Raúl García-Barrios y David Mayer

Justice and Efficiency in Economic Relations: Explaining Collaboration and Conflict in the Firm and Choice in Ultimatum Games
División de Economía
Número 62

Raúl García-Barrios y David Mayer

Justice and Efficiency in Economic Relations: Explaining Collaboration and Conflict in the Firm and Choice in Ultimatum Games
JUSTICE AND EFFICIENCY IN ECONOMIC RELATIONS:
explaining collaboration and conflict in the firm
and choice in ultimatum games

(Please do not cite without permission of the authors)
December 1995

Raúl García-Barrios
David Mayer-Foulkes

División de Economía
Centro de Investigación y Docencia Económicas, A.C.
Carretera México-Toluca # 3655
Apartado Postal 10-883
Col. Lomas de Sta. Fé,
México, D.F., 12080
garcia@dis1.cide.mx
mayerfou@dis1.cide.mx
SUMMARY

Individual preferences are sensitive to particular economic interactions. Complex psychological phenomena affect the meaning and ordering of their objects according to different institutional patterns and strategical arrangements. We provide an agency model that explains preference endogenity in reference to higher order normative principles and individual moral judgment. We use this model to explain the organization of labor regimes in the firm and the formation of preferences and choice in ultimatum games. In the firm, workers judge the fairness of the distribution of the gains from cooperation and consequently adopt working attitudes that determine the equilibrium level of information asymmetry, transaction costs and inefficiency. We explore from this perspective the formation and distribution of gains from cooperation, and their relation with the adoption of different labor regimes. A regime in which the parties maintain fair working relations clears up potential market and information failures and may generate a collaborative equilibrium enhancing maximum global efficiency. Other distributive strategies, however, uncover such failures and reduce efficiency. "Efficiency wage" regimes, while having the potential for increasing fairness and efficiency, may evolve into relations generating transfers of value in favor of the firm considered unfair by the workers; and produce resource misallocation and an equilibrium where strongly dissonant workers constitute a "moral hazard" for the firm. We analyze the conditions that determine the firm's choice.
I. INTRODUCTION

Workers' attitudes towards sharing resources and information with firm managers are usually very sensitive to the organization of capital-labor interaction. Managers may use the high sensibility of workers' beliefs and preferences to enhance labor productivity and efficiency (Akerlof, 1982; Doeringer et al, 1986; Doeringer, 1991). Such working relations and managerial practices are referred to as "superefficient". They include good human relations, participative decision making, job enrichment, work rotation, and a "fair" pay. These practices will stimulate individual propensity to work, collaborate, share information and receive new training (Brown, 1980; Mowday, Porter, and Steers, 1982; Deming, 1991; Kohn, 1993).

Many firms adopt instead practices that reinforce conflictive capital-labor relationships and produce welfare loss (Reynolds, 1951; Doeringer and Piore, 1985; Doeringer, 1991). The violation of the workers' legal or customary distribution or procedural rights, the systematic criticism of their performance and intense supervision of their work, the combined use of "sticks and carrots", the physical or psychological isolation of workers to control potential organization, etc., constitute restrictions to effective socialization, suppress intelligence, self-esteem and creativity, and produce in the workers feelings of vulnerability and marginalization. As a consequence, workers accumulate aversion towards supervisors and managers, turn their energies to frustrate the projects of the firm (MacIntyre, 1985), become opaque and unpredictable to the firm, increasing information imperfections (Scott, 1985) and build up market power affecting other workers (e.g., they hoard specific labor skills and information; Thurow, 1984). Workers may also experience lack of energy, guilt, identity crisis, self-depreciation, and profound apathy or depression (Fromm & Maccoby, 1973), which will reflect in a lack of commitment and productivity (Buunk, et al, 1993; Ostroff, 1993).

