

**John F. Kennedy School of Government  
Harvard University  
Faculty Research Working Papers Series**

**More Order with Less Law:  
On Contract Enforcement, Trust and Crowding**

**Iris Bohnet  
Bruno S. Frey  
Steffen Huck**

**July 7, 2000**

**KSG Working Paper No. 00-009**

**This paper can be downloaded without charge from the  
Social Science Research Network Electronic Paper Collection:  
[http://papers.ssrn.com/paper.taf?abstract\\_id=240389](http://papers.ssrn.com/paper.taf?abstract_id=240389)**

**The views expressed in the KSG Faculty Research Working Paper Series are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or Harvard University. All works posted here are owned and copyrighted by the author(s). Papers may be downloaded for personal use only.**

## More order with less law: On contract enforcement, trust, and crowding\*

### Abstract

Most contracts, whether between voters and politicians or between house owners and contractors, are incomplete. “More law,” it typically is assumed, increases the likelihood of contract performance by increasing the probability of enforcement and/or the cost of breach. This paper studies a contractual relationship where the first mover has to decide whether she wants to enter a contract without knowing whether the second mover will perform. We analyze how contract enforceability affects individual performance for exogenous preferences. Then we apply a dynamic model of preference adaptation and find that economic incentives have a non-monotonic impact on behavior. Individuals perform a contract when enforcement is strong or weak but not with medium enforcement probabilities: Trustworthiness is “crowded in” with weak and “crowded out” with medium enforcement. In a laboratory experiment we test our model’s implications and find support for the crowding prediction. Our finding is in line with the recent work on the role of contract enforcement and trust in formerly Communist countries.

Iris Bohnet (corresponding author)

Assistant Professor, Kennedy School of Government, Harvard University  
79 JFK Street, Cambridge, MA 02138, USA

Phone: +1 (617) 495 5605, Fax: +1 (617) 496 5747, email: Iris\_Bohnet@Harvard.edu

Bruno S. Frey

Professor, Institute for Empirical Economic Research, University of Zürich  
Bluemlisalpstreet 10, 8006 Zürich, Switzerland

Phone: +41 (1) 634 3731, Fax: +41 (1) 634 4907, email: bsfrey@iew.unizh.ch

Steffen Huck

Senior Lecturer, Department of Economics, Royal Holloway College, University of London  
Egham, Surrey TW20 0EX, United Kingdom

Fax: +44 (1784) 439 534, email: s.huck@rhbnc.ac.uk

---

\*We thank Linda Babcock, Sam Bowles, Uwe Dulleck, Ernst Fehr, Simon Gächter, Werner Güth, Ernan Haruvy, Owen Jones, Michael Mandler, Jörg Oechssler, Elinor Ostrom, Fred Schauer and Richard Zeckhauser as well as participants of the 1998 American Law and Economics Association meetings at Berkeley, the 1999 Economic Science Association meetings at Lake Tahoe, the 1999 Symposium on Behavioral Economics at the Levy Institute, the 1999 SPUDM meetings in Mannheim, the 2000 Public Choice Association meetings in Charleston, the 2000 meeting of the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms in Chicago, and seminars at Harvard, Royal Holloway and the University of Zürich for many helpful comments and suggestions. Further thanks are due to Dorothea Kübler for her help in conducting the experiments, to Joanna Veltri for her editorial support and to Michael Zürcher for his legal advice. Financial support of the Schweizerischer Nationalfonds and the Kennedy School of Government Dean’s Fund are gratefully acknowledged. The third author also acknowledges financial support of the German Science Foundation (DFG).

# 1 Introduction

Trust can increase efficiency in the economic and the political spheres. Recent studies using aggregate data have suggested the existence of an efficiency-enhancing feature of trust for countries and organizations.<sup>1</sup> This paper attempts at providing a microfoundation for some of these findings. It investigates whether trustworthiness can have an economic payoff at the individual level. Important domains for trust and trustworthiness include the relationship between representatives and their constituents, such as between politicians and voters, managers and shareholders or attorneys and clients. In all these situations, principals have to decide whether they want to enter a contract with an agent, knowing that the contract will be incomplete. Offering a contract, thus, is a matter of trust, and performing it, a matter of trustworthiness.

The problem of trust is more pronounced in large, anonymous societies than in small groups where participants frequently interact. In the latter case, reputation matters and folk theorem-like arguments can sustain cooperation even in the absence of genuine trustworthiness. In the former case, similar arguments tend to fail, creating the need for institutions, such as the law, to facilitate efficient outcomes.<sup>2</sup> The law, however, may affect behavior not only by creating incentives but also by influencing preferences. Rational choice theory has focused on the first aspect. In this paper, we propose a model integrating both effects. We analyze how the enforceability of a contract affects individual performance in the short run

---

<sup>1</sup> See, for example, Putnam (1993), Fukuyama (1995), La Porta *et al.* (1997) and Knack and Keefer (1997).

<sup>2</sup> See, for example, Milgrom, North, and Weingast (1990) and Greif, Milgrom, and Weingast (1994). For a proof of the folk theorem in the prisoner's dilemma with large groups and anonymous interaction, see Ellison (1994). The behavioral relevance of repetition and anonymity have been studied in many laboratory experiments, see, for example, Andreoni (1988) and Bohnet and Frey (1999). For a survey, see Ostrom (1998) who also discusses how experimental results relate to political science.

with given preferences and in the long run when preferences adapt to the new environment. Our analysis builds on an evolutionary approach.<sup>3</sup> We present analytical results and test their implications in a laboratory experiment.

A contractual relationship is represented by a game in which the first mover has to decide whether she wants to enter a contract without knowing whether the second mover will perform. If the second mover breaches, a chance move decides whether he is held liable for the cost of breach. Standard economic analysis of law predicts that the higher the expected cost of breach is, the more likely second movers are to perform. We show that, when preferences are subject to change, this need not be the case. More specifically, we find that the probability with which a contract is enforced has a *non-monotonic* impact on behavior: Second movers' performance rates are not only high when the expected cost of breach is sufficiently large but also when it is sufficiently small.

We focus on preferences for contract performance and assume that individuals may experience psychological costs when breaching a contract. Such trustworthy individuals are said to have a preference for honesty.<sup>4</sup> Based on the idea that a specific preference is more likely to be maintained and to flourish if it proves to be economically successful, we study a dynamic process in which preferences can change over time. Legal rules can “crowd in” as well as “crowd out” preferences. We find that intermediate levels of contract enforcement lead to crowding out while low levels induce crowding in. The intuition for this is rather

---

<sup>3</sup> See, for example, Becker (1976), Axelrod (1984), Boyd and Richerson (1985), Bendor and Swistak (1997), Bowles 1998, or Güth and Kliemt (1999).

<sup>4</sup> We purposely remain vague here because neither does our model depend on the specifics of the psychological costs incurred nor does our experiment test for different kinds of such costs. Our results are compatible with intrinsic motivation (see, for example, Frey 1997), inequality aversion (see, for example, Fehr and Schmidt 1999 or Bolton and Ockenfels 1999), reciprocity (see, for example, Rabin 1993) and psychological contracts (see, for example, Harrison and Robinson 1997 or Rousseau 1995).

straightforward. Suppose you are a first mover who has to decide whether to enter a contract. If you know that the legal system is inefficient, i.e., that contracts are rarely enforced, you will be extremely cautious. Clearly, you would like to enter if you knew your partner was trustworthy. If you received a signal about your partner's trustworthiness, you would only enter if the signal was "sufficiently good." This caution protects you from being exploited too often as well as making trustworthy second movers more successful than others, because on average, they will get more contracts than others. Hence, honesty will be crowded in.

Contractual relationships with weak enforcement are typical for many organizational settings. Some firms purposely create a low enforcement environment where interactions are not guided by the expected cost of breach but by intrinsic motivation. At the same time, these firms heavily invest in screening of potential future employees, stressing that character is more important than the possession of specific skills.<sup>5</sup> Similarly, most micro-finance institutions (e.g. Grameen Bank or Accion) lending money to poor clients without physical collateral, focus on "character-based lending". In the absence of external enforcement mechanisms, the intrinsic trustworthiness of their clients is one of the key variables making a contract between borrowers and lenders possible (Murdoch 1999).

The same pattern applies to many other domains: The more leeway agents have - whether these are employees, borrowers, legislators, judges or executives - the more careful principals are when deciding whom to offer a contract. That the leeway for politicians can be considerable becomes clear in Rose-Ackerman's (1999) analysis of corruption. She points out that (not even) in the United States the law is strict enough to deter elected officials from being

<sup>5</sup> See for a summary of such high-commitment human resource management practices, Baron and Kreps (1999).

corrupt. “The criminal penalties are ‘not more than three times the monetary equivalent of the thing in value (i.e. the bribe) or imprisonment for not more than fifteen years, or both’ (18 USC § 201 (a)). This is appropriate for officials who receive bribes except that multiplying by three may be a poor measure of the risk of detection and punishment. The actual probability of catch is likely to be well below one-third.” (p. 55) Whether this probability is low in enough to induce crowding-in is an empirical question. If the probability of contract enforcement is higher but not high enough to deter all second movers from breaching, we expect crowding-out.

