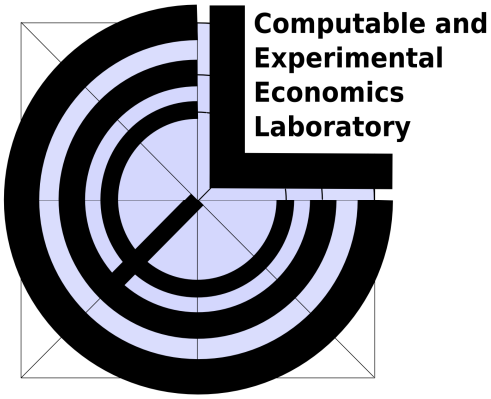



Francesco Guala and Luigi Mittone

# An Experimental Study of Conventions and Norms

CEEL Working Paper 8-10

Computable and Experimental Economics Laboratory

Via Inama, 5 38100 Trento, Italy

http://www-ceel.economia.unitn.it
tel. +39.461.882246

# An Experimental Study of Conventions and Norms*

Francesco Guala
(Department of Sociology and Philosophy, University of Exeter and CEEL, University of Trento)

Luigi Mittone
(Department of Economics and CEEL, University of Trento)

**Abstract**

Although it is now recognized that norms play an important role in many economic decisions, compliance with conventions is generally considered to be driven by rational self-interest only. We report instead experimental data showing that (1) 'external' norms of fairness sustain social conventions that have emerged from repeated play of simple coordination games; and (2) with repetition such conventions acquire an 'intrinsic' normative power of their own. This creates pressure towards conformity, and patterns of regular behaviour that are far stronger and more stable than those that would be generated by mere self-interest and rationality.

* Research for this paper was financed by the ESRC grant RES-000-22-1591 and the Computable and Experimental Economics Laboratory of the University of Trento. Early drafts were presented at the universities of Amsterdam, Cambridge, and Trento, a seminar of the British Society for Philosophy of Science, meetings of the International Network of Economic Method and of the Italian Society for the History of Economic Thought. We are grateful to members of these audiences, Ivan Moscati, Mario Gilli, Matteo Ploner, Marco Failla and two anonymous ESRC reviewers for their constructive feedback. Tim Miller and Marco Tecilla provided invaluable help in designing the software and running the experiments. The usual disclaimers apply.

## 1. Introduction

Economists and psychologists generally agree that norms of equality and reciprocity play an important role in many socio-economic contexts, for instance by helping subjects to achieve Pareto-superior outcomes or fairer allocations of goods than is predicted by standard rational choice theory (Fehr and Fischbacher 2002). In many other areas, however, individualistic instrumental rationality or some close variant thereof is considered to be prevalent. Consider *conventions*, for example: following David Lewis (1969), it has become customary to model conventions as solutions to repeated coordination games with multiple equilibria. Thomas Schelling (1960) convincingly argued that the problem of equilibrium selection in many coordination games is solved by the existence of 'salient' solutions or 'focal points'. In the case of conventions, salience is provided by *precedence*: the strategy followed by the majority of players up until now becomes a focal point that we can all use to successfully coordinate from now on. When this is the case, an equilibrium becomes established as *the* convention in a given population and the coordination problem is solved (Lewis 1969).

Repeated coordination games constitute the vast majority of social interactions that we face in everyday life. The Lewis-Schelling approach therefore seems to vindicate the analytical power of the rational choice approach: individual self-interest *cum* a theory of induction and focal points may go a long way towards explaining this ubiquitous feature of social life. This reading, however, is not entirely correct. Lewis (1969) explicitly highlighted the importance of social norms in sustaining conventions by proscribing deviations from established behavioural regularities. Other theorists, like Margaret Gilbert (1989), have gone even further and questioned the very distinction between norms and conventions. They have argued that conventions are intrinsically normative phenomena, and that a standard rational-choice analysis based on individual rationality and narrow self-interest is unable to capture their essential features.

In order to adjudicate between these different views, we report a series of experiments that investigate for the first time the relation between norms and conventions. The

evidence, as we shall see, largely confirms Lewis' insights concerning the importance of norms in sustaining coordination, but also points in the direction of an intrinsic normative element in conventions.

## 2. Conventions and norms

The idea that norms and conventions play an important role in the functioning of markets is as old as economics itself, and in various guises this insight has re-emerged periodically in the history of economic theory. In the contemporary debate it sits at the core of the so-called 'social capital' theory and the 'new' institutional economics, but also drives the work of many experimental and behavioural economists. But what, exactly, are conventions and norms? The standard approach in contemporary rational choice theory has been to model these important institutions as solutions to coordination and social-dilemma problems, respectively (e.g. Ullmann-Margalit 1977, Schotter 1981, Sugden 1986). In this section we briefly review the standard analysis of conventions and norms, and lay out some hypotheses for experimental investigation.

A simple coordination problem is represented in Table 1. As usual, the first number in each cell represents the payoff of the row player, the second one of the column player. Each Nash equilibrium (Left-Left and Right-Right) is a possible solution of this game. A perfectly rational calculator cannot do better in this game than by flipping a coin and choosing a strategy at random. Schelling (1960) however noticed that in many real-life situations human beings seem to be much more successful at coordinating than purely rational calculators. He argued that seemingly irrelevant features of the environment, such as the position of the objects of choice or the way they are labelled, can function as 'cues' that help us to converge on a common solution. A strategy that is made 'salient' by such features is called a 'focal point', and in the course of repeated interaction is likely to become a point of attraction for the individuals in a given population (see also Sugden 1986).

| | Left | Right |
|---|---|---|
| Left | 1, 1 | 0, 0 |
| Right | 0, 0 | 1, 1 |

Table 1: A simple coordination game

Schelling's hypothesis was mainly based on anecdotal examples and rudimentary experiments, but recent more systematic studies have confirmed its validity (Mehta et al. 1994). The first important application of focal points theory dates back to 1969, with the publication of a seminal book on *Convention* by the philosopher David Lewis. Lewis, who was mainly interested in the conventional nature of language, introduced the idea that a given strategy may be made salient simply by precedence, or the fact that it was played by a sufficiently large number of people in previous rounds of the game. When this happens, we shall say that a convention has emerged in a given population.