Why would firms choose to adopt strategies that generate moral hazard? This paper explains this apparent paradox by exploring how labor contracting and management affect the formation and distribution of gains of cooperation between labor and capital. To narrow our focus, we analyze situations in which both the workers' effort and productivity are non-contractible, and compare the traditional "efficiency wages" regime (as described by Shapiro & Stiglitz, 1985) with "superefficient" labor regimes having similar contractual elements (i.e., a fixed wage rate and a working standard). In this framework, we analyze the role of the principles of distributive justice and reciprocity, which are determinant in the formation and destruction of labor commitment1.

Given this framework, we provide theoretical support for the following propositions: (1) There are situations in which workers' moral deficiency (defined as their incapacity to behave according to their own values of justice given the power that internal or external incentives exert over them) justify the use of "efficiency wages" as a short-run method to increase efficiency and maximize profits. This proposition corresponds to the standard idea behind efficiency wages (see Shapiro & Stiglitz, 1985; Weiss, 1990). (2) However, the firm's observance of the workers' values of justice will also increase efficiency and sometimes maximize profits. (3) If the workers are morally sound (incentives are not sufficient to induce them to deviate from their beliefs), such observance will take the economy, in the absence of coordination problems, to a unique efficient equilibrium that will attain welfare maximization. (4) The purpose of an "efficiency wages" regime may not be to correct a problem of opportunism in workers' behavior and

---

1 These principles of economic normative behavior are well recognized by social psychologists and anthropologists (Malinowski, 1921; Thurnwald, 1932; Polanyi, 1975) but have only recently been associated with economic analysis (for a survey, see Hausman & McPherson, 1993). Economists have increasingly accepted the importance of the notions of fairness in market regulation (Kahneman, Knetsch and Thaler, 1986), price and wage stickiness (Hirschman, 1970; Akerlof, 1982), and rigidities in the operation of customer markets (Okun, 1981).
increase efficiency, but to generate a transfer of value running from the workers to the firm that may maximize profits but, because it is perceived as unfair by the workers, increases informational failure and supervision costs and reduces efficiency. We refer to such process as exploitation.

We analyze the foundations of these propositions in the following two sections. Section II proposes a model of human agency that incorporates normative reasoning and is useful for game-theoretical analysis. Considerable evidence derived from experimental ultimatum games supports the hypotheses that economic agents consider the fairness of distribution and may be willing to forego income opportunities when they imply transactions perceived as unfair, or alternatively they may be willing to pay a cost to punish unfair allocators, i.e., they practice reciprocity (Kahneman, Knetsch and Thaler, 1986; Camerer and Thaler, 1995). In this section, we provide analytical examples of how a more developed model of normative reasoning may explain these behaviors.

Section III enunciates and proves the central theorem of the paper supporting our four propositions. Proposition (4) deserves some introduction. Most economists think of "efficiency wages" as a solution and not a cause of inefficiency. More generally, they believe that strategies chosen by a profit maximizing firm will not reduce their internal productive efficiency and increase transaction costs. However, the firm is an institution regulating capital-labor cooperation, and according to Douglass North: "Institutions are not necessarily or even usually created to be socially efficient; rather they, or at least the formal rules, are created to serve the interests of those with the bargaining power to devise new rules" (North, 1992, page 16). In the firm, the unfair exercise of bargaining power will generate incompatibility between the pursuit of the firm's interests and the well-being of the cooperative collectivity, provoking in the workers behavioral adjustments governed by the rules of reciprocity. The use of "efficiency wages" as a mechanism of exploitation will demand the destructuring of social relations that satisfy workers' claims on fair treatment, making it necessary for the firm to control their power to retribute through a costly system of supervision. To reduce the costs, the firm may institutionalize labor stratification to combine collaborative and exploitative regimes. In this paper, we explore the workers' reactions to exploitation and the institutional structures and transaction costs they generate, thus developing at the micro level Polanyi's tradition of "substantive economics" first developed in *The Great Transformation* (Polanyi, 1975).

