Journal of Institutional Economics (2008), 4: 3, 327–349 Printed in the United Kingdom
© *The JOIE Foundation 2008* doi:10.1017/S1744137408001112

Two notions of conventions: an experimental analysis

LUIS M. MILLER*

Max Planck Institute of Economics, Jena, Germany

Abstract: This paper aims at analyzing the interaction between both economic and sociological notions of convention. To this end, it starts by distinguishing conceptually between *specific convention*, i.e. an arbitrary but stable social regularity, and *general convention*, i.e. a principle of action prescribing how to behave in a certain class of situations. A game theoretical framework to represent the interrelation between both concepts is then introduced. Finally, this relation is studied experimentally. The main results of the experiment are: (1) *general conventions* have to be commonly known and adopted among subjects in order to work as guides to coordinate on *specific conventions*; (2) once subjects follow a *general convention* they are highly consistent with it in a repeated environment; (3) efficiency concerns are focal in the class of games studied in this paper.

1. Introduction

When considering the whole of modern social sciences, we find that there is no unequivocal and commonly accepted technical definition of the notion of social convention. Consequently, very important theoretical and empirical questions remain unsolved. For instance: Which regularities observed in human behavior can be labeled as conventions? How do conventions emerge, endure, and evolve? Why do humans conform to conventions? These and other related questions find only partial answers mainly from economic theory, analytical philosophy, and, less frequently, sociological theory (see Sugden, 1998; Vanderschraaf, 1998; Biggart and Beamish, 2003, respectively).

Despite this fact, the publication of the book *Convention: A Philosophical Study* by the analytical philosopher David Lewis (Lewis, 1969) has recovered the use of the concept of convention for two main reasons. On the one hand, a number of theorists, mainly economists, have not only accepted the definition put forward by Lewis but have incorporated this notion into their

*Correspondence to: Strategic Interaction Group, Max Planck Institute of Economics, Kahlaische Straße 10, 07745 Jena, Germany. Email: miller@econ.mpg.de

The author is grateful to Fernando Aguiar, Andrés de Francisco, Werner Güth, Ana León, José A. Noguera, Christoph Vanberg, and five referees for constructive advice that induced major revisions of the article. Preliminary versions of this article were presented at the Indiana University Workshop in Political Science and Policy Analysis and the Max Planck Institute of Economics. Financial support received from the Spanish Ministry of Education (SEJ2006-00959/SOCI) is also acknowledged.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

conceptual repertoire. On the other hand, other theorists, mainly philosophers and sociologists, have criticized the Lewis model of convention and have recovered a notion of convention anchored in the methodological foundations of the Sociology of Max Weber (Weber, 1978). The aim of this paper is to explore these two approaches to convention from both a theoretical and an empirical-experimental standpoint. In a previous paper (Büchner *et al.*, 2005), the labels *specific convention* and *general convention* were used to refer to the first and the second approach to convention respectively. Both theoretical approximations to convention are briefly reconstructed in Section 2. In Section 3 a game theoretical framework to represent both types of conventions is defined, while an experimental setting to empirically study this framework is introduced in Section 4. Finally, the experimental results are presented in Section 5 and discussed in Section 6.

2. Two notions of convention

The aim of this section is to briefly reconstruct the two ideal notions of convention introduced above, namely the economic notion of convention, or *specific convention*, and the sociological notion, or *general convention*.

Specific convention

In the last few decades, a number of analytical philosophers and economic theorists have recovered the original notion of convention proposed by the Scottish philosopher David Hume (1978, 1975).[1] Although there is no consensus regarding an exact definition of convention among these theorists, all of them agree on at least three basic characteristics of convention, namely a convention is an actual behavioral regularity that is arbitrary as well as stable.