In one important respect Lewis' theory sits firmly in the rational choice tradition. Our main motivation to follow a convention is strictly selfish: we drive on the left because we want to avoid accidents; we say 'cat' rather than 'tac' because we want to be understood by our interlocutors; we wear black at funerals because we want to communicate our grief. Lewis' approach thus leads naturally to a neat separation between social norms and conventions. A social norm always comes with an intrinsic 'ought', and is usually backed up by a system of sanctions. The sanctions are meant to change the payoffs of the game: for example, to change a mixed-motives game (like a prisoner's dilemma) into a coordination game (Figure 1).[1]

[1] For a seminal game-theoretic account of social norms along these lines, see Ullmann-Margalit (1977).

| | Left | Right |
|---|---|---|
| Left | 2, 2 | 0, 3 |
| Right | 3, 0 | 1, 1 |



| | Left | Right |
|---|---|---|
| Left | 2, 2 | 0, 0 |
| Right | 0, 0 | 1, 1 |

Figure 1: Transforming a Prisoner's Dilemma game into a Coordination game.

The transformation of (3, 0) and (0, 3) into (0, 0) may take place in different ways. If the payoffs represent utility values, as it is usually the case in game theory, the reduction of the 'free-riding' payoffs (Right-Left and Left-Right) may be due to a feeling of guilt or shame: the other player had trusted my cooperation and I have let her down, for example. But in many societies there are external mechanisms that can reduce our payoffs both at the psychological and the material level: a verbal reproach and ostracism from business are examples of how normative pressure helps in attaining socially superior equilibria in the game of life.

Roughly, then, a social norm exists when every individual (1) prefers to conform to the norm provided that (almost) everybody else does the same; (2) it is common knowledge that one ought to conform; and (3) this normative expectation is backed up by sanctions.[2] Lewis somewhat misleadingly claims that 'conventions are a species of norms'. But it is important to realize that Lewis-Conventions are *not* norms in the sense specified by conditions (1)-(3). Rather, conventions are supported by *extrinsic* normative considerations: one follows a convention because (a) it is individually rational to do so, and (b) deviance from conventions is usually sanctioned by *other* independent social norms. [3] Although a convention does not *per se* imply a commitment to conformity, in other words, normative pressure to abide is exercised by a set of 'external' social norms that tend to reduce deviance from an established regularity.

[2] Although Lewis does not analyze social norms in depth, 'Ludovician' theories of norms can be found e.g. in Pettit (1990) and Bicchieri (2006).

[3] See in particular Lewis (1969, p. 98). A detailed discussion of this aspect of Lewis' theory can be found in Gilbert (1989, especially p. 354).

Suppose, for example, that two players have been playing repeatedly the coordination game in Table 1. With repetition, Left-Left has emerged as a convention, and both players expect the other to continue to comply with it. Imagine that the situation were to change in such a way that one of the players now has an incentive to deviate from the established convention.[4] By doing that she would damage the other player, an action that is sanctioned by various norms in our society. The social consequences of violating these norms may be serious enough to compensate for the gain, and thus deter the player from deviating. In this sense, 'external' norms play an important role in sustaining social conventions.

In Lewis' theory, therefore, norms enhance the resilience of social conventions in the presence of potential deviants. Given that repeated coordination is the prevalent form of social interaction, this analysis attributes much more importance to the existence of social norms than is commonly recognised by rational choice theorists. The importance of norms in sustaining coordination however has never been empirically investigated so far. The first question we will try to answer in our experiments is:

> (Q1) Do 'external' social norms help sustaining conventions in the way envisaged by Lewis?

Some authors however have argued that Lewis' theory *underestimates* the normativity of convention. Margaret Gilbert (1969) has argued forcefully that social conventions and related concepts (customs, tradition, rules) must be analysed in terms of a more primitive notion of collective agency. In particular, conventions result from a "quasi-agreement" among members of a group to pursue a certain line of action that will attain a specific

[4] Notice that the new incentives need not be material: in some cases a mere change in one's preference structure creates the temptation to defect from a convention. A suicidal person, for example, may be tempted to kill herself by driving on the wrong side of the road. The pressure of moral norms however may induce her to find an alternative way of committing suicide that does not cause harm to others. It is also possible that an individual may simply fail to conform to a convention by mistake. In such cases ('trembling hands'), the 'external' norms and associated sanctions will make sure that absent-minded deviants pay more attention in the future.

collective goal. Such quasi-agreements need not be formulated explicitly, but are often inferred from the mere observation that people do pursue a certain line of action that serves the goals of the relevant group. Collective intentions result in a *joint commitment* that cannot be unilaterally breached by an individual group member. This is why, according to Gilbert, we usually feel the need to excuse and justify a breach of convention in front of other group members. [5]

Gilbert's 'contractarian' theory blurs the very distinction between conventions and norms that is at the core of the standard rational choice approach. Conventions are not only supported by external norms that sanction deviance for independent reasons, but *are* themselves norms. Failure to comply violates a normative principle that stands at the core of conventional behaviour. This view is not implausible, in the light of what we know about human tendencies towards conformity. Classic experiments in social psychology show that deviance from conventional behaviour is psychologically costly, to such an extent that people prefer to give false perceptual reports to going against a group's majority (Sherif 1936, Asch 1951, 1956). Thus in the context of coordination problems the mere formation of a majority and the collective repetition of a task may create a normative pressure on members of a group to conform to a behavioural regularity, even when they have an individualistic incentive to deviate.