With such medium enforcement, the expected payoff of entering is higher than the payoff of abstaining even if you know that the contract will be breached. Accordingly, you will enter regardless of your beliefs about your partner’s trustworthiness. This unconditional trust makes expected monetary payoff-maximizing second movers more successful than honest types who forsake profitable opportunities to breach. Therefore, honesty will be crowded out causing the aforesaid non-monotonicity: With high levels of contract enforceability, all second movers perform because they are deterred regardless of their preferences, and all first movers enter the contract; preferences are irrelevant and outcomes are efficient. With intermediate levels, honesty is crowded out; more second movers breach and resources are wasted in trials. With low levels, trustworthiness is crowded in; more second movers perform even though they would have an incentive to breach without a preference for honesty and efficiency increases.

Contract enforcement probabilities which are too small to deter breach may be due to a badly functioning legal system with weak state protection, corrupt governments and judicia-

ries, or high enforcement costs. So far, it has mainly been discussed how informal institutions, such as social norms, may substitute for an ineffective legal system, and whether shame and ostracism can replace imprisonment and fines.<sup>6</sup> Our model focuses on formal law and intrinsic dispositions and shows how the effectiveness of the two motivational factors depends on each other. By providing a specific legal enforcement regime, the state affects the degree of trust and trustworthiness in a society. Including formal institutions into the analysis of trust and trustworthiness addresses an aspect so far little touched in the current debate on trust and social capital.<sup>7</sup> Our findings may help understand two tendencies currently observed in many countries of the former Soviet bloc: On the one hand, there is a demand for “more law” in order to enforce contracts and to secure property rights. As the state is unable to provide levels of enforcement high enough to deter breach, the demand for protection is satisfied privately. This is one of the explanations for the rise of the Mafia in Sicily (Gambetta 1993) and may also account for its thriving in Russia (Varese 1994). On the other hand, a reemergence of the demand for and supply of trust-based relationships can be observed in the very same countries. In his chapter on “Economics of ethnic capital formation and conflict,” Wintrobe (1995: 46-47) writes: “The absence of enforceability generates a demand for trust. The costs of trust formation are lower, when the two parties share common traits, such as a common language, ethnicity, and so on.” While his analysis is different from ours, his conclusions

---

<sup>6</sup> Social norms have been shown to confine minor crimes such as trespassing (Ellickson 1991) or the overuse of common pool resources (Ostrom, Gardner, and Walker 1994). For the legal debate about “alternative sanctions” see Kahan (1996) and for a general discussion of how social norms and the law interact Sunstein (1996) and Cooter (1998).

<sup>7</sup> See Tarrow (1996: 395) who asks: “Can we be satisfied interpreting civic capacity as a home-grown product in which the state has played no role?” Schneider *et al.* (1997) are among the few who discuss the influence of institutions, namely of the extent parents can choose a public school, on social capital. Those interested in the influence of institutions mainly focus on the relationship between institutions and trust rather than between institutions and trustworthiness (see, e.g., Brehm and Rahn 1997, Nye, Zelikow, and King 1997 and Norris 1999).

are similar: More order can be achieved by relying on trust-based relationships where the parties are able to predict each other's likelihood of cooperation.

Our analysis provides a possible rationale for the coexistence of these two tendencies. People cannot help asking for “more law” if trustworthiness has been crowded out. On the other hand, if it has been crowded in, people can rely on trust-based interactions. Simis (1982) and Varese (1994) suggest that the former Soviet system was characterized by corruption rather than organized crime affecting only specific sections of the population, namely the “nomenclatura.” It is these groups who in the absence of a credible state are unable to engage in trust-based interactions and thus, demand private protection. On the other hand, for ordinary people, trust was and still is the basis of their contractual relationships (Wintrobe 1995). Overall, our model predicts a pattern also described by a Latin American quip: “A los amigos todo, a los enemigos nada, al extrano la ley.” (For friends everything, for enemies nothing, for the stranger the law.)<sup>8</sup>

The model's empirical validity is tested in the laboratory where we implement the theoretical setup as closely as possible. The experimental results support the crowding predictions. The remainder of the paper is organized as follows: In Section 2 we first study how agents behave with fixed preferences, then analyze the crowding dynamics and discuss the model's main implications. Section 3 presents the design and the experimental results. Section 4 concludes.

---

<sup>8</sup> We found this proverb in Rose-Ackerman (1999, p.97).



## 2 Theory

### 2.1 The contract game

We model a situation in which two players have the opportunity to produce a joint surplus. There is, however, an asymmetry between the players. Player 1 has to enter the contract without knowing whether player 2 will perform. Therefore, player 1's decision whether to enter the contract is a matter of trust.<sup>9</sup>

We denote her trusting move with  $T$  and her non-trusting move with  $\bar{T}$ . In the case where she trusts, player 2 can perform (move  $P$ ) or breach ( $\bar{P}$ ). The game ends if either player 1 does not enter, which yields zero-payoffs for both parties, or if player 1's trust is rewarded by player 2's cooperation. This yields payoffs of 1 for both players.<sup>10</sup> In case of breach, we assume that there is a final chance move capturing a litigation process. With probability  $p$  player 2 will be held liable (move  $L$ ). The surplus is divided as in the case of performance but player 2 has to bear the costs of the trial  $c > 0$ , i.e. we assume that perfect expectation damages place the trusting party in the position it would have been in if the other party had performed and that all legal fees are paid by the losing party.<sup>11</sup> Thus, the payoffs are 1 for player 1 and  $1 - c$  for player 2. If player 2 is not held liable ( $\bar{L}$ ) he profits from breach and receives a payoff of  $1 + b$  (with  $b > 0$ ) while player 1 suffers a loss. Her payoff is  $-a$  with  $a \geq c$  as she bears the legal cost and is not compensated for any investments she made by entering the contract.

---

<sup>9</sup> For related games on trust, see Berg, Dickhaut and McCabe (1995), Burnham, McCabe and Smith (1999), Gueth and Kliemt (1999), or Van Huyck, Battalio, and Walters (1995).

<sup>10</sup> This specification of the payoffs is a simple normalization without loss of generality.

<sup>11</sup> This corresponds to the English legal cost allocation rule.

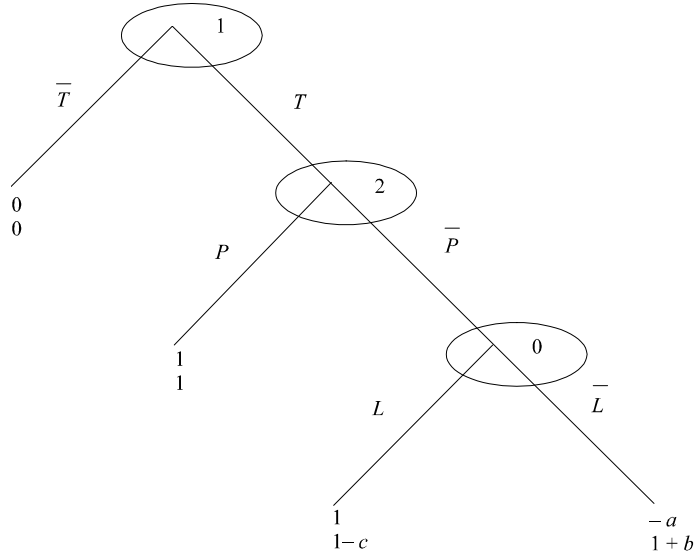


Figure 1: The contract game with monetary payoffs.

Breach is never efficient. The benefits from breach are not large enough to compensate the first mover, i.e.,  $b < 1 + a$ . Figure 1 shows this game in its extensive form.

We assume that all payoffs are monetary. However, in order to solve the game we need utilities associated with the various outcomes. To map outcomes into (cardinal) utilities we assume that there are two possible preferences a player can have, i.e., players can be of two different types. One type ( $M$ ) is only interested in (expected) *monetary* payoffs. This means that for this type the monetary payoffs in Figure 1 represent utilities. The second type ( $H$ ) is assumed to have a preference for *honesty*. In particular, we assume that this type suffers from psychological costs when breaching a contract. These costs are assumed to be sufficiently high such that a player of type  $H$  would never betray regardless of the monetary gain  $b$ .<sup>12</sup>

---

<sup>12</sup> This assumption simplifies the analysis without altering its results. All results would still hold if we allowed for a larger set of preferences incorporating different levels of psychological costs. Notice that in the case of a continuous type space—possibly ranging from infinite (psychological) costs to infinite gains—all that

Having introduced the set of possible preferences  $\{M, H\}$  we can transform the ‘money game’ of Figure 1 into a standard game where the payoffs are von Neumann–Morgenstern utilities. This is done by replacing the payoff of player 2 after path  $T\overline{P}L$  by  $1 + b - \delta$  where  $\delta = 0$  for type  $M$  and  $\delta > b$  for type  $H$ .

