual and predicted average effort levels

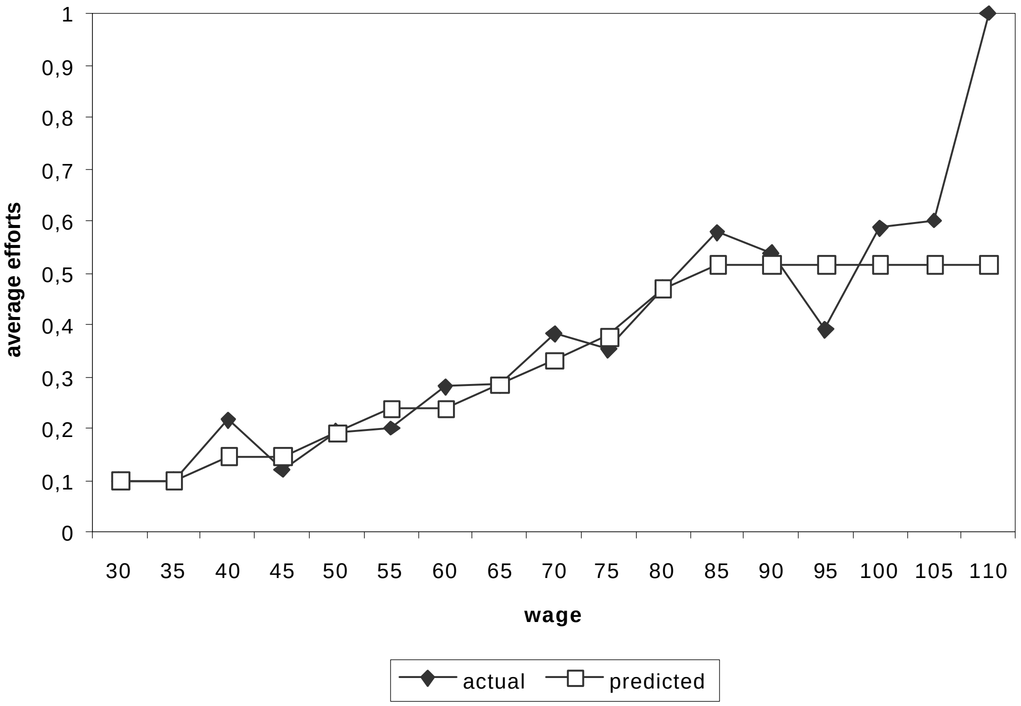


Figure 5(b). Actual and predicted average payoffs of the workers

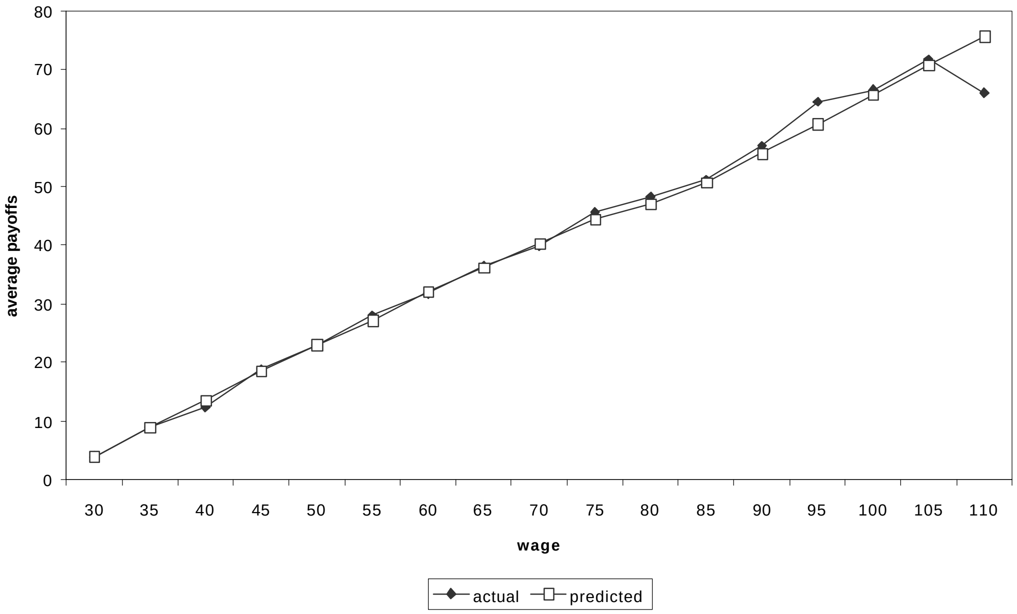


Figure 5(c). Actual and predicted average payoffs to the firms, and actual wage offers