[5] There are by now several theories of collective agency in the scientific and philosophical literature. The labels vary from discipline to discipline, from 'collective intentionality' in philosophy (Searle 1990, Bratman 1993, Tuomela 1995), to 'group identity' in social psychology (Tajfel and Turner 1989), and 'team reasoning' in economics (Sugden 2000, Bacharach 2006). Here we focus on Gilbert's version because of its contractarian aspect and its accent on the intrinsic normativity of convention. The unifying trait of theories of all theories of collective agency is the claim that both individuals *and* groups can be legitimately taken as units of agency. These theories however are still individualistic in the sense that collective preferences and beliefs are supposed to be 'stored' in the minds of the individual members of the group. The idea is that each individual carries different preference/beliefs profiles that are put at use in different circumstances. Depending on the context, one can act either on the basis of her own individualistic preferences and beliefs, or on the basis of collective preferences and beliefs *qua* group member.

Evolutionary game theorists and cultural anthropologists also assign an important role to conformity in their models. Consider the phenomenon of *altruistic punishment*, for example: there is extensive evidence that many experimental subjects are willing to punish deviations from established norms of fairness by destroying the earnings of free riders, even if this act is costly for the punishing individuals (Fehr and Gachter 2000, 2002; Gintis 2000). This form of 'strong reciprocity' can be used to stabilize pro-social norms of fairness and cooperation. Such norms, in turn, may confer a fitness advantage to a group of cooperators, and take over an entire population via processes of group selection. Group selection works only under relatively strict conditions, though: it is crucial in particular that variation among groups is maintained in spite of migration. Henrich and Boyd (2001) have shown that a moderate amount of conformist learning can sustain the effect of group selection, by making sure that homogeneity of behaviour is not threatened by new immigrants.

Although conformity may hold the key to understanding the evolution of human cooperation,[6] this hypothesis has never been directly tested so far. The second question we try to answer in our experiments then is the following:

> (Q2) Do conventions tend to evolve into social norms? Or, in other words, does repeated convergence on a coordination solution create an intrinsic normative pressure to conform that may override individual incentives to deviate?

We investigated these two questions using the standard methods of experimental economics. We observed the effect of repeated group coordination on individuals who are occasionally exposed to free-riding opportunities. As we shall explain below, not only this setting is appropriate to investigate Q1 and Q2, but it is also more realistic than many experimental settings that have been studied in depth by economists and psychologists. Social life is neither a constant prisoner's dilemma nor a pure coordination game, but a mixture of the two. In the next two sections we describe the design of our experimental study and the main results obtained.

[6] See also Richerson and Boyd (2005: 203-206), Gintis et al. (2003).

## 3. Study 1

Our first study was run at the University of Exeter (UK) and was designed to answer the first question (Q1) outlined above. Subjects were recruited using email lists from the population of graduate and undergraduate students. Volunteers registered for one of the experimental sessions and, as they arrived at the lab, were seated randomly at one of 18 computer terminals separated by partitions. After signing a consent form, they were asked to read the experimental instructions illustrating the main features of the task. The task varied depending on the experimental condition. Each subject participated in one condition only, and all comparisons took place across subjects. Our first study consisted of a comparison between two conditions, labelled 'Baseline' and 'Norms' respectively. In both conditions subjects played in groups of three players, with random selection of group membership, anonymity, and without the possibility of communication. They received a show-up fee, and on top of that received whatever they earned in the experimental task. Individual payoffs were calculated in terms of 'experimental tokens' which were converted into real money at the exchange rate of 3 *p* per token (or £1 = 33 tokens).[7]

### 3.1. *Design*

The main task in the Baseline condition was a repeated coordination game with two options labelled 'Red' and 'Blue'. At each round each subject chose an option by clicking a button on the screen and then receives feedback regarding the choices of the other two group members. If all members chose the same option, they earned 10 tokens each; otherwise, they received nothing. The payoffs of this game are represented in Table 2. For simplicity, we use a 2 x 2 matrix were the row player is 'You', and the column player is 'Others'.

[7] Average earnings ranged between £9 and £11, depending on the condition, for less than one hour of experimentation.

| | Red | Blue |
|---|---|---|
| Red | 10, 10 | 0, 0 |
| Blue | 0, 0 | 10, 10 |

Table 2: Coordination task

After a few rounds the overwhelming majority of the groups converged on a common strategy, and continued to coordinate for the rest of the game.[8] Subjects were told in the instructions that the game would last for ten rounds. They were also told that the payoffs could change during the course of the game, and that not all players would necessarily be informed in advance if this happened (although they would all be fully briefed *after* one of these 'special rounds' took place). No specific details were provided regarding the payoff structure of these special rounds.

As a matter of fact, in the tenth and last round all groups faced a special task with the payoff structure represented in Table 3. Within each group, only one player (that we will call the 'potential deviant') was made aware in advance of this change in payoff structure. In Table 3 the potential deviant is the row player, and again for simplicity the other two group members are represented as the column player. The potential deviant expects that the other group members will continue to coordinate on the established convention, lacking any reason to do otherwise. She knows, however, that the other players will be told at the end of this round that the payoff structure had changed, and that she was aware of this change.