Assuming that player 1 recognizes which type player 2 is, we have, for each possible match of players, a well-specified standard game.<sup>13</sup> We solve this game by backward induction.<sup>14</sup>

Obviously, player 2 will breach if his expected payoff from breach exceeds 1, i.e. if

$$p(1 - c) + (1 - p)(1 + b - \delta) > 1. \quad (1)$$

As assumed, for type  $H$  this is never fulfilled. A player 2 with a preference for honesty will always choose  $P$ . For type  $M$ , we can insert  $\delta = 0$  into inequality 1 and rewrite it as

$$p < \frac{b}{b + c}. \quad (2)$$

Next, we can focus on the decision of player 1. If she is confronted with an  $H$ -type she will surely trust. The same is true if she is confronted with an  $M$ -type and  $p > \frac{b}{b+c}$  holds. But if she is confronted with an  $M$ -type and (2) is fulfilled she will enter only if her expected payoff from entering exceeds her outside-option payoff, i.e., if  $p - a(1 - p) > 0$ . This can be rearranged as

$$p > \frac{a}{1 + a}. \quad (3)$$

---

matters is whether the costs are larger or smaller than the monetary gain  $b$ .

<sup>13</sup> If player 2 is of type  $M$  the game is identical to the one of Figure 1. If player 2 is of type  $H$  the payoff after path  $T\overline{P}L$  is  $1 + b - \delta < 1 + b$ .

<sup>14</sup> Admittedly, the assumption of perfect type recognition is strong. However, the results of our analysis hold as long as players receive sufficiently informative signals about their opponents’ types. This is also discussed in section 2.3 and illustrated in Appendix B. The main effect of assuming perfect signals is that it simplifies notation and analysis significantly. Recent experimental evidence supports the notion that type recognition is possible after pre-play communication in a prisoners’ dilemma game (see Orbell and Dawes 1991, or Frank, Gilovich, and Regan 1993). Kikuchi, Watanabe and Yamagishi (1995) found that trustworthy types were better in predicting others’ types than non-trustworthy types. More generally, see Frank (1988).

From this follows

**Proposition 1** *The (unique) subgame perfect equilibrium (SPE) is*  
 $(T, P)$  if player 2 is of type H or if player 2 is of type M and  $p > \frac{b}{b+c}$  [High  $p$ ];  
 $(T, \bar{P})$  if player 2 is of type M and  $\frac{a}{1+a} < p < \frac{b}{b+c}$  [Medium  $p$ ];  
 $(\bar{T}, \bar{P})$  if player 2 is of type M and  $p < \min \left\{ \frac{a}{1+a}, \frac{b}{b+c} \right\}$  [Low  $p$ ].

In what follows we shall assume that  $ac < b$ . Otherwise  $(T, \bar{P})$  would never be an SPE and the crowding analysis would be less rich and less interesting. Imposing this requirement means, informally speaking, that the loss player 1 incurs after having been betrayed and not compensated must be relatively small in comparison to the profit of player 2 and the legal costs.

## 2.2 Crowding

The previous subsection showed how individuals with given preferences behave under different legal regimes. We now allow for preferences to adapt to the contractual situation and study the implications of “preference crowding” for our model. Economically successful preferences are crowded in, unsuccessful preferences are crowded out. Formally, this means that the share of types with a certain preference grows faster than another share, if and only if the average material earnings of the former exceed the average earnings of the latter. In the context of our model and in the absence of a fully fledged theory of preference formation, this assumption seems natural. Contracts are closed to secure material benefits and the outcomes of our contract game are exclusively characterized by different resource allocations.<sup>15</sup>

---

<sup>15</sup> For earlier studies analyzing endogenous preferences in a similar framework, see, for example, Bester and Güth (1998), Huck and Oechssler (1999), To (1999), and especially Güth and Kliemt (1999) who also deal with the issue of trust. Pioneering studies suggesting that endogenous preferences may result from payoff-driven (evolutionary) dynamics are Becker (1976) and Hirshleifer (1977). However, we do not wish to generalize too much here as different mechanisms may apply if one deals with attitudes towards drugs, religion or other domains in which outcomes are less tangible. For a survey of alternative approaches, see Bowles (1998).

The assumption that successful “traits” spread is often associated with models of genetic evolution, although it can be justified differently.<sup>16</sup> One justification is offered in Appendix A which briefly illustrates *a stochastic model of individual preference adaptation*.<sup>17</sup> In our model, this implies the following: If honesty leads to forsaking profitable opportunities such that a typical  $H$ -type earns less than a typical  $M$ -type, honesty will be crowded out. On the other hand, honesty will be crowded in if individuals are in an environment favoring honesty such that, on average,  $H$ -types earn more than  $M$ -types.

In order to calculate how successful the two different types (of preferences) are, we assume that individuals are randomly matched to play the contract game, and that there are enough individuals such that the law of large numbers can be applied. This allows us to take expected values as a measure of success.

Proposition 1 shows that what happens when two players interact depends on the value of  $p$ , the probability that the contract is enforced. As indicated in the proposition we can distinguish three regimes. For each regime we can now study how preferences are crowded in and out.

*High  $p$  ( $p > \frac{b}{b+c}$ )*

In this case, all individuals, regardless of type and matching, will play the SPE  $(T, P)$ . This means that individuals of both types receive the same expected monetary payoff. Thus, there will not be any crowding and regardless of the numbers of  $H$ - and  $M$ -types, the large

<sup>16</sup> See Börgers and Sarin (1997) and Schlag (1998) for two such justifications.

<sup>17</sup> The probabilities with which preferences are adapted are assumed to depend on two factors, economic success and conformity. The more an individual’s environment favors a certain preference and the more common a certain preference is, the more likely it is that an individual will develop the same preference. (Note, however, that adaptation is not deliberate. People cannot consciously decide about their preferences. For a different approach, see Cooter 1998.) We show that this process behaves, under certain additional regularity assumptions, like a growth-monotonic (evolutionary) process.

enforcement probability ensures performance.

$$\text{Medium } p \left( \frac{a}{1+a} < p < \frac{b}{b+c} \right)$$

With medium  $p$ , behavior in a match depends on the type of player 2. If player 2 is of type  $H$ , the SPE is  $(T, P)$ . If he is of type  $M$ , the SPE is  $(T, \bar{P})$ . Since  $M$  maximizes expected monetary payoffs while  $H$  does not, it follows that average earnings of  $M$ -types exceed average earnings of  $H$ -types. In the role of player 2,  $M$ -types earn on average  $p(1-c) + (1-p)(1+b)$ , which is strictly greater than 1, the payoff of an  $H$ -type in the role of player 2. When in the role of player 1,  $M$ -types and  $H$ -types do equally well on average, since the expected payoff of a player 1 is independent of her own type. Thus, with medium  $p$ ,  $M$ -types always earn more than  $H$ -types (regardless of the current number of them). Accordingly, honest preferences will be crowded out. This implies, asymptotically for the long run, that  $H$ -types will completely vanish and that all individuals will play equilibrium  $(T, \bar{P})$ .

$$\text{Low } p \left( p < \frac{a}{1+a} \right)$$

We proceed as in the above case, but now take into account that  $(\bar{T}, \bar{P})$  is the SPE if player 2 is of type  $M$  and that  $(T, P)$  is the SPE if player 2 is of type  $H$ . In this case, the earnings of  $H$ -types exceed the earnings of  $M$ -types. When in the role of player 2, the former always receive 1, the latter always 0. When in the role of player 1, the average payoffs of the two types are again identical. They do not depend on their own type. Thus, with low  $p$ , preferences for honesty are crowded in and, in the long run, type  $M$  will vanish such that all individuals will play the trust-rewarding equilibrium  $(T, P)$ .

We summarize our results in

**Proposition 2** *In the case where*

$\frac{a}{1+a} < p < \frac{b}{b+c}$	$M$	$H$
$M$	$1 + b - a + p(1 - b + a - c)$	$2 + b - p(b + c)$
$H$	$1 - a + p(1 + a)$	$2$

Table 1: The evolutionary game with medium  $p$ .

$p < \frac{a}{1+a}$	$M$	$H$
$M$	$0$	$1$
$H$	$1$	$2$

Table 2: The evolutionary game with low  $p$ .

$p > \frac{b}{b+c}$  there is no crowding; in the long run both types may be present in the society and all individuals play  $(T, P)$ ;

$\frac{a}{1+a} < p < \frac{b}{b+c}$  trustworthiness is crowded out; in the long run only type  $M$  will be present in the society and all individuals play  $(T, \bar{P})$ ;

$p < \frac{a}{1+a}$  trustworthiness is crowded in; in the long run only type  $H$  will be present in the society and all individuals play  $(T, P)$ .

The long-run stable states can also be derived by analyzing evolutionary games in which the two types compete. For  $p > \frac{b}{b+c}$  this game is trivial as both types behave identically and receive identical payoffs. For the other cases the two matrix games in Tables 1 and 2 emerge. (Note that the payoffs are based on the assumption that both types are equally likely to become player 1 or 2.)

As  $p < \frac{b}{b+c}$ , it is easy to see that the game has a unique evolutionarily stable strategy (see Maynard Smith 1982). There is a unique equilibrium  $(M, M)$  and the equilibrium is strict. Hence, the unique evolutionarily stable strategy is  $M$ . This mirrors the second result of Proposition 2: In the long run only  $M$ -types survive.

An even simpler matrix emerges in case of  $p < \frac{a}{1+a}$ . Clearly, the unique equilibrium of this game is  $(H, H)$  and as the equilibrium is strict, the unique evolutionarily stable strategy is  $H$ . This mirrors the third result of Proposition 2.