Firstly, a *specific convention* is assumed to be an actual regularity in the behavior of a given population because it implies the behavioral pattern actually followed by this population in a recurrent situation of social interaction (Lewis, 1969: 42; Schotter, 1981: 10; Young, 1996: 105). Thus, conventions are empirically observable behavioral patterns. For instance, to say a simple 'good morning' when one arrives at work is the established convention in my workplace; but, if I meet a colleague outside the workplace, the convention is just 'shaking hands' if the encounter is between two men and 'giving two kisses' if one of the two parties is a woman. In general, all the rules related to 'good manners' are considered conventions because they assume predominant behavioral regularity in a particular situation of social interaction. Conventions, however, are not only related to politeness and civic-mindedness, but to many other social and economic aspects as well. Recurrent examples of conventions

1 Among others, see Lewis (1969), Schotter (1981), Sugden (1986, 1989), Skyrms (1996), and Binmore (1998, 2005).

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

include adhering to a particular dress code, using a word with its conventional meaning, or many features related to the market and economic activity such as the use of different means of exchange, industrial standards or accounting rules. What is important to keep in mind is that in all of these cases the convention is a customary practice.

Secondly, a *specific convention* is assumed to be arbitrary in the sense of being the realization of only one of the multiple potential regularities that could emerge in a recurrent situation of social interaction. Coleman (1990: 248) provides the following example: 'In some cases the selection of the focal action is largely arbitrary, but in others it is not. The former is exemplified by the convention of driving on the right side of the road (or, in England and Australia, on the left). It is arbitrary whether the action defined as correct is driving on the right or driving on the left' (see also Binmore, 1992; and Lewis, 1969: 70). In game theory terms, this arbitrariness is captured by the representation of convention as a stable solution to a recurrent coordination problem; that is, a convention is just one of the multiple equilibria that a coordination game displays. Therefore, by definition, there always exists an alternative behavioral regularity to the established convention (Lewis, 1969: 68). Perhaps the most straightforward example is driving on the right side or driving on the left side of the road. Other examples could be those related to dress codes, for instance wearing black or white clothes when attending a funeral.

Thirdly, a *specific convention* is assumed to be a stable behavioral regularity. Although the majority of *specific convention* theorists agree with this statement, not all of them agree on the exact meaning of this stability criterion. Hence, whereas authors like Lewis (1969) justify the stability of conventions through the reconstruction of the actual processes of reasoning of those involved in the social interaction situation in which a particular convention emerges (see Cubitt and Sugden, 2003), other authors, such as Sugden (1986, 1989) and Skyrms (1996), base their explanation on the development of behavioral dispositions through an adaptive process, which sooner or later leads individuals to coordinate on a particular convention. This last interpretation has led many theorists to interpret conventions as combinations of evolutionary stable strategies. For instance, Sugden (1986: 32) defines a convention as 'any stable equilibrium in a game that has two or more stable equilibria'.

The notion of *specific convention* sketched above is, however, silent about the origin of conventions. Among economic theorists, authors such as Sugden (1986, 1989) or Skyrms (1996) consider conventions 'the product of human action and not the result of human design'. However, other authors, such as Young (1996), also include in their account of conventions those established by central authority, as the following quotation proves: 'We may discern two ways in which conventions become established. One is by central authority ... the second mechanism by which conventions become established [is]: the gradual accretion of precedent' (p. 106).

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

As Section 3 will show, the three outlined characteristics are sufficient to develop an accurate game theoretical representation of *specific convention*. However, many sociologists have argued that a behavioral notion of convention does not capture the whole scope of convention. According to them, one of the main aspects of convention, which is almost absent from the economic debate, must be considered: its normative character. In the next subsection a sociological notion of conventions will be explored.

General convention

In 1989, two decades after the publication of Lewis' influential work (Lewis, 1969), the philosopher Margaret Gilbert published the book *On Social Facts*. In Chapter 6, the author criticizes the definition of convention put forward by Lewis and proposes a definition that is closer to the one introduced by Max Weber in his *Basic Concepts in Sociology*. The main point of Gilbert's argument is the claim that conventions are of a normative character; that is, individuals have an independent reason to conform to conventions. Several authors have tried to develop this normative idea, for instance the so-called French School of the Economy of Conventions (see, among others, Batifoulier, 2001; Favereau and Lazega, 2002; Orlean, 2004), as two recent reviews have put forward (Biggart and Beamish, 2003; Dequech, 2005). In the following, we present and reformulate this broader notion of convention.