[8] For curiosity: the dominant convention emerging from these early stages was Red (played 66% of the time). A possible explanation suggested by recent neuroeconomic studies is that when faced with a choice between two indifferent options, we tend to choose the one upon which our sight fixated first, and it's twice as likely that we fixate on the left-hand side than on the right-hand side object. (Rangel 2007).

| | Red | Blue |
|---|---|---|
| Red | 200, 200 | 300, 0 |
| Blue | 300, 0 | 200, 200 |

Table 3: Incentive to deviate in the 10th round

Notice that the game in Table 3 is *not* a coordination game. Since the other two players are going to continue to play the convention ('Red-Red', say), the potential deviant faces a straightforward choice between maximizing the earnings of the group vs. her own income. With this design we can detect the effect of norms on individual behaviour by observing whether the experimental subjects are willing to forego individual earnings in order to conform to the regularity that has evolved during the previous rounds of group play. *The normativity of convention is the (normative) expectation that you ought to bear the possible costs of non-deviance, because I am planning my choices based on the (plain) expectation that you will conform.* It is thus manifested in the decision to 'leave some money on the table' and privilege the group's earnings with respect to one's own private gain.

As a matter of fact, in our experiment the majority of potential deviants decided to conform to the established convention and thus maximize the group's earnings.[9] The mere fact that some subjects are willing to conform to the convention and forego individual gains however merely tells us that there is some norm at work, but does not indicate exactly *what kind* of norm we are observing. In the previous section we have distinguished among two possible source of normativity: (i) 'external' norms of fairness, altruism, or cooperation that prescribe to play cooperatively in a situation such as the one represented in Table 3; and (ii) the 'intrinsic' normativity of convention, that is, the pressure to conform to collective group behaviour that has emerged from repeated team play over the first nine rounds of the game. Our two experimental studies were designed to separate these effects as observed in the Baseline condition.

[9] The data are presented in detail and discussed in section 3.3 below.

*3.2. Norms condition*

In the 'Norms' condition we tried to manipulate the 'external' norms that may help to sustain convention-abiding behaviour. Intuitively, three 'external' considerations may prompt subjects to seek the maximization of group rather than individual earnings: *altruistic* considerations prescribing to increase the earnings of the other players, *fairness* considerations prescribing to avoid inequitable outcomes, and/or *utilitarian* considerations prescribing to maximize the sum of the individual payoffs of group members. In either case, such considerations encourage a conformist behaviour with respect to the established convention: altruist, egalitarian, and utilitarian players should all choose Red-Red or Blue-Blue in round ten of the Baseline condition. To observe the force of these 'external' norms at work, we decided to compare behaviour in the Baseline condition with behaviour in a 'Norms' condition when one of these factors (fairness, intended as inequality aversion) has been controlled for by design.

The Norms condition is in all respects identical to the Baseline, except that the payoffs in the 10th ('special') round are those represented in Table 4. The main difference between this payoff structure and the one in Table 3 is that following the convention now carries a higher cost for the potential deviant. The choice therefore is not between collective and individual gain, but between others' gain and my gain. Since 'my loss is your gain' (and vice-versa), the decision to stick to the convention can be supported only by a norm of altruism or a utilitarian norm, but not by a norm of fairness (or inequality-aversion) as in Table 3. If external norms of fairness did play a role in sustaining conventions, we would expect to observe lower compliance with the convention in the Norms than in the Baseline condition.

| | Red | Blue |
|---|---|---|
| Red | 0, 200 | 300, 0 |
| Blue | 300, 0 | 0, 200 |

Table 4: Payoffs 10th round of the Norms condition

3.3. *Results*

Experiment 1 was designed to answer the question:

> (Q1) Do egalitarian, altruistic, and utilitarian norms provide 'external' support to social conventions?

If the answer to this question were positive, we would expect to observe *more* deviance in Norms than in the Baseline condition. If, in contrast, the answer was negative, there should be no significant difference both in deviance and in punishment rates across the two conditions.

The data were collected at the Financial and Experimental Economics Laboratory of the University of Exeter (FEELE) between October 2006 and February 2007. In total 273 subjects participated in the experiment (141 in the Baseline and 132 in the Norms condition). In each condition we obtained one observation regarding potential deviants per each group of three players. A first process of data-analysis was run to identify and discard those subjects who had misunderstood the nature of the experimental task, based on their answers in the pre-experimental questionnaire and/or failure to coordinate during the first nine rounds of the coordination game. This left us with 72 observations regarding potential deviants (38 in Baseline, 34 in Norms).

The deviation data for each condition are summarized in Table 5. Overall, out of 72 potential deviants, 33 (45.8%) decided to maximize their own monetary gains instead of coordinating with the other members of their group. Although 54.2% may seem a relatively high proportion of players displaying 'socially-oriented' or 'norm-abiding' behaviour, these figures are consistent with highly replicated results in Public Goods, Dictator, and similar experimental games.[10] Conformity to the convention however is distributed unevenly across the two conditions. While in the Baseline only 28.9% of

[10] Charness and Rabin (2002) provide data from a wide range of settings.

subjects deviate, when egalitarian concerns are removed by design (in the Norms condition) the rate of deviance surges to 64.7% of the sample.

| Condition | N | Deviants | % |
|---|---|---|---|
| Baseline | 38 | 11 | 28.9 |
| Norms | 34 | 22 | 64.7 |
| Total | 72 | 33 | 45.8 |

Table 5: Experiment 1: deviation rates

Comparison between the Baseline and the Norms condition therefore indicates that (1) 'external' social norms of fairness and cooperation do sustain conventions, and can be manipulated experimentally by changing the relative payoffs of the game. Statistical analysis confirms that the null hypothesis ($H_0$: 'there is no effect from manipulating payoffs in the Norms condition') is rejected using an ordinary chi-square test ($\chi^2 = 9.242$, $df = 1$, $p = .002$) as well as more demanding non-parametric tests (Mann-Whitney $Z = -3.019$, $p = .003$; Kolmogorov-Smirnov $Z = 1.515$, $p = .020$).