## 2.3 Discussion

There are two main implications of Proposition 2. The most fundamental one is that it is impossible to predict behavior in a group of agents playing the contract game without knowing their history. Individuals' preferences are subject to change, and outcomes in one round affect the distribution of types in the next. And, since individuals' preferences depend on past regimes, so do their actions. The longer a group of agents has played in a low-probability environment, the more individuals with a preference for honesty are present and the less breach is observed. Vice-versa for a medium- $p$  environment. Also it may be possible that groups have experienced regime changes in the past in which case one does not only need to know how long the group has been playing under the current regime, but also how long under the preceding one. The total crowding history matters.

The second significant implication of the proposition is that the probability of contract enforcement exerts a non-monotonic effect on behavior. Such a non-monotonicity would not occur in standard models with fixed preferences.<sup>18</sup> The worst legal regime is not one where contracts cannot be enforced but one with an intermediate level of contract enforceability. With an intermediate  $p$  there is no need for first movers to care about with whom they interact, since entering the contract is better than staying with the outside option, even if the contract is breached. This lack of caution makes those who are dishonest economically successful. Accordingly, the share of dishonest types will grow. This is the situation which gave the present study its title: There are two alternatives—“more order with more law” or

---

<sup>18</sup> The first authors highlighting the possibility of such non-monotonicities are, to the best of our knowledge, Akerlof and Dickens (1982). They study a model with cognitive dissonance which may induce players to re-evaluate outcomes.



“more order with less law”.

With less law, first movers have to be extremely cautious. They have to think about their partners’ trustworthiness which makes honesty a successful preference. In our model, first movers receive a perfect signal about their opponents’ type and their decision rule is simple: “Only enter a contract if your opponent is trustworthy.” With a stochastic signal the rule would be very similar: “Only enter if the signal is good enough.” This illustrates how our results would extend to the more general case of imperfect but informative signals. With perfect signals,  $M$ -types in the role of player 2 are *never* offered contracts when  $p$  is low while  $H$ -types *always* get contracts. With imperfect signals, some  $M$ -types would get contracts while some  $H$ -types would not (namely whenever the signal is wrong). However, if the signal is sufficiently informative  $H$ -types will get *more* contracts than  $M$ -types which is required for crowding in. Appendix B elaborates on the case of imperfect signals further.

Comparing the two alternative policies for replacing a medium probability regime (“more or less law”), two differences can be observed. The first one is due to the dynamic nature of our analysis. With “more law”, more order is instantaneously achieved since performance becomes rational for everybody and so does entering. This is different in the case of “less law” because after the change of the regime the crowding process needs some time. Though our experimental results below indicate that adjustments can be pretty fast in small groups, the behavior does not change instantaneously. This is an argument in favor of the standard law-and-order approach. However, “less law” is less costly. In our model, we disregard all fixed costs of legal contract enforcement and also variable costs being a function of  $p$ . Increasing  $p$  costs resources. Decreasing  $p$  saves resources.

Here we do not wish to make any judgement about what is the better policy. Instead we test our model’s empirical validity in the laboratory.

### 3 Experiment

#### 3.1 Design

The experimental design tries to implement our model as closely as possible. Subjects participated in a 2–person contract game and were randomly matched. Six sessions with altogether 154 subjects were conducted and the game was repeated several times. Focussing on crowding in (“more order with less law”), subjects in all sessions were confronted with a low contract enforcement probability during the last rounds. In order to create different histories, the legal regime in the first few rounds was varied. If behavior was driven by incentives only, behavior should be independent of the history created in earlier rounds and should only depend on the current legal regime. If preferences adapt to the legal regime, on the other hand, the history of earlier rounds should also affect the likelihood of performance in later rounds.

Experimental subjects were confronted with identical payoff tables and the payoffs were

- 50 cents for each player whenever the first mover chose  $\bar{T}$  ;
- 150 cents for each player in case of  $TP$ ;
- 150 cents for the first and 120 cents for the second mover in case of  $T\bar{P}L$ ;
- 20 cents for the first and 250 cents for the second mover in case of  $T\bar{P}\bar{L}$ .

Thus, the normalized payoff variables were  $a = .3$ ,  $b = 1$ , and  $c = .3$ . The probabilities

were .1, .5, and .9.<sup>19</sup>

Subjects received payments for each round. Instructions (see Appendix C) were neutrally framed. After each round, aggregate information on outcomes was provided, i.e. first and second movers knew how many contracts were offered and performed in the previous round. Providing information on the distribution of types only serves as a conservative test of our model, as individuals' types could not perfectly be detected. As the crowding model assumes random matching, we implemented a stranger treatment in five sessions and used a fixed-pair matching in one control session. Table 3 presents an overview of all sessions.

The experiments were conducted at two different universities (labelled B and H).<sup>20</sup> Participation was voluntary; interested students could sign up for an experimental session and were paid a show-up fee of \$5. The experiments took approximately 45 minutes. Average earnings were approximately \$15.

Subjects were identified by code numbers only and anonymity was fully preserved. After having signed a consent form, participants were randomly assigned to the roles of first and second mover and given a written instruction and an envelope containing a code number sheet and nine decision sheets, all marked with the subject's code number. Instructions were repeated orally, allowing subjects to ask questions and to control that everybody faced the same decision task. In all but session 4, they were truthfully assured that they would be randomly matched with a different person after each round. They were told that nine rounds would be played and that they would publicly be informed on the aggregate outcome after

---

<sup>19</sup> Subjects carried out the chance moves themselves. After each round a randomly chosen participant picked a card from a pile of red and black cards.

<sup>20</sup> This labelling follows the APSR guide lines for double-blind reviewing.

Session	Matching	Prob. 1–3	Prob. 4–9	# subjects	Univ.
1 (RLLB)	Random	Low ( $p = .1$ )	Low ( $p = .1$ )	20	B
2 (RMLB)	Random	Medium ( $p = .5$ )	Low ( $p = .1$ )	28	B
3 (RHLB)	Random	High ( $p = .9$ )	Low ( $p = .1$ )	28	B
4 (FMLB)	Fixed	Medium ( $p = .5$ )	Low ( $p = .1$ )	16	B
5 (RLLH)	Random	Low ( $p = .1$ )	Low ( $p = .1$ )	34	H
6 (RMLH)	Random	Medium ( $p = .5$ )	Low ( $p = .1$ )	28	H

Table 3: Experimental sessions.

each round. Individual results were private information. Subjects could not anticipate the regime change in round 4.<sup>21</sup> However, when they were informed about the new conditions, they were also told that they would play under the new regime for the remaining six rounds.

### 3.2 Results

Our theory predicts history-dependent behavior. The longer individuals are confronted with a low- $p$  environment, the more trustworthiness should be crowded in, and the higher performance rates should be. High enforcement probabilities are expected to be crowding-neutral, while trustworthiness should be crowded out with medium probabilities.<sup>22</sup>

Appendix D presents the results for all sessions in all rounds. First, we briefly examine session 4 with fixed pairs. This is a simple control session as the requirements for crowding are not fulfilled. There is no random matching and therefore no interaction on the group level. Instead, subjects play a finitely repeated game. Experiments on other games with cooperative gains (e.g., repeated public goods games or gift exchange games) reveal that with this kind of matching, cooperation rates are typically higher than standard theory expects. Furthermore, cooperation rates seem to be relatively stable over time, but break

<sup>21</sup> The instructions neither told them that the environment would remain constant nor that it would change.

<sup>22</sup> However, these predictions cannot be viewed as deterministic since the law of large numbers does not apply in the lab (see also Appendix A).

down towards the end of the game when the “shadow of the future” loses its power and reputation no longer plays a role. This strong drop in the last rounds has been called an “end game effect” (Selten and Stoecker 1983).<sup>23</sup> If the contract game is comparable to these games, we should observe a similar pattern. Inspection of Appendix D confirms this expectation: We find that in the fixed-pairs session cooperation drops from 100% to 0% in the last round.

In contrast, our crowding theory predicts increasing cooperation over time and rules out an end-game effect. After all, it predicts that trustworthy second movers perform because they receive less utility from breaching than from performing, even though breaching leads to a higher monetary payoff. Table 4 shows aggregate data for rounds 4 to 9 in all random-matching sessions. It suggests that there is a trend towards more cooperation which does not break down. On the contrary, the performance rate (the number of contracts performed divided by the number of contracts offered) reaches its maximum in the last round which is in line with the crowding prediction. We summarize by

**Result 1** In the low-probability environment with random matching, performance rates increase over time and there is no end-game effect.

Our model assumes that it is most efficient if all second movers perform. We expect high efficiency rates instantaneously when enforcement is strong, and only slow increases in efficiency rates over time with low enforcement probabilities. Figure 2 presents the efficiency rates (the number of contracts performed divided by the number of all possible contracts), the percentage of performed contracts, for the sessions with random matching. In the short

---

<sup>23</sup> For additional evidence, see Andreoni (1988) and Croson (1996) who differentiate between cooperation based on reputation and reciprocity.