To develop our notion of *general convention*, we start by considering Gilbert's definition of convention. She defines *convention* as 'a jointly accepted principle of action, a group fiat with respect to how one is to act in certain situations' (Gilbert, 1989: 377). The acceptance of this principle, or fiat, is tacit since the principle does not need to be made explicit in order to be followed by a particular population. Moreover, the principle is jointly accepted in the following manner: we accept principle *P*, in so far as I am one of us, I accept principle *P*. In this sense, the knowledge of the principle has to be shared among those who take part in it. Secondly, 'there is a fairly strong analogy between agreements and social conventions' (Gilbert, 1989: 369). In both cases, one cannot argue individually that an agreement or a convention no longer applies, nor can one unilaterally establish a new agreement or convention. In this sense, both concepts have a collective character. For instance, even though I consider that wearing formal clothes is not necessary for me to carry out my daily work, I cannot establish by myself that the convention prescribing wearing formal clothes no longer holds in my workplace. Thirdly, although a convention can be backed up by moral reasons, the normative aspect of conventions does not always coincide with the normative character of any moral argumentation. Thus, whereas moral norms distinguish between 'good' and 'bad' actions, *general conventions* 'provide a basis for judging the appropriateness of acts by self and others' (Biggart and Beamish, 2003: 444). Since *general conventions* are interpreted as a general principle of action, the normativity of general *convention* can be explained in

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

the same way Gilbert does for principles: 'As a matter of logic, if such-and-such is your principle, then, from your own point of view, you have a reason to conform' (Gilbert, 1989: 374).

Despite its suggestive formulation, recent papers have pointed out the limitation of Gilbert's approach to widen the account of conventions proposed by Lewis. Thus, several authors have made the argument that the formalization of a wider notion of rationality is necessary to endow the agents in Lewis' definition of convention with a sufficient reason to coordinate their behavior, without resorting to Gilbert's 'fiats' or 'independent reasons' (see Cubitt and Sugden, 2003; Sillari, 2005). Bicchieri (2006), for instance presents a definition of convention closer to Lewis's definition of social convention than to Gilbert's. Bicchieri's definition needs not posit any 'group fiat', and the reason to act is endogenous to the definition of convention, given that also cognitive capabilities of the agents are taken into account. According to Bicchieri, individuals only need *empirical* or *plain expectations* of the conformity of others to a particular convention to actually conform to this convention. In our proposal of *general convention*, not only do individuals have expectations about what action is going to be followed in a specific situation, but they have expectations about which general principle will be used to make recurrent choices in similar situations. Therefore, subjects are not supposed to make decisions in a case by case manner, nor are they supposed to expect that others do so, but they apply general principles of decision making and expect that others apply these principles as well. In this sense our definition of *general convention* does not rely on individuals' intrinsic motivations; subjects only need to expect that others are following a given principle to actually follow it.

Why is it important to distinguish between a combination of actions (a *specific convention*) and a general principle (a *general convention*)? In her seminal book, Ullmann-Margalit (1977) offers three arguments regarding the usefulness of this type of general coordination principles: first, they are 'capable of regulating and channeling the expectations ... of anonymous participants' (p. 85); second, 'while a regularity extracted from past events might sometimes be continued in more than one way, a [general principle[2]] will provide the principle of continuation which will resolve potential ambiguities in most future events' (p. 87); and, third, 'there is a higher degree of articulation and explicitness associated with a [general principle] than with a mere regularity of behavior' (p. 87).

This notion of *general convention* captures Gilbert's claim that conventions must be understood as general principles of actions. However, rather than interpreting these principles as 'joint commitments', we consider them as 'understandings, often tacit but also conscious, that organize and coordinate

2 She uses the concept of *coordination norm* with a similar meaning to our concept of *general principle* or *general convention*. When presenting her points, we have written *principle* (between brackets) instead of *norm*.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

action in predictable ways. Conventions are agreed-upon, if flexible, guides for economic interpretation and interaction' (Biggart and Beamish, 2003).