## 4. Study 2

Our second study was designed to test whether conventions impose an *intrinsic* normative obligation, independently of the influence of external norms. To discriminate between external norms and the intrinsic normativity of conventions, we compared behaviour in the Baseline condition with behaviour in a structurally similar one-shot game, where subjects face a similar payoff structure to the one of Table 3, but *without being exposed to a history of successful coordination in the previous rounds of the game*.[11]

[11] To control for portfolio effects, the One-shot condition was preceded by an incentivised questionnaire where subjects were given the chance of earning an amount of tokens roughly equivalent to that earned by the participants to the Baseline condition during the first nine ("normal") rounds of that game. The same questionnaire (aimed at facilitating understanding of the game) was also administered before the Baseline condition, but without monetary incentives.

4.1. *Design*

In the 'One-shot' condition subjects again played in groups of three, anonymously and without communication. The payoff structure was transparent for all players this time, and the game was played in two stages as follows: in the *first stage*, two players had to choose simultaneously between two options (Red and Blue); if they chose the same option, the game proceeded to stage two; if in contrast they failed to coordinate, the game ended and all players earned zero tokens. In the *second stage* of the game, the third player (the potential deviant) faced a choice between: (a) playing the same colour chosen by the other players in the previous round, or (b) playing a different colour. The payoff structure is represented in Figure 2, as seen from the viewpoint of the potential deviant (or 'Self' player).

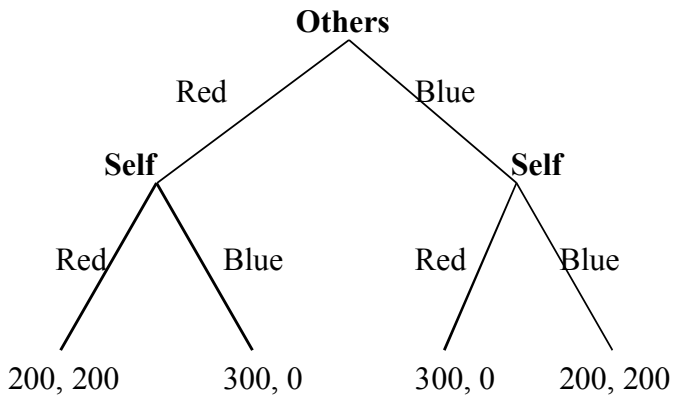



Figure 2: The One-shot condition

Although the One-shot condition and the final round of the Baseline condition are not strictly speaking isomorphic from a game-theoretic point of view, the choice faced by the potential deviant becomes identical if the potential deviants in the Baseline condition expect the other players to play the established convention in the 10th round of the game (which is confirmed by the behavioural data). In either condition, then, the potential deviant must take the choices of the other players as given, and faces a straightforward

dilemma between an individualistic and a cooperative choice. The relevant difference, for our purposes, lies in the fact that in the Baseline the potential deviant may be influenced by precedence – a history of repeated convergence on a common strategy – whereas in the One-shot condition (which is similar to a Dictator game) this element has been removed by design. If individualistic behaviour is more common in the One-shot than in the Baseline condition, then, we would obtain direct evidence that repeated group play generates an independent, intrinsic normative pressure on individuals to conform to the group's conventions. This normativity, combined with external norms of altruism and fairness, should increase the resilience of conventions to changes in individual incentives.

4.2. *Results*

For reasons to be explained shortly, we ran this study twice: at the University of Exeter during the winter of 2006-07 and at the Computable and Experimental Economics laboratory of the University of Trento in June 2008. We report for completeness the Exeter data first, and explain why we felt that a replication with a slightly modified design was required.

Overall 207 observations were collected in the Exeter run of the One-shot condition; since several players failed to coordinate in the first stage of the game, the proportion of useful data-points concerning potential deviants was lower here than in the other conditions (46 out of a total maximum of 69). The relevant data can be found on the left-hand side of Table 6. Recall that the study aimed at answering the following question:

> (Q2) Are conventions endowed with intrinsic normative power that significantly reduces deviance rates, independently of these external norms?

If the answer were positive, we would expect to observe *more* deviance in One-shot than in the Baseline condition. If the answer were negative, in contrast, there should be no significant difference both in deviance and in punishment rates across the two conditions.