Aggregate	Round					
Data	4	5	6	7	8	9
<i>No Trust</i> ( $\bar{T}$ )	40	38	29	34	27	35
<i>Trust &amp; Breach</i> ( $T, \bar{P}$ )	17	12	16	11	14	8
<i>Trust &amp; Perf.</i> ( $T, P$ )	12	19	24	24	28	26
<i>Performance Rate</i>	.41	.61	.63	.69	.67	.77

Table 4: Aggregate random-matching data for rounds 4 to 9

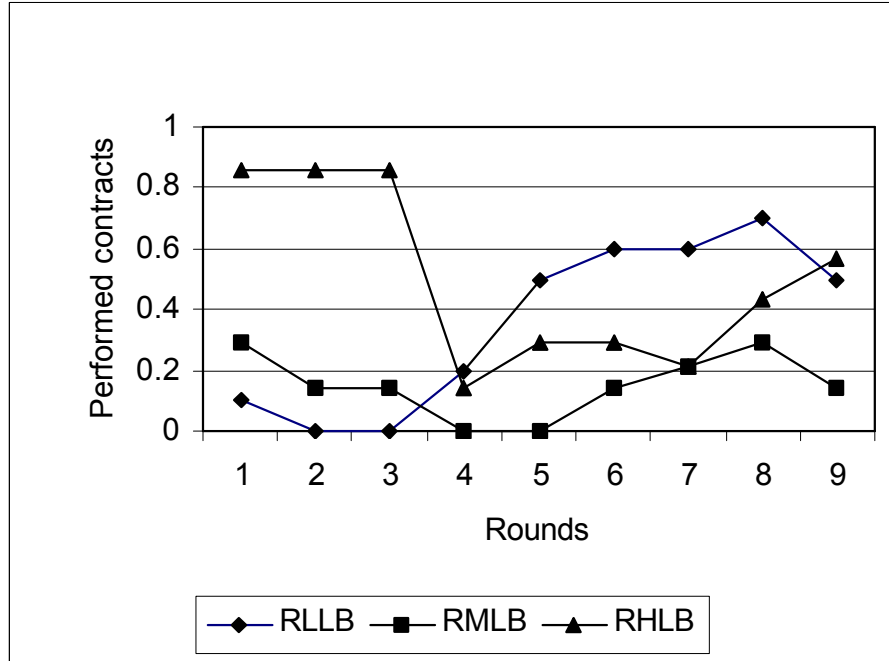


Figure 2: Efficiency rates over time (Berkeley only).

run, high enforcement probabilities lead to the most efficient outcomes. In the medium term, when the crowding dynamics start to become relevant, the low- $p$  environment is most efficient. The longer subjects have been confronted with a low- $p$  environment, the less the differential effects of their respective crowding histories apply.

**Result 2** In the short run, efficiency rates are highest when enforcement is strong; in the medium term, they are highest in the low-probability environment; in the long run,

the differential effects of enforcement and crowding tend to vanish.

In order to analyze the data more thoroughly, we next estimate binary choice models for first movers' propensity to enter and second movers' propensity to perform, and control for the relevance of crowding compared to economic incentives and for university fixed effects.

In order to measure crowding let

$$\gamma_j^t = \begin{cases} 1 & \text{if } p \text{ is small in group } j \text{ in round } t \\ 0 & \text{if } p \text{ is high in group } j \text{ in round } t \\ -1 & \text{if } p \text{ is medium in group } j \text{ in round } t \end{cases}$$

and let

$$CROWD_j^t = \sum_{h=1}^{t-1} \gamma_j^h.$$

The variable  $\gamma_j^t$  indicates whether our theory predicts crowding in (+1), crowding out (-1), or no crowding (0) and the variable  $CROWD_j^t$  summarizes the “crowding history” of group  $j$  up to round  $t$ . If the theory is relevant we would expect that  $CROWD_j^t$  helps explain the propensity of second movers to perform.<sup>24</sup>

Besides  $CROWD$  we also include the following variables as covariates:

- $PERFORM_j^{t-1}$ —the performance rate in group  $j$  in round  $t - 1$ , measured as the number of contracts performed divided by the number of contracts offered;
- $ENTER_j^{t-1}$ —the rate of entering in group  $j$  in round  $t - 1$ , measured as the number of contracts entered divided by the number of first movers in group  $j$ ;

---

<sup>24</sup> One could argue that this definition is somewhat arbitrary and we agree. By no means, do we claim that  $CROWD$  captures the “true story”, and certainly we do not claim that crowding is linear. Probably it is not. However, we think that this a “Bayesian approach”. In the absence of any specification which is *a priori* more natural than another, the above definition is the most simple one and can be justified by taking expected values over equally probable alternatives. Furthermore, if we find a significant effect of  $CROWD$  in its current form, any “optimal” definition would increase its significance.

- $INCENT_j^t$ —a dummy variable indicating whether the first mover has an incentive to enter the contract if only monetary payoffs play a role, i.e.  $INCENT_j^t = 1$  if  $p$  is medium or high and  $INCENT_j^t = 0$  otherwise.
- $INCPERF_j^t$ —a dummy variable indicating whether the second mover has an incentive to perform the contract if only monetary payoffs play a role, i.e.  $INCPERF_j^t = 1$  if  $p$  is high and  $INCPERF_j^t = 0$  otherwise.
- $UNIV_j$ —a dummy variable indicating the university group  $j$  belonged to;  $UNIV_j = 1$  for H and  $UNIV_j = 0$  for B.

Furthermore, we also run (logistic) regressions including subject dummies  $S_{ji}$ .<sup>25</sup>

The estimated model without subject dummies is

$$\begin{aligned} \ln \frac{q_{rji}^t}{1 - q_{rji}^t} = & \alpha + \beta_0 CROWD_j^t \\ & + \beta_1 PERFORM_j^{t-1} + \beta_2 ENTER_j^{t-1} \\ & + \beta_3 INCENT_j^t + \beta_4 INCPERF_j^t \\ & + \beta_5 UNIV_j + \varepsilon_{ji} \end{aligned}$$

with  $q_{rji}^t$  indicating either first mover  $i$ 's propensity to enter or second mover  $i$ 's to perform.

Table 5 presents the results for first movers' decisions, Table 6 for second movers' decisions.

All sessions with random matching are included.

We find that first movers' behavior is mainly driven by the economic incentives they face and by the performance rate of the last round. In the estimation without subject dummies

---

<sup>25</sup> We use nested effect coding (see Königstein 1998) as the subject dummies are nested in the university dummy.



<b>First movers</b>						
	without subject dummies			with subject dummies		
	Coeff.	Sign.	$R$	Coeff.	Sign.	$R$
$CROWD_j^t$	.048	.175	.000	-.103	.189	.000
$PERFORM_j^{t-1}$	<b>2.062</b>	.000	.160	<b>5.441</b>	.000	.218
$ENTER_j^{t-1}$	.530	.247	.000	-.263	.729	.000
$INCENT_j^t$	<b>2.553</b>	.000	.173	<b>7.955</b>	.000	.185
$INCPERF_j^t$	4.649	.684	.000	5.914	.856	.000
$UNIV_j$	-.518	.015	-.072	-1.838	.724	.000
Const.	-1.296	.000		-1.999	.514	

Table 5: Logistic regression for first movers' propensity to enter (552 observations).  $R$  is the *partial contribution* measuring the relative importance of each variable.

<b>Second movers</b>						
	without subject dummies			with subject dummies		
	Coeff.	Sign.	$R$	Coeff.	Sign.	$R$
$CROWD_j^t$	<b>.125</b>	.030	.077	<b>.792</b>	.000	.183
$PERFORM_j^{t-1}$	<b>1.493</b>	.034	.080	-.959	.566	.000
$ENTER_j^{t-1}$	-.864	.223	.000	-1.731	.319	.000
$INCENT_j^t$	-.177	.743	.000	-1.454	.303	.000
$INCPERF_j^t$	<b>2.057</b>	.006	.115	15.225	.708	.000
$UNIV_j$	<b>.594</b>	.034	.077	2.588	.915	.000
Const.	-.505	.365		-.484	.972	

Table 6: Logistic regression for second movers' propensity to perform (312 observations).  $R$  is the *partial contribution* measuring the relative importance of each variable.

we also find a significant effect of the university dummy: on average, H students enter less often than B students. Since the university dummy loses its significance when subject dummies are introduced we are confident that the difference between H and B is not due to some differences in the experimental procedures (of which we were not aware). Rather, the difference is on the individual level.

The same is true for second movers where we find that H students are more trustworthy than B students. Again, this effect disappears when we control for differences between individuals. In the estimation without subject dummies we find three additional variables which help explain second movers' decisions: economic incentives (less breaching if breach is deterred), last round's performance rate (such that there is some inertia), and finally, the variable capturing the crowding history of groups (the longer subjects interacted in a low- $p$  environment, the more likely they are to be trustworthy). The latter variable is the only one which remains significant when we include subject dummies. We summarize by

**Result 3** First movers' propensity to enter mainly depends on their monetary incentives and on second movers' previous performance rate. Second movers' propensity to perform mainly depends on their crowding history.