Once we have analyzed the economic implications of moral reasoning, we consider how this form of intentional behavior develop (Section IV) and why it may lead to moral deficiency (Section V). The purpose of analyzing the ontogenesis of normative reasoning is twofold. First, to situate from a common developmental perspective the different models of human morality used in economic analysis. Our model captures as special cases two extreme models of morality: opportunism (as described by Williamson, 1985) and conventionalism (i.e., consistency with beliefs and norms irrespective of outcomes; see Kahneman et. al., 1986; Frey, 1986; Neurborg, 1991; Miller, 1992). The general case will correspond to situations in which individuals may deviate, if rewards or punishments are sufficiently strong, from the dictates of their moral judgment, itself based on general moral values. Second, to emphasize that the development of normative reasoning is a complex process. Morality evolves in multidimensional and interactionist patterns, and most adults maintain areas of normative reasoning that are contaminated by processes that characterize earlier stages of moral judgment, facilitating the emergence of moral deficiency and internal conflict.

Moral deficiency and conflict are pervasive phenomena in modern societies (McIntyre, 1985). A full discussion of their multiple and complex causes is beyond the scope of this paper. However, we explore several situations that produce systemic moral deficiency and internal conflict in workers. We show the considerable complexity economic analysis may acquire once preferences are no longer exogenous: criteria of efficiency and justice become analytically inseparable, and information and institutional failures become endogenous.
II. THE PSYCHOLOGY OF NORMATIVE REASONING AND THE FORMATION OF PREFERENCES

1.- Normative judgment: general values and adaptable behavioral norms

Adult individuals usually guide their normative interaction with others through a complex process of normative judgment based on the acceptance of general values. These are higher order normative rules (i.e., rules that specify the norms individuals must apply in each specific context) whose reference is very broad (Smelser, 1963). They provide premises that reflect the states of social well-being that socialized individuals believe are ultimately desirable and which they use to form their daily normative judgments. Although historically determined, and group specific (they vary with gender, social class, and culture, (Van Lange, 1992)), individuals believe them to be "universal" and fundamental in the regulation of human interaction, and use them consistently for the definition of individual and collective rights and the evaluation of the legitimacy of human actions. Examples of general values that function in the working place are the claim for equitable distribution of income, the respect for individual dignity, the right to social security and the right of the people to resist injustice (Max Neef, et.al., 1986; Punzo & Meara, 1993).

Normative judgments depart from general values to generate more specific rules of behavior (i.e., norms in the usual sense). Clearly, such behavioral norms depend on the values providing the normative premises and on the situation under normative analysis. Behavioral norms resulting from individual normative reasoning may become social through human communication, or result in the qualification of existing social norms, thus becoming a source of legal or customary change.

Behavioral norms carry the motivational force of the general values from which they derive, and thus their internal reasons for individual commitment. Social psychologists have focused on understanding how and why they may constitute a force that modifies individual attitudes and actions. Well-established theories that provide light on this issue are those based on congruence models of human behavior such as Heider's balance theory (Heider, 1958) and Festinger's cognitive dissonance theory (Festinger, 1957; 1964). According to congruence models, adult people tend to maintain a balance amongst their emotional, cognitive and behavioral elements, and seek to maintain congruity with the general circumstances that surround their life (Ostroff, 1993; Sánchez, 1994). This comes from their capacity of perceptual attribution, that is, their rational evaluation of their own beliefs and actions (Harvey & Weary, 1981). Such evaluation may lead to incongruence in the elements in our own personal perceptions, preferences, beliefs, sentiments, attitudes or actions, generating emotional instability and a feeling of uneasiness referred to as dissonance. The greater dissonance becomes, the greater the urgency to acquire the necessary balance or consonance. As a consequence of dissonance, individuals will experience conscious and unconscious mental processes tending to change their desires for specific external objects, relationships and social situations, which will be reflected in the modification of their attitudes and actions.

We therefore expect the dictates of moral judgment to affect human action. Normative dictates and actions, however, may not coincide, since normative reasoning is not the only mechanism of preference formation in adult individuals. Other factors, such as prudence, greed or fear, which respond to immediate external incentives, may have important effects on individual choice (Perlman & Cozby, 1983; Glassman, 1994). Normative dictates, however, have powerful emotional content, and individuals who do not follow them remain dissonant, suffering in the long run mental confusion, unhappiness, regret and guilt, lower self-esteem, anger, anxiety or depression. Many individuals will eventually respond by changing the importance they attribute to the external factors that make them deviate from their behavioral norms, to increase the consistency of
their actions with their own normative beliefs. Even if actions may be permanently at odds with behavioral norms, they will not escape their reference.

2. The adoption of values of justice and reciprocity

Two important general values are the adoption of values of justice and reciprocity. These lie at the core of many fundamental social interactions, such as the family, patron-client relationships, kinship and community, and conform a structural basis for many social organizations in which loyalty or prestige are important, such as religious churches and political states (Polanyi, 1974; Herkovitz, 1954).

The adoption of values of justice implies the acceptance of the duty of just play in daily interaction. This reflects the recognition in each participant of the needs, aspirations and interests of the others. It usually implies the recognition of legitimate differences in income and privileges amongst individuals, social groups or nations belonging to different status or having different economic power. In modern western societies, principles of equity tend to profoundly affect beliefs on the allocation of the benefits of cooperation (Adams, 1965; Lind & Tyler, 1988). Other principles, however, compete with equity in governing the distribution and redistribution of resources and welfare in such societies, such as the principles of equality, social security and difference (Miller, 1992; Mitchel, et al, 1993). In pre-capitalist societies, the institutions of centricity and symmetry, which were essential components of the existing systems of labor division, foreign trade, taxation for public purposes, and defense provisions (Polanyi, 1975), were based on different principles of justice (e.g., redistribution and moral economy; Thurnwald, 1932; Wolf, 1966; Watts, 1984).

The recognition of differences in power and status in the definition of justice principles does not imply that they must coincide with the potential outcomes of self-interested bargaining. Individuals systematically refrain (and are expected to refrain) from using their full power capacity during joint activity due to moral constraining forces. The rationality of such ethical forces has been subject to a long-lasting philosophical debate. They may be grounded on a specific telos, as in Aristotle, Bhudda and modern utilitarianism, or on the autonomous pursuits of universal rational principles, as in Kant, on some type of radical uncertainty, as in Rawls, or on other factors (for surveys of theories of justice see Barry, 1989, and Hausman & McPherson, 1993). Adult individuals, however, have powerful self-seeking motivations to deviate from the principles that dictate the use of power. The operation of another ethical principle, reciprocity, regulates the reaction of people to deviant conducts that seek illegitimate opportunity or make excessive use of economic and social power.

Reciprocity is the principle of normative behavior that defines the actions necessary to restore any deviation from fair distribution or procedure and to reduce psychological dissonance. In modern societies, people feeling distress due to inequity promptly act to restore equity one way or another. In most cases, this is true even for those benefiting from inequity (Adams, 1963). Reciprocity operates by norming those actions, usually through the exchange of gifts or punishments among social and economic parties (Scott, 1985). Such exchange is usually important in economic life, due to its role in the exchange of individual advantages in technology, information and economic power among economic agents. Reciprocity may thus affect the efficiency of cooperative action by reducing or increasing market and other institutional failures and thus transaction costs.

Let us compare the operation of these moral principles with more orthodox economic means of regulating institutional failure, such as self-interest joint ventures and contracts. Contracts, as defined by most economists, are products of hard bargaining or competition that motivate cooperation by acting on the individual's enlightened self-

---

2 Reciprocity also regulates the actions prompted by sentiments different from injustice distress but which may accompany it, such as envy or excessive generosity and altruism.
interest. Economic analysis focus on the *stricti juris* terms of reward and punishment (e.g., pay incentives and threats of discharge, terms of repeated interaction) or on customary and social sanctions (e.g., gossiping). In contrast, moral principles motivate action through self-imposed commitment, which depends critically upon the emotions created by identification, trust and mutual respect. (In orthodox contract theory, trust and prestige have no emotional content and only reflect the subjective probabilities bayesian rational agents attribute to particular actions of other agents.) Even in modern societies, individuals interpret many practices of bargaining and competition as expressions of greed or opportunism, and attempts to make over-explicit the terms of interaction as signs of distrust or disrespect for the dignity or social status of other participants. The emotional reactions to these actions generally undermine voluntary compliance and destabilize production and exchange. All societies discourage or regulate these practices, usually through institutions connected with moral, tradition, etiquette, kinship, social hierarchy and religion (Mauss, 1954; Polanyi, 1977).

Self-enforcement contributes to one of the distinguishing features of normative behavior, that is, its diffuse and personal character. The rules of behavior derived from the principles of justice and reciprocity usually lack the temporal and spatial discreteness of formal contracts. They function more as a continuing relationship between or among persons based upon opportunity to help (or punish) and mutual expectations (Oakerson, 1988). This personal and diffuse character constitutes an important source of comprehensiveness and strength in flexible human organizations, such as the family. However, it may also cause intractable coordination problems, and become a handicap in complex production or exchange systems (Bloch, 1961).

3. Normative judgement in ultimatum games: a basic model

By modelling ultimatum games we may clarify our previous discussion. Following Camerer & Thaler (1995), we may describe these games as follows. Two players (who we call Proposer and Responder) are allotted a sum of money. Proposer offers some portion of money to Responder. If Responder accepts, she gets what was offered, and Proposer gets the rest. If Responder rejects the offer, both get nothing. In the case that Proposer and Responder only care for the monetary pay-offs of the game, game-theoretical analysis predicts a subgame perfect equilibrium in which Responder will accept any share offered by Proposer, who will offer the smallest unit of currency available. Furthermore, Responder's threats to reject any other offer will not be credible. However, the following conducts have robust empirical support: (see Kahneman, Knetsch and Thaler, 1986 and Camerer & Thaler, 1995):

(a) Usually, individuals playing the role of Proposer will offer a larger share than predicted.
(b) Individuals playing the role of Responder will usually be reluctant to accept positive offers which imply a very unequal and therefore unfair distribution.
(c) Responders will be willing to pay a cost for punishing unfair offers.
(d) Proposers may take advantage of information asymmetries to increase their share (appearance of fairness is enough).

The first three conducts have been interpreted as evidence of the existence of externalities affecting the decisions of the players. The last one, however, is in tension with this interpretation. To explain the process of preference formation involved, capturing the whole set of evidence in a single analytical framework, we consider two different games. First, we analyze the classical ultimatum game, as described by Camerer & Thaler, but our friends, Proposer and Responder, now have some type of morality. They may be morally strong, morally deficient or opportunistic. In the second game, in addition, we allow Responder to punish Proposer at a certain cost. In both games, perfect equilibria will vary according to the type of morality of the players.
Suppose both Proposer and Responder, indistinctly of their moral strength, share a value of equality and believe that in an ultimatum game payoffs should be equal. Suppose they play to distribute one dollar. Let $S_R$ be the share offered by Proposer (if Responder accepts, Proposer would obtain with $S_P = 1 - S_R$). Define the distributive gap function as:

$$T(S_P, S_R) = S_P - S_R$$

The principle of equal distribution of the dollar would mean that:

**Proposition 1:** "$T$ should equal 0".

We assume that ethical values are relevant, i.e., both players will reason their behavioral norms independently of the incentive structure of the model. They will then believe that Proposer should satisfy:

**Proposition 2:** "$S_R$ should make $T(S_P, S_R) = 0$",

or equivalently, that $S_R$ should equal 0.5.