Comparison between the One-shot and Baseline conditions at Exeter suggests that mere repetition of a collective task enhances the conformity to a convention. Statistical testing however provides a weak result, with the null hypothesis ($H_0$: 'there is no effect from repeated group play') coming close only to rejection at the 10% level in a chi-square ($\chi^2 = 2.463$, $df = 1$, $p = .116$) and in a Mann-Whitney test ($Z = -1.560$, $p = .119$), but failing by a good margin in a Kolmogorov-Smirnov non-parametric test ($Z = .762$, $p = .607$).

| | Exeter | | | Trento | | |
|---|---|---|---|---|---|---|
| **Condition** | **N** | **Deviants** | **%** | **N** | **Deviants** | **%** |
| Baseline | 38 | 11 | 28.9 | 29 | 9 | 31.0 |
| One-shot | 46 | 21 | 45.6 | 28 | 19 | 67.8 |
| Total | 84 | 32 | 38.0 | 57 | 28 | 49.1 |

Table 6: Study 2: deviation rates

The One-shot condition as implemented in the Exeter experiments however was flawed in one important respect. In the first stage of the game the first two members of the group have the option of ending the game immediately by failing to coordinate, which would result in an equal payoff of zero tokens to each player. By deciding *not* to derail the game (that is, by successfully coordinating on either Red or Blue), the first two players thus send an implicit message to the potential deviant that may be read as an offer of cooperation. There is extensive evidence from the experimental literature that 'nice' moves of this kind can trigger norms of reciprocation.[12] It is possible, then, that some third players would respond 'nice' with 'nice' and choose to conform to the first two players' choice. Intentions in contrast cannot play a similar role in the Baseline. Here the first two players cannot see the opportunity of ending the game by failing to coordinate. On the contrary, they stick to the convention because they think that this is the best course of action for all members of the group. Their success in coordinating, therefore, cannot be read by the potential deviant as a 'nice' intentional move to be reciprocated by conforming.

[12] See e.g. Charness and Rabin (2002), Falk et al. (2003), McCabe et al. (2003), Cox (2004).

It can be argued then that our original design failed to control for one important motivation. This convinced us to run a new, improved experiment where intentions and reciprocity have been properly shielded by design. In this replication we simply eliminated the opportunity for the first two players to end the game after the first stage: instead of making their choice autonomously, a computer programme assigned a colour to both of them before proceeding to the second stage of the game. As in the Baseline condition, the third player now could not read their moves as a 'nice' intentional offer but only as an accident of the game.

Our first task in Trento was to replicate the evidence obtained in the Baseline condition at Exeter, so as to compare it with the newly designed One-shot condition. In order to make comparisons possible, we kept the design of the Baseline condition as well as the general procedures as similar as possible to those implemented in the Exeter part of the experiment. In total 171 subjects participated in two sessions at Trento (84 in the Baseline and 87 in the One-shot condition).

The deviation data for each condition are summarized on the right-hand side Table 6. Notice, first of all, that the results of the Baseline condition replicated those obtained at Exeter. But comparison between the One-shot and Baseline conditions in Trento indicates strongly that mere repetition of a collective task does enhance the conformity to a convention. The null hypothesis ($H_0$: 'there is no effect from repeated group play') is rejected in a two-tailed chi-square test ($\chi^2 = 6.33$, $df = 1$, $p = .0119$), as well as in a Mann-Whitney ($Z = -2.649$, $p = .008$) and in a Kolmogorov-Smirnov non-parametric test ($Z = 1.336$, $p = .056$).

These data confirm our insight that the original design was indeed flawed. While qualitatively we observed in Trento the same surge in deviance rates in the One-shot condition that we had observed in Exeter, the effect was much stronger once reciprocal motives had been controlled for by design. On the basis of these data we feel confident to

conclude that repeated coordination play leads to the emergence of a convention that is normatively binding for the members of the group.

## 5. Conclusions

Coordination games are a ubiquitous form of social interaction, and understanding their functioning should be a priority in the foundations of social science. In contrast, most experimental activity in economics has been devoted to studying problems of cooperation, especially in prisoner's dilemma-like situations. These situations are important and allow one to observe the effect of social norms at work, but have also prompted excessively pessimistic conclusions regarding the possibility of cooperation. If life were a constantly repeated prisoner's dilemma game, then society would probably not exist. Instead of being constantly exposed to free-riding temptations, we usually play repeated coordination games with the members of our social group, and we are only occasionally exposed to free-riding opportunities. Our Baseline design captures this dyanamics nicely and can be used to study the way in which conventions acquire normative power during repeated team play.

The evidence presented in this paper confirms the hypothesis that social norms play an important role not only in the solution of mixed-motives games, but also in sustaining conventional behaviour in coordination games. Lewis' (1969) claim that norms of fairness and altruism provide 'external' support to conventions is vindicated by our experimental data. The data also indicate that conventions have a tendency to turn into social norms, acquiring an intrinsic normative power that further reduces deviation rates from an established behavioural regularity.[13] At the practical level, our results suggest

[13] This result goes well beyond the expectations of those social scientists who have highlighted conformity as an important element in the evolution of pro-social behaviour. Richerson and Boyd, for example, have claimed that conformity is likely to play a role 'only if individuals have difficulty evaluating the costs and benefits of alternative cultural variants' (2005: 206). If this were true, we should expect pressure towards conformity to be rather weak in presence of well-defined individual incentives. As we have seen, in contrast, the effect of conformity is strong even with clearly stated payoffs and immediate rewards.

that conventional modes of behaviour such as traditions, habits, and routines may be more difficult to disrupt by changing individual incentives than one would assume based on standard rational choice analysis. Since the texture of norms that keeps societies together is probably far stronger than traditionally recognized by economists, the theoretical and experimental study of norms and their functioning is a most urgent task in economics and social science more generally.

**References**


Asch, S.E. 1951. "Effects of Group Pressure upon the Modification and Distortion of Judgements", in Guetzkow, H. (ed.) *Groups, Leadership and Men*. Pittsburgh: Carnegie Press.

Asch, S.E. 1956. "Studies of Independence and Conformity: A Minority of One against a Unanimous Majority", *Psychological Monographs: General and Applied* 70: 1-70.