The second part of Result 3 is the key result of this study as it confirms the qualitative predictions of the crowding theory. It is, however, a result on the aggregate level, and it seems worthwhile to investigate whether the theory also predicts individual behavior well. The main prediction of our model on the individual level is that in the low- $p$  environment, subjects who have breached contracts once are much more likely to switch to honest behavior than vice-versa. In order to analyze this hypothesis we simply count how many second movers

Session	<i>breach</i> $\rightarrow$ <i>perform</i>	<i>unpredicted switches</i>	<i>behavior fixed</i>	<i>N</i>
1 (RLLB)	7	–	3	10
2 (RMLB)	6	1	7	14
3 (RHLB)	9	–	5	14
4 (RLLH)	6	3	8	17
5 (RMLH)	5	–	9	14
Total	33	4	32	69

Table 7: Individual behavior in the low-probability environment.

switch from breaching to performing once  $p$  is low. Additionally, we count how many second movers switch in unpredicted directions and how many never switch once  $p$  is low. Table 7 presents the results. The picture is clear: 53.6% of all subjects changed their behavior in the low-probability environment. And of these, 89.2% switched in the direction which is predicted. In other words, 33 second movers breached in the early phases of the experiment and then performed later, even though the money-maximizing strategy would have been to stick to the old behavior. None of these 33 subjects switched back to breaching. They became trustworthy, even though they started by breaching. On the other hand, there are only very few unpredicted switches.<sup>26</sup> The null-hypothesis that the switching of behavior is random can be rejected at a significance level of .01%. Thus, the crowding theory’s predictions are also confirmed by analyzing individual data.

## 4 Conclusion

We have studied a model where the (legal) rules of a game do not only have short-run but also long-run effects on behavior because they affect preferences. In a contractual relationship, economic incentives have a non-monotonic impact on contract performance. Our theoretical model complements recent approaches studying how rules and preferences interact (see, for

<sup>26</sup> Notice that, as the theory is stochastic, these instances are not entirely unpredicted.

example, Fehr and Schmidt, 1999) as the latter only allow for differences in preferences but not for preferences to change. Our model suggests that the rules of the game determine which preferences dominate. More specifically, it predicts that low levels of legal contract enforcement crowd in trustworthiness. As first movers cannot trust the legal system, they only enter a contract if they can trust the second movers. They are careful when deciding whether to enter a contract, thus making trustworthiness a successful trait.

Arguing from a different perspective, others come to very similar conclusions. Mansbridge (1999), e.g., discusses various ways of encouraging trustworthiness and trust and concludes: “When the trustworthiness of a population is too low to sustain a general stance of initial trust, and when geographic and social mobility make reputational, kin and local sanctions less viable, the trustworthy members of a given population will benefit from finding ways of distinguishing themselves and other trustworthy individuals from the untrustworthy (...) the trustworthy would find it useful to train themselves to recognize subtle signs of trustworthiness in others and also to develop in themselves signs that could not easily be mimicked.” (p. 305). Differences in incentives to learn about others’ dispositions may account for some of the cross-cultural differences in behavior found in laboratory experiments. Yamagishi, Cook, and Watanabe (1998), e.g., argue that Japanese are less trusting and less trustworthy than American subjects because contract enforcement mechanisms and assurance structures are more prevalent in Japan than in the United States.

With high levels of contract enforcement, when contracts are completely specified, interpersonal trust is replaced by institutional trust in the legal system. First movers enter a contract because second movers are deterred from breaching.

While previous work on crowding focused on the relevance of preferences when contracts are complete,<sup>27</sup> we show that it is not the complete contract which crowds out trustworthiness but the semi-specified contract. With intermediate levels of contract enforcement, second movers are not yet deterred from breaching, and for first movers entering is financially more attractive than the outside option. Interpersonal trust is replaced by institutional trust in the legal system and genuine trustworthiness is crowded out. Such semi-specified contracts cause a non-monotonicity of behavior: More order can result from less law, yielding a “motivation-compatible” environment (Bohnet and Frey 1997), as well as from more law yielding an incentive-compatible environment.

Closely related to this finding are results by Huang and Wu (1994) and Huck (1998). Using psychological game theory<sup>28</sup> and modeling games very similar to ours, Huang and Wu (1994) show that if players’ payoffs also depend on beliefs, different levels of order may result from the same level of law; i.e., there is a multiplicity of equilibria. Simply speaking, there is one equilibrium in which everybody believes that the society is well-functioning and that trust is rewarded, and this becomes self-fulfilling. If everybody believes the opposite is true, then it also becomes self-fulfilling. A crucial difference between Huang and Wu’s approach and ours is that preferences (over payoffs and beliefs) are fixed in their model. Thus, there are also no obvious dynamics leading from one state to another and there is no straightforward link between institutional design and behavior. This is different in Huck

---

<sup>27</sup> See for policy examples, the early research by Titmuss (1970), Frey (1997) for the crowding out of tax morale, Frey and Oberholzer-Gee (1997) for the crowding out of civic duty in a siting context and Gneezy and Rustichini (1999) in every day life. For experimental evidence, see the extensive survey of studies in psychology by Deci, Koestner, and Ryan (1999) as well Falk, Gächter and Kovacs (1999). For the disruption of “implicit agreements” in the organizational context, see Arrow (1974).

<sup>28</sup> See Geneakoplos, Pearce, and Stacchetti (1989).

(1998) who shows in the context of criminal law that if preferences are allowed to change, socially desirable behavior can be induced with lower levels of (monetary) punishments than one would conclude assuming fixed preferences.

The implications of our model were tested in an experiment in which we tried to map the theoretical assumptions into a laboratory environment as precisely as possible. The results support our qualitative predictions. If there is enough time for the crowding dynamics to unfold, low enforcement–probability environments can produce outcomes as efficient as high levels of contract enforcement.

To the best of our knowledge, our results provide the first empirical evidence for long–run effects of legal rules on behavior. While experiments simplify reality, we tried to weaken the trade–off between external and internal validity as much as we could. Our design of the contract game is informed by real life institutions and represents a situation where legal enforcement leads to perfect expectation damages, and transaction costs are allocated according to the English rule. The “long run” in our experiment is nine rounds (or 45 minutes) which we interpret as a conservative test of crowding. In fact, we were surprised that the crowding dynamics unfolded as quickly and that our subjects’ inclination to trust and to be trustworthy changed in such a short time span. Our experimental results, thus, support the institutionalists’ view that institutional changes affect behavior. However, the paper tried to show that, by affecting behavior, institutions also affect preferences.

## References

- Akerlof, G.A. (1997): Social distance and social decisions, *Econometrica* 65, 1005-1027.
- Akerlof, G.A., and W.T. Dickens (1982): The economic consequences of cognitive dissonance, *American Economic Review* 72, 307-319.

- Andreoni, J. (1988): Why free ride? Strategies and learning in public goods experiments, *Journal of Public Economics* 37, 291-304.
- Arrow, K. (1974): *The Limits of Organization*, New York: Norton.
- Axelrod, R. (1984): *The Evolution of Cooperation*, New York: Basic Books.
- Baron, J.N., and D. Kreps (1999): *Strategic Human Resources*, New York: Wiley.
- Becker, G.S. (1976): Altruism, egoism, and genetic fitness: Economics and sociobiology, *Journal of Economic Literature* 14, 817-826.
- Bendor, J., and P. Swistak (1997): The evolutionary stability of cooperation, *American Political Science Review* 91, 290-307.
- Berg, J., J. Dickhaut, and K.A. McCabe (1995): Trust, reciprocity, and social history, *Games and Economic Behavior* 10, 122-142.
- Bernheim, B.D. (1994): A theory of conformity, *Journal of Political Economy* 102, 841-877.
- Bester, H., and W. Güth (1998): Is altruism evolutionarily stable?, *Journal of Economic Behavior & Organization* 34, 193-209.
- Bohnet, I., and B.S. Frey (1997): Rent leaving, *Journal of Institutional and Theoretical Economics* 153, 711-721.
- Bohnet, I., and B.S. Frey (1999): Social distance and other-regarding behavior in dictator games: Comment, *American Economic Review* 89, 335-339.
- Bolton, G.E., and A. Ockenfels (1999): ERC: A theory of equity, reciprocity, and competition, *American Economic Review*, forthcoming.
- Börgers, T., and R. Sarin (1997): Learning through reinforcement and replicator dynamics, *Journal of Economic Theory* 77, 1-14.
- Bowles, S. (1998): Endogenous preferences: The cultural consequences of markets and other economic institutions, *Journal of Economic Literature* 36, 75-111.
- Bowles, S. (1999): Cultural evolution and human social structure, *mimeo*, University of Massachusetts.
- Boyd, R., and P.J. Richerson (1985): *Culture and the Evolutionary Process*, Chicago and London: University of Chicago Press.
- Brehm, J., and W. Rahn (1997): Individual-level evidence for the causes and consequences of social capital, *American Journal of Political Science* 41, 999-1023.
- Burnham, T., K. McCabe, and V. Smith (1999): Friend-or-foe intentionality priming in an extensive form trust game, *mimeo*, Harvard University and University of Arizona.