In the next section, a *general convention* will be formally defined as a principle of action prescribing how to behave in certain situations. Recall that the only function of this principle is to coordinate individuals' (empirical) expectations in recurrent situations. Assuming this simple definition, we will try to study the interaction of both *general* and *specific conventions* in a game theoretical framework.[3]

3. Game theory framework

This section defines a game theoretical framework that can be useful for analyzing the interaction between the two notions of conventions presented above. To this end, we are going to consider a subclass of 2×2-bimatrix games (two players, each of whom faces two different actions). Four payoff parameters a, b for player B and c, d for player A satisfying $1 > a > b > 0$ and $1 > c > 0 > d > -1$ determine the payoff matrix

		B	
		X_B	Y_B
A	X_A	**1, a**	d, 0
	Y_A	0, b	**c, 1**

with payoffs in the usual order. In addition to the two strict equilibria $X = (X_A, X_B)$ and $Y = (Y_A, Y_B)$, there exists a mixed strategy equilibrium (p^*, q^*) with $p^* = \Pr\{X_A\}$, $q^* = \Pr\{X_B\}$ and payoffs $U_i\,(p^*, q^*)$

$$p^* = \frac{1-b}{1-b+a}, \quad q^* = \frac{c-d}{1+c-d}$$

$$U_A\,(p^*, q^*) = \frac{c}{1+c-d}, \quad U_B(p^*, q^*) = \frac{a}{1-b+a}$$

It is natural (see Harsanyi and Selten, 1988) to begin by first considering the two strict equilibria as solution candidates.[4] Due to the coexistence of X and Y

3 The two notions of conventions presented in this section have been considered, using different terms, by other authors, especially some of those within the so-called French School of the Economic of Conventions. Thus, Favereau (1986) uses 'Convention$_1$' to refer to a 'social or collective representation', and 'Convention$_2$' to refer to an 'inter-individual behavioral pattern'. Batifoulier (2001) distinguish between 'conventional modes of evaluation' (conventions of thought) and 'conventional rules' (behavioral conventions). Finally, Orléan (1999) uses the notion of 'convention of interpretation' to capture the normative or evaluative component of conventions.

4 Note that in this type of game each strict equilibrium represents a *specific convention* in the sense introduced in the preceding section. Thus, each equilibrium implies a different behavioral regularity, which is arbitrary and stable by definition.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

as strict equilibria, this alone would not resolve strategic uncertainty. Thus even this most refined equilibrium idea (Van Damme, 1987) does not help players to solve their coordination problem. The problem becomes even worse when the attractiveness of strict equilibria is not commonly accepted and known. In this case, other criteria as well as other candidates might become relevant.

A selection principle, or *general convention*, would solve the coordination on how to play one of the games described above, e.g. in the sense of recommending a (strict) equilibrium or, more generally, one of the four possible pure strategy vectors. Such coordination, of course, requires that both players are commonly aware which coordination principle should be applied. This requirement is one of the main aspects of the experiment presented below.

To analyze how people use principles when deciding in the class of games presented above, we define a limited number of principles prescribing different behaviors in this class of games. The selection of principles has not been completely arbitrary. These principles represent general patterns of behavior observed in previous experimental papers where subjects confronted games similar to our basic game class. We cite below previous experimental investigations on the use of these principles. In this sense, the principles proposed do not correspond to *real-life* type of principles, but to general criteria suited for the laboratory game that participants are going to play. Therefore, the selection of principles is instrumental to the aim of the experiment: to study how individuals use general principles to coordinate on specific actions.[5] In the following, we describe the four principle selected.

When studying the emergence and evolution of conventions, the most convincing concept seems to be *risk dominance* (R), which, for the type of games considered, is characterized by the three axioms of best reply invariance, isomorphic invariance, and monotonicity (Harsanyi and Selten, 1988).[6] The *risk dominance* principle suggests to play:

- $X = (X_A, X_B)$ if $a > (c - d)(1 - b)$ and
- $Y = (Y_A, Y_B)$ if $a < (c - d)(1 - b)$,

where we will guarantee applicability in the sense of $a \neq (c - d)(1 - b)$ by our parameter choices in the experiment.