Bacharach, M. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton: Princeton University Press.

Boyd, R. and Richerson, P.J. 2005. *The Origin and Evolution of Cultures*. Oxford: Oxford University Press.

Bratman, M. 1993. "Shared Intention", *Ethics* 104: 97-113.

Burlando, R.M. and Hey, J.D. 1997. "Do Anglo-Saxons Free Ride More?" *Journal of Public Economics* 64: 41-60.

Charness, G. and Rabin, M. 2002. "Understanding Social Preferences with Simple Tests", *Quarterly Journal of Economics* 117: 817-69.

Cox, J. 2004. "How to Identify trust and Reciprocity", *Games and Economic Behavior* 46: 260-81.

Falk, A., Fehr, E. and Fischbacher, U. 2003. "On the Nature of Fair Behavior", *Economic Inquiry*, 41: 20-6.

Fehr, E. and Fischbacher, U. 2002. "Why Social Preferences Matter – The Impact of Non-selfish Motives on Competition, Cooperation and Incentives." *Economic Journal* 112: C1–C33.

Fehr, E. and Gachter, S. 2000. "Cooperation and Punishment", *American Economic Review* 90: 980-994.

Fehr, E. and Gachter, S. 2002. "Altruistic Punishment in Humans", *Science* 415: 137-40.

Gilbert, M. 1989. *On Social Facts.* London: Routledge.

Gintis, H. 2000. "Strong Reciprocity and Human Sociality", *Journal of Theoretical Biology* 206: 169-179.

Gintis, H., Bowles, S., Boyd, R., Fehr, E. 2003. "Explaining Altruistic Behavior in Humans", *Evolution and Human Behavior* 24: 153-172.

Henrich, J. and Boyd, R. 2001. "Why People Punish Defectors: Weak Conformist Transmission Can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas", *Journal of Theoretical Biology* 208: 79-89.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E. and Gintis, H. (eds.) 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies.* Oxford: Oxford University Press.

Lewis, D.K. 1969. *Convention: A Philosophical Study.* Cambridge, Mass.: Harvard University Press.

McCabe, K., Rigdon, M. and Smith, V. 2003. "Positive Reciprocity and Intentions in Trust Games", *Journal of Economic Behavior and Organization* 52: 267-275.

Mehta, J., Starmer, C. and Sugden, R. 1994. "The Nature of Salience: An Experimental Investigation of Pure Coordination Games", *American Economic Review* 84: 658-673.

Rangel, A. 2007. "The Role of Visual Attention in Simple Choices", California Institute of Technology.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M. and Zamir, S. 1991. "Bargaining and Market Behavior in Jerusalem, Lubljana, Pittsburgh and Tokyo: An Experimental Study", *American Economic Review* 81: 1068-95.

Schelling, T. 1960. *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.

Schotter, A. 1981. *The Economic Theory of Social Institutions*. Cambridge: Cambridge University Press.

Searle, J.R. 1990. “Collective Intentions and Actions”, in Cohen, P.R., Morgan, J. and Pollack, M.E. (eds.) *Intentions in Communication*. Cambridge, Mass.: MIT Press.

Sherif, M. 1936. *The Psychology of Social Norms*. New York: Harper.

Smith, V. L. 1976. “Experimental Economics: Induced Value Theory,” *American Economic Review* 66: 274–7.

Sugden, R. 1986. *The Economics of Rights, Cooperation and Welfare*. Oxford: Blackwell.

Sugden, R. 2000. “Team Preferences”, *Economics and Philosophy* 16: 174-204.

Tajfel, H. and Turner, J.C. 1986. “The Social Identity Theory of Inter-Group Behavior”, in S. Worchel and L. W. Austin (eds.), *Psychology of Intergroup Relations*. Chicago: Nelson-Hall.

Tuomela, R. 1995. *The Importance of Us*. Stanford: Stanford University Press.

Ullmann-Margalit, E. 1977. *The Emergence of Norms*. Oxford: Clarendon Press.

APPENDIX

# Experimental Instructions*

You are now taking part in an economic experiment which has been financed by various foundations for research purposes. The instructions which we have distributed to you are solely for your private information. It is prohibited to communicate with the other participants during the experiment. Should you have any questions please ask us. If you violate this rule, we shall have to exclude you from the experiment and from all payments.

At the end of the experiment you will receive a fixed wage of £ 5 [7 Euros] simply for participating in the experiment (in exchange for your time). On top of that, you will have the opportunity to earn some extra money; the exact amount that you will earn will depend on your decisions and the decisions of the other participants in the experiment.

During the experiment we will not speak of Sterling but rather of tokens. During the experiment your entire earnings will be calculated in tokens. At the end of the experiment the total amount of tokens you have earned will be converted into Sterling at the following rate:

1 token = 3 pence [5 cents] (£1 = 33 [1 Euro = 20] tokens, in other words)