- Cooter, R. (1998): Expressive law and economics, *Journal of Legal Studies* 27, 585-608.
- Croson, R. (1996): Partners and strangers revisited, *Economics Letters* 53, 25-32.
- Deci, E.L., R. Koestner, and R. Ryan (1999): Extrinsic rewards and intrinsic motivation: A clear and consistent picture after all, *Psychological Bulletin*, forthcoming.
- Ellickson, R.C. (1991): *Order Without Law: How Neighbors Settle Disputes*, Cambridge/Mass.: Harvard University Press.
- Ellison, G. (1994): Cooperation in the prisoner's dilemma with anonymous random matching, *Review of Economic Studies* 61, 567-588.
- Fehr, E., and K. Schmidt (1999): A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics* 114, 817-868.
- Frank, R.H. (1988): *Passions within Reason: The Strategic Role of the Emotions*, New York: Norton.
- Frank, R.H., T. Gilovich, and D.T. Regan (1993): The evolution of one-shot cooperation: An experiment, *Ethology and Sociobiology* 14, 247-256.
- Frey, B.S. (1997): A constitution for knaves crowds out civic virtues, *Economic Journal* 107, 1043-1053.
- Frey, B.S., and F. Oberholzer-Gee (1997): The costs of price incentives: An empirical analysis of motivation crowding out, *American Economic Review* 87, 747-755.
- Fukuyama, F. (1995): *Trust: Social Virtues and the Creation of Prosperity*, New York: Free Press.
- Gambetta, D. (1993). *The Sicilian Mafia*, Cambridge: Harvard University Press.
- Geneakoplos, J., D. Pearce, and E. Stacchetti (1989): Psychological games and sequential rationality, *Games and Economic Behavior* 1, 60-79.
- Gneezy, U., and A. Rustichini (1999): A fine is a price, *Journal of Legal Studies*, forthcoming.
- Greif, A., P. Milgrom, and B.R. Weingast (1994): Coordination, commitment, and enforcement: The case of the merchant guild, *Journal of Political Economy* 102, 745-776.
- Güth, W., and H. Kliemt (1999): Evolutionarily stable co-operative commitments, *Theory and Decision*, forthcoming.
- Hirshleifer, J. (1977): Economics from a biological viewpoint, *Journal of Law and Economics* 20, 1-52.



- Huang, P.H., and H.-M. Wu (1994): More order without more law: A theory of social norms and organizational cultures, *Journal of Law, Economics, & Organization* 10, 390-406.
- Huck, S. (1998): Trust, treason, and trials: An example of how the evolution of preferences can be driven by legal institutions, *Journal of Law, Economics, & Organization* 14, 44-60.
- Huck, S., and J. Oechssler (1999): The indirect evolutionary approach to explaining fair allocations, *Games and Economic Behavior* 28, 13-24.
- Kahan, D.M. (1996): What do alternative sanctions mean?, *University of Chicago Law Review* 63, 591-653.
- Kikuchi, M., Y. Watanabe, and T. Yamagishi (1996): Accuracy in the prediction of others' trustworthiness and general trust: An experimental study, *Japanese Journal of Experimental Social Psychology* 37, 23-36.
- Knack, S., and P. Keefer (1997): Does social capital have an economic payoff? A cross-country investigation, *Quarterly Journal of Economics* 112, 1251-1288.
- Königstein, M. (1998): Measuring treatment-effects in experimental cross-sectional time series, *mimeo*, Humboldt University, Berlin.
- La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R.W. Vishny (1997): Trust in large organizations, *American Economic Review* 87, 333-338.
- Mansbridge, J. (1999): Altruistic trust, in: Warren, M.E. (ed.): *Democracy and Trust*, Cambridge: Cambridge University Press, 290-309.
- Milgrom, P., D. North, and B. Weingast (1990): The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs, *Economics and Politics* 2, 1-23.
- Morrison, E.W., and S.L. Robinson (1997): When employers feel betrayed, *Academy of Management Review* 22, 226-256.
- Murdoch, J. (1999): The microfinance promise, *Journal of Economic Literature* 37, 1569-1615.
- Norris, P. (ed.) (1999): *Critical Citizens: Global Support for Democratic Governance*, Oxford: Oxford University Press.
- Nye, J.S., P.D. Zelikow, and D. King (eds.) (1997): *Why People Don't Trust Government*, Cambridge: Harvard University Press.
- Oostveen T., R. Knibbe, and H. DeVries (1996): Social influences on young adults' alcohol consumption: Norms, modeling, pressure, socializing, and conformity, *Addictive Behaviors* 21, 187-197.

- Orbell, J., and R. Dawes (1991): A “cognitive miser” theory of cooperators’ advantage, *American Political Science Review* 85, 515-528.
- Ostrom, E. (1998): A behavioral approach to the rational choice theory of collective action: Presidential address, *American Political Science Review* 92, 1-22.
- Ostrom E., R. Gardner, and J. Walker (1994): *Rules, Games, and Common-Pool Resources*, Ann Arbor: University of Michigan Press.
- Putnam, R.D. (with R. Leonardi and R.Y. Nanetti) (1993): *Making Democracy Work*, Princeton: Princeton University Press.
- Rabin, M. (1993): Incorporating fairness into game theory and economics, *American Economic Review* 83, 1281-1302.
- Rose-Ackerman, S. (1999): *Corruption and Government: Causes, Consequences and Reform*, Cambridge: Cambridge University Press.
- Rousseau, D.M. (1995): *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*, Thousand Oaks: Sage.
- Schlag, K.H. (1998): Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits, *Journal of Economic Theory* 78, 130-56.
- Schneider, M., P. Teske, M. Marschall, M. Mintrom, and C. Roch (1997): Institutional arrangements and the creation of social capital: The effects of public school choice, *American Political Science Review* 91, 82-93.
- Selten, R., and R. Stoecker (1983): End behavior in finite prisoner’s dilemma supergames, *Journal of Economic Behavior and Organization* 7, 47-70.
- Simis, K. (1982): *USSR: The Corrupt Society*, New York: Simon and Schuster.
- Smith, J.M., and P.A. Bell (1994): Conformity as a determinant of behavior in resource dilemma, *Journal of Social Psychology* 134, 191-200.
- Sunstein, C.R. (1996): Social norms and social roles, *Columbia Law Review* 96, 903-968.
- Tarrow, S. (1996): Making social science work across space and time: A critical reflection on Robert Putnam’s Making Democracy Work, *American Political Science Review* 90, 389-397.
- Titmuss, R.M. (1970): *The Gift Relationship*, London: Unwin.
- To, T. (1999): Risk and evolution, *Economic Theory* 13, 329-343.
- Urberg, K.A. (1992): Locus of peer influence: Social crowd and best friend, *Journal of Youth and Adolescence* 21, 439-450.

Van Huyck, J.B., R.C. Battalio, and M.F. Walters (1995): Commitment versus discretion in the peasant–dictator game, *Games and Economic Behavior* 10, 143-170.

Varese, F. (1994): Is Sicily the future of Russia? Private protection and the rise of the Russian Mafia, *Archives Européennes de Sociologie* 35, 224-258.

Wintrobe, R. (1995): Some economics of ethnic capital formation and conflict, in: Breton, A., G. Galeotti, P. Salmon, and R. Wintrobe (eds.): *Nationalism and Rationality*, Cambridge: Cambridge University Press, 43-70.

Yamagishi, T., K.S. Cook, and M. Watanabe (1998): Uncertainty, trust and commitment formation in the United States and Japan, *American Journal of Sociology* 104, 165-194.

## A A stochastic model of individual preference adaptation

Consider a large population of individuals who may be heterogenous with respect to their preferences. Let  $\Omega$  be the finite set of possible preferences and  $\omega$  a typical element of this set. When individuals interact, their behavior and earnings depend on the exact situation they face and on their types. For a given situation (e.g. a game which specifies only monetary payoffs) we can denote the monetary payoff of individual  $i$  as  $\pi_i(\omega_i, \omega_{-i})$  where  $\omega_{-i}$  denotes the types of all individuals  $i$  is interacting with. (For situations with multiple equilibria, this implies that there is a ready selection criterion at hand.) Individuals are matched by a matching scheme  $S$ , i.e.,  $S$  maps the set of individuals itself. The average payoff individuals of a certain type  $\omega'$  earn depends on  $\omega'$ ,  $S$ , and the current profile of types, denoted by the cumulative distribution function  $F(\omega)$ . It can be written as  $\Pi(\omega', S, F(\omega))$  or as  $\Pi(\omega')$ . For convenience, we restrict ourselves to discrete time,  $t$  being the time index. Accordingly, let  $f^t(\omega)$  denote the share of individuals who are of type  $\omega$  at time  $t$ . The expression  $\frac{f^{t+1}(\omega) - f^t(\omega)}{f^t(\omega)} = g(\omega)$  reflects the growth rate of type  $\omega$ .

We assume that preference changes occur stochastically and within individuals. Let  $q(\omega', \omega'')$  be the probability that an individual's preference  $\omega'$  changes to  $\omega''$ . Obviously,  $\sum_{\omega''} q(\omega', \omega'') = 1$ . Furthermore, we assume that these probabilities depend on two factors, economic success and conformity.<sup>29</sup> To capture the role of conformity we assume

---

<sup>29</sup> The role of conformity in shaping cultural evolution has been pointed out by Boyd and Richerson (1985). For a discussion of this and an alternative modelling approach see Bowles (1999). How conformity can affect consumption preferences is, for example, illustrated in Urberg (1992) who deals with smoking, or in Oostveen, Knibbe, and DeVries (1996) who deal with drinking. Experimental evidence in the context of resource dilemmas can be found, for example, in Smith and Bell (1994). For economic models incorporating conformity, see Bernheim (1994) or Akerlof (1997).

**Assumption 1 (conformity)** Ceteris paribus,  $q(\omega', \omega'')$  is proportional to  $f(\omega'')$ .

Assumption 1 implies that  $q(\omega', \omega'')$  can be written as  $f(\omega'')$  times some other function  $Q(\omega', \omega'')$ . In order to embed the role of economic success we assume the following.