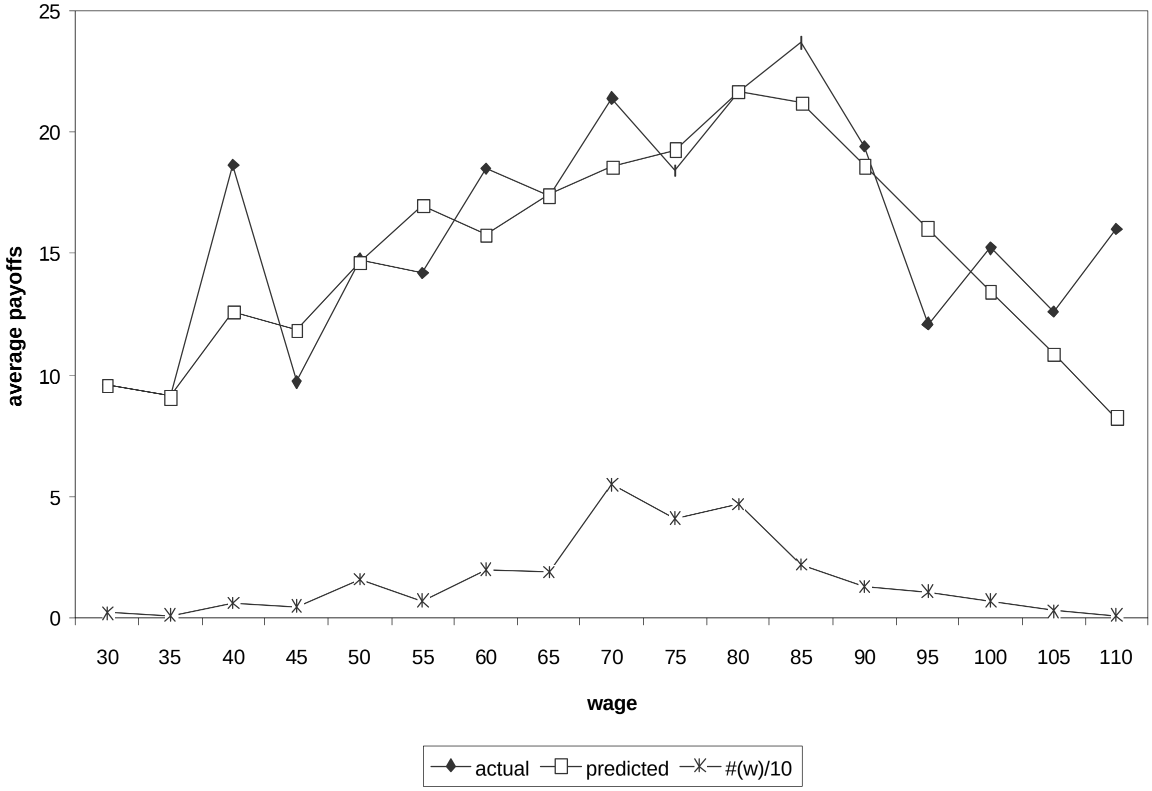


Figure 6(a). Probability first mover cooperates (by round)

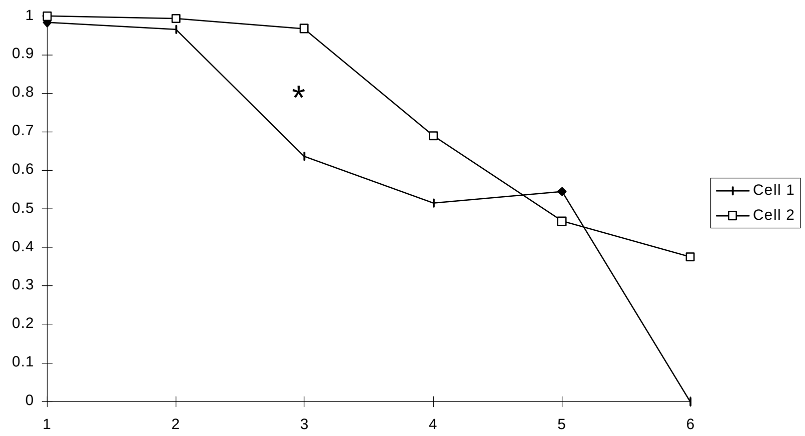


Figure 6(b). Probability second mover cooperates given first mover cooperates (by round)

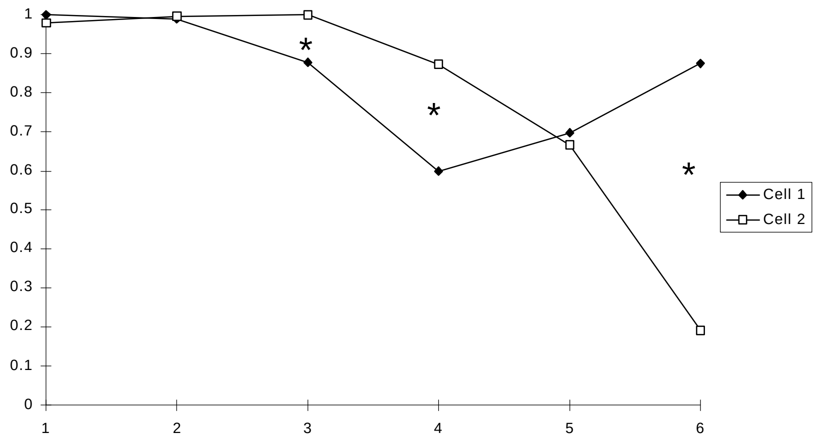


Table 1. A comparison of payoffs for the mini-games.

mini-	*proposer*	*left*		*right*	
games	*responder*	*reject*	*accept*	*reject*	*accept*
ultimatum	y_P	0	2	0	3
game	y_R	0	2	0	1
impunity	y_P	0	2	3	3
game	y_R	0	2	0	1
best shot	y_P	1	1	0	3
game	y_R	3	1	0	1

Table 2. One-shot prisoner's dilemma payoff matrix

		2	
	y_1, y_2	cooperate (C)	defect (D)
1	cooperate (C)	2m, 2m	m, 1+m
	defect (D)	1+m, m	1, 1

m = marginal per capita return (mpcr) $\in (0.5,1)$

Table 3. Sequential stage game

		2			
	y_1, y_2	cooperate (C)	defect (D)	proportions	sessions
1	cooperate (C)	40, 60	-100,150 -100,0	.66 .33	Camerer and Weigelt
			-100,150 -100,0	.66 .33	Neral and Ochs cell 1
			-100,100 -100,0	.66 .33	Neral and Ochs cell 2
	defect (D)	10,10	10,10		