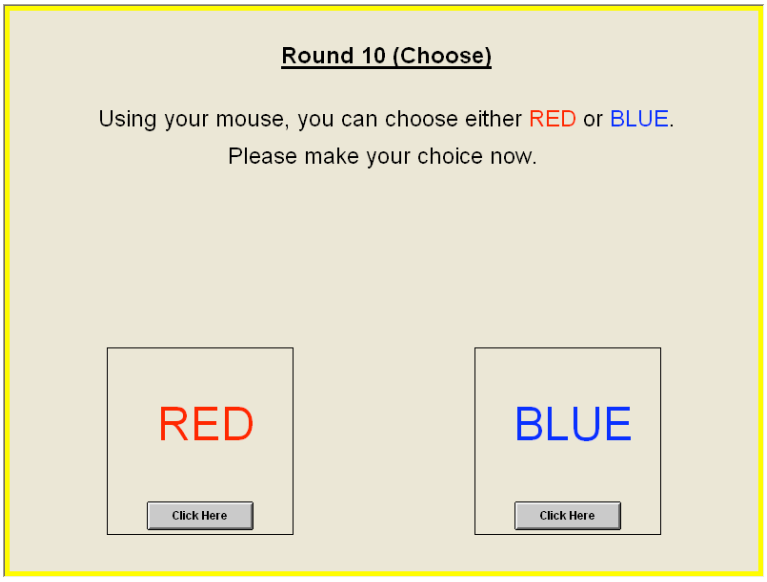
All participants will be divided into groups of three members. Except the experimenters, nobody knows who is in which group. We will preserve anonymity throughout and after the experiment: all the money that you earn will be paid to you privately at your desk when the experiment is over.

**The decision situation**

We first introduce you to the basic decision situation. You will be a member of a group of 3 people. The experiment consists of 10 rounds; the membership of the group will be determined randomly before the first round and will remain the same throughout the 10 rounds.

At each round you will be asked to choose one of two options (named "Red" and "Blue") by clicking on one of two buttons on a screen like this:

* This set of instructions was used for both the "Baseline" and the "Norms" condition in Experiment 1. An Italian translation was used in the Baseline condition of Experiment 2 (we have indicated between square brackets the relevant variations).




At each round, if all three members of the group choose the same colour, they will earn 10 tokens each. If, in contrast, one or more member(s) choose differently from the rest of the group, all three members will earn 0 tokens each. After each round, you will be told automatically how many members have chosen each colour, and the number of tokens earned during that round. For example:

"In round 1, 2 players chose Red, 1 player chose Blue. You have earned 0 tokens in this round."

Then you will be introduced to the following round, and so forth, for the 10 rounds of the experiment.

**Special rounds**

It is possible that the payoff structure will change during the course of the experiment. You may or may not be told in advance that one of these "special rounds" is about to take place. After a "special" round, however, you will be told if the group members were facing a different payoff structure. As usual, you will also be told how many players have chosen which colour, and the number of tokens you have earned in that round.

At the end of the experiment you will be asked to fill in a short questionnaire; when you have finished, leave the completed questionnaire on the table, and wait for the experimenter to bring to your desk the money you have earned. You will be asked to sign a receipt, and then you will be allowed to leave the room quietly. We would also be grateful if you did not discuss the experiment with the other participants outside the laboratory.

Before the experiment begins, a series of questions will appear on your screen. The aim of these questions is to help you familiarise with the logic of the game and with the earnings that follow from various decisions.

# Experimental Instructions*

You are now taking part in an economic experiment which has been financed by various foundations for research purposes. The instructions which we have distributed to you are solely for your private information. It is prohibited to communicate with the other participants during the experiment. Should you have any questions please ask us. If you violate this rule, we shall have to exclude you from the experiment and from all payments.

At the end of the experiment you will receive a fixed wage of 7 Euros simply for participating in the experiment (in exchange for your time). On top of that, you will have the opportunity to earn some extra money; the exact amount that you will earn will depend on your decisions and the decisions of the other participants in the experiment.

During the experiment we will not speak of Sterling but rather of tokens. During the experiment your entire earnings will be calculated in tokens. At the end of the experiment the total amount of tokens you have earned will be converted into Sterling at the following rate:

1 token = 5 cents (1 Euro = 20 tokens, in other words)

All participants will be divided into groups of three members. Except the experimenters, nobody knows who is in which group. We will preserve anonymity throughout and after the experiment: all the money that you earn will be paid to you privately at your desk when the experiment is over.

## Description of the experimental task

We first introduce you to the basic decision situation. You will be a member of a group of 3 people. During the game each player will have two options, named "Red" and "Blue". The experiment will consist of three rounds:

In the **first round**, the computer will randomly assign a role ("first player", "second player", or "third player") to each member of the group. While the first two players have a "passive" role in the experiment, the third one will have the opportunity of making an important decision.

In the **second round**, the computer will assign randomly the same colour (Red or Blue) to the first and then the second player. The assigned colour will be communicated to all members of the group (including the third player) and cannot be rejected by the first two players.

* This set of instructions was used for both the "One-Shot" condition in Experiment 2. We have translated the instructions from Italian.

In the **third round**, the third player will have the opportunity of choosing between the same two colours (Red and Blue), represented by two buttons that will appear on a screen like this one:


The choice will determine the following payoffs:

- If the third player will choose a <u>different</u> colour from the one assigned to the first two members of the group, the third player will earn 300 tokens and the other members will earn 0 tokens each.
- If the third player will choose the <u>same</u> colour that was assigned to the first two players, every member of the group will earn 200 tokens.

When the third player will have made his choice, the experiment will end.

At the end of the experiment you will be asked to fill in a short questionnaire; when you have finished, leave the completed questionnaire on the table, and wait for the experimenter to bring to your desk the money you have earned. You will be asked to sign a receipt, and then you will be allowed to leave the room quietly. We would also be grateful if you did not discuss the experiment with the other participants outside the laboratory.

Before the experiment begins, a series of questions will appear on your screen. The aim of these questions is to help you familiarise with the logic of the game and with the earnings that follow from various decisions. For each correct answer you will earn 15 tokens that will be paid with the other earnings at the end of the experiment.