To understand what would be a normative reaction governed by the principle of reciprocity, suppose that after the distribution of the dollar Responder could choose a mechanism (i.e., a transference or a costly punishment), $\Theta$, to correct any fairness anomaly in the payoffs of the game. A simple logical operation would result in the following behavioral norm:

**Proposition 3:** "$\Theta$ should make $T(S_P, S_R; \Theta(S_P, S_R)) = 0$".

Conformance to behavioral norms will not result from purely normative-logical reasoning, but from the players' psychological need to reduce dissonance to maximize utility. Dissonance will be an increasing function of normative incongruence as measured by the distributive gap function, i.e., $D_j = D_j(T)$, $D' > 0$. Proposer and Responder have one of three types of morality (A, B and C) defined by the following dissonance functions:

- $D_A(T) = T^2 + 1.51T1$,  
- $D_B(T) = T^2$,  
- $D_C(T) = 0$

Utility will be a function on the individuals share and dissonance:

$$U_{i,j} = S_i - D_j(T), \quad i = P, R; \quad j = A, B, C$$

Utility maximization may permit normative incongruence ($T\neq 0$) and thus a positive "optimal" dissonance. Players of type A are an example of a strong morality because their decisions will implement $T=0$ under some range of external incentives. (If Proposer is of type A, he will freely choose to be fair.) Players of type B are morally deficient because they will never implement $T=0$. Nevertheless, they will show some degree of moral response. (If Proposer if of type B, he will not be fair but will nevertheless offer a meaningful share.) Players of type C will have no moral response. Moral reasoning will have no effect on their behavior; they are opportunistic.

Finally, suppose the reservation utility of all players is zero, and that their type, strategic set and payoffs are common knowledge.
Game 1: Classical Ultimatum Game (no punishment is allowed).

The game between Proposer and Responder has nine different outcomes according to their moral types. Each outcome will result from a unique perfect equilibrium, whose payoffs are summarized in Table 1. Important results emerge from its analysis. Independently of the morality of his opponent, Proposer's monetary payoffs will tend to decrease the stronger his normative response, while Responder's payoffs will tend to increase. This results from a stronger willingness to forego unfair opportunities and from a stronger reluctance to accept unfair offers.

Game 2: Ultimatum Game with Costly Punishment.

We consider now situations in which Responder may impose a costly punishment if Proposer is unfair. We address the question of how much punishment Responder will demand.

Suppose punishment is discrete and that each unit of punishment costs one cent and consists in withdrawing δ cents from Proposer's payoff. Clearly, δ is a measure of Responder's punishment power. Let x be Responder's demand for punishment. The distributive gap and utility functions may be respecified as follows:

\[ T(S, R, x) = (S - 6x) - (R - x) \]
\[ U_{P, i} = (S - 6x) - D_i(T), \quad i = A, B, C \]
\[ U_{R, j} = (R - x) - D_j(T), \quad j = A, B, C \]

while the rest of the model remains as in Game 1.

In Game 2, Proposer makes an offer, and Responder determines whether to reject the offer or to accept and demand a punishment. The game may be modeled as a Stackelberg game with Proposer as leader. To describe its solutions, we must first make several definitions, together with some comments:

**Definition 1:** Let \( x_{R, j}^{**} = x_{R, j}^{**}(S, \delta), j = A, B, C, \) be the Optimal Punishment, i.e., the solution to Responder's problem:

\[
\max_{x} U_{R, j}(S, \delta, x) = (S - x) - D_j((1 - S) - \delta x - (S - x)) \\
\text{s.t. } U_{R, j} \geq 0, \quad j = A, B, C.
\]

A necessary (but not sufficient) condition for \( x_{R, j}^{**}(S, \delta) > 0, j = A, B \) is \( T > 0 \), since punishing if \( T \leq 0 \) will both reduce income and increase dissonance. If Responder is opportunistic she will never punish an unfair Proposer since her willingness to punish will be nil.

**Definition 2:** Let Responder's Participation Share be the function \( S_{R, j}^P = S_{R, j}^P(\delta) \) defined by

\[ U_{R, j}(S_{R, j}^P(\delta), \delta, x_{R, j}^*(S_{R, j}^P(\delta), \delta)) = 0. \]

For every \( \delta \), Proposer's offer should exceed the corresponding Participation Share. Otherwise, Responder will reject the offer.

**Definition 3:** Let Responder's Critical Share be the function \( S_{R, j}^C = S_{R, j}^C(\delta) \) defined by