**Assumption 2 (economic success)**  $Q(\omega', \omega'')$  only depends on  $\Pi(\omega'')$  and is strictly increasing in it.

This gives rise to a stochastic process where  $F(\omega)$  develops over time. Below we will show that under certain additional (regularity) assumptions such a process behaves like a growth-monotonic evolutionary process, i.e., like a process assumed in the main body of the paper where shares of types grow according to their relative economic success. These assumptions are

**Assumption 3 (regularity)** (a) The matching scheme  $S$  specifies random matching. (b)

The population is large enough for the law of large numbers to apply.

**Theorem 3** *Under Assumptions 1 to 3 the dynamic process of individual preference adaptation behaves like a growth-monotonic evolutionary process, i.e.  $g(\omega') > g(\omega'') \Leftrightarrow \Pi(\omega') > \Pi(\omega'')$ .*

**Proof.** With Assumptions 1 and 3 we can write

$$f^{t+1}(\omega') = \sum_{\omega} f^t(\omega) Q(\omega, \omega') f^t(\omega').$$

Therefore,

$$g(\omega') > g(\omega'') \Leftrightarrow \sum_{\omega} f^t(\omega) Q(\omega, \omega') > \sum_{\omega} f^t(\omega) Q(\omega, \omega'').$$

Due to Assumption 2  $Q(\omega, \omega') > Q(\omega, \omega'') \Leftrightarrow \Pi(\omega') > \Pi(\omega'')$ . Hence, the claim follows. ■

The widely used replicator dynamics belong to the class of growth-monotonic evolutionary processes and it is easy to see when a process of individual preference adaptation behaves like the replicator dynamics.

**Corollary 4** *If  $q(\omega', \omega'') = \frac{\Pi(\omega'')}{\sum_{\omega} f(\omega)\Pi(\omega)} f(\omega'')$  the process of individual preference adaptation behaves like the replicator dynamics.*

**Proof.** Straightforward. ■

## B A simple case with imperfect signals

Suppose first movers receive, prior to making their decision whether or not to enter the contract, a signal  $s \in \mathbf{R}$ . The signal technology is the following: If player 2's true type is  $H$ , the signal is  $1 + \varepsilon$ , with  $\varepsilon$  being normally distributed with mean zero and variance  $\sigma^2$ . If player 2's true type is  $M$ , the signal is  $0 + \varepsilon$ , with  $\varepsilon$  coming from the same normal distribution.

How will player 1 decide about whether to enter the contract? Nothing changes in case of  $p > \frac{a}{1+a}$ : She will always enter. (This follows from the fact that she even enters when she knows for sure that player 2 is of type  $M$ .) In case of  $p < \frac{a}{1+a}$  player 1 has to update her beliefs about the type of her partner by using Bayes' rule. And if and only if the probability for player 2 being of type is large enough, she will enter. Hence, there will be a critical value  $\bar{s}$  such that player 1 enters if  $s \geq \bar{s}$  and stays out otherwise.

Given  $\bar{s}$  and the signal technology one can now compute the probabilities with which the two different types are offered contracts. And using these probabilities one can calculate the expected payoffs of both types or, in a population model, the average payoff of both types. Obviously,  $H$ -types will always get more contracts than  $M$ -types. (That signal  $s$  exceeds  $\bar{s}$  is more likely for  $H$ -types.) On the other hand,  $M$ -types benefit from profitable breach. Which of these two effects is the stronger one critically depends on how noisy the signal is, i.e., on the variance  $\sigma^2$ .

There are two boundary cases: (i)  $\sigma \rightarrow 0$ , the case of a perfect signal, where the first effect is stronger than the second (and where trustworthiness is crowded in). (ii)  $\sigma \rightarrow \infty$ , the case without a signal where the second effect is stronger (and trustworthiness is crowded out).

Obviously, one can now find a critical standard deviation,  $\bar{\sigma}$ , inducing identical expected monetary payoffs for both types. If the standard deviation is above this level,  $M$ -types will earn more than  $H$ -types and the third result of Proposition 2 would be reversed. If the standard deviation is, however, below this level the original result is resurrected.



## C Sample instructions

*for random matching, medium probability, player 1 and the first phase of the experiment*

Welcome to this research project!

You are participating in a study in which you have the opportunity to earn cash. The actual amount of cash you will earn depends on your choices and the choices of other persons. At the end of the study, the amount of cash earned will be added to your show-up fee and paid to you in cash.

What the study is about:

The study is on how people decide. You are randomly matched with another person present in this room. You and the other person have to choose between two alternatives. The payoff table tells you how much money you earn depending on what you choose and what the other person chooses.

How the study is conducted:

The study is conducted anonymously, **without communication between the participants**, and repeated 9 rounds. Participants are only identified by a letter or a number called “code number”. Neither the other participants nor the researcher will ever know how you decided. You are randomly matched with another person after each round. **You will never interact with the same person again.**

*You are person 1.*

Start of the study

Round 1:

The payoff table reads as follows:

First you have to choose between A and B.

If you choose A, you and the other person receive 50 cents each.

If you choose B, the other person 2 gets to choose between Y and Z.

If person 2 chooses Y, you and person 2 receive 150 cents each.

If the other person chooses Z, chance decides about your earnings.

You earn 150 cents with probability 0.5 ( $\alpha$ ) and 20 cents with probability 0.5 ( $\beta$ ), i.e. your expected earnings after a chance move are 85 cents.

The other person earns 120 cents with probability 0.5 ( $\alpha$ ) and 250 cents with probability 0.5 ( $\beta$ ), i.e. his or her expected earnings after a chance move are 185 cents.

Payoff Table

Who decides	Alternatives	Earnings for 1	Earnings for 2
Person 1	A	50	50
	B	→	→
Person 2	Y	150	150
	Z	→	→
Chance	$\alpha$ (prob = .5)	150	120
	$\beta$ (prob = .5)	20	250

Now, please open your envelope. It contains 9 decision sheets and a code number sheet. Please take everything out of the envelope. Keep the code number sheet. Then **choose between A and B**. Indicate your choice on the decision sheet marked “Round 1”, put this decision sheet back into the envelope and put it into the box which we will pass around. Keep all other decision sheets.

Persons 2 are randomly allocated an envelope and asked to look at your decision and—if they get to make a choice—indicate their choice of either Y and Z on the decision sheet. Decision sheets will be put back into the envelope and into the box.

We collect all decision sheets and count how many people in this room chose A, B, Y, and Z, respectively, and inform all of you of the aggregate outcome of the first round. We then give you the envelope back. Please take the decision out. The information on the decision sheet is private. Please do not share it with anybody else.

Chance now decides whether  $\alpha$  or  $\beta$  will be realized in this round. For this purpose we draw a card from a pile with 5 red and 5 black cards. Red implies  $\alpha$ , black implies  $\beta$ .

We determine your earnings according to your choice and the choice of the other person after the study is over. Your earnings will be paid to you in cash.

End of round 1.

## D Data

Table 8 summarizes behavior for all sessions over time.

Session	Round								
	1	2	3	4	5	6	7	8	9
<b>1 (RLLB)</b>									
<i>No Trust</i> ( $\bar{T}$ )	4	8	10	8	3	2	3	1	5
<i>Trust &amp; Breach</i> ( $T, \bar{P}$ )	5	2	0	0	2	2	1	2	0
<i>Trust &amp; Perf.</i> ( $T, P$ )	1	0	0	2	5	6	6	7	5
<b>2 (RMLB)</b>									
<i>No Trust</i> ( $\bar{T}$ )	0	0	4	10	10	10	8	4	8
<i>Trust &amp; Breach</i> ( $T, \bar{P}$ )	10	12	8	4	4	2	3	6	4
<i>Trust &amp; Perf.</i> ( $T, P$ )	4	2	2	0	0	2	3	4	2
<b>3 (RHLB)</b>									
<i>No Trust</i> ( $\bar{T}$ )	0	0	0	5	8	2	7	6	4
<i>Trust &amp; Breach</i> ( $T, \bar{P}$ )	2	2	2	7	2	8	4	2	2
<i>Trust &amp; Perf.</i> ( $T, P$ )	12	12	12	2	4	4	3	6	8
<b>4 (FMLB)</b>									
<i>No Trust</i> ( $\bar{T}$ )	0	1	0	2	0	0	4	0	8
<i>Trust &amp; Breach</i> ( $T, \bar{P}$ )	1	0	0	0	0	2	0	0	0
<i>Trust &amp; Perf.</i> ( $T, P$ )	7	7	8	6	8	6	4	8	0
<b>5 (RLLH)</b>									
<i>No Trust</i> ( $\bar{T}$ )	6	3	9	6	7	7	8	9	10
<i>Trust &amp; Breach</i> ( $T, \bar{P}$ )	5	6	2	5	3	3	3	3	2
<i>Trust &amp; Perf.</i> ( $T, P$ )	6	8	6	6	7	7	6	5	5
<b>6 (RMLH)</b>									
<i>No Trust</i> ( $\bar{T}$ )	0	1	1	11	10	8	8	7	8
<i>Trust &amp; Breach</i> ( $T, \bar{P}$ )	8	7	8	1	1	1	0	1	0
<i>Trust &amp; Perf.</i> ( $T, P$ )	6	6	5	2	3	5	6	6	6

Table 8: Behavior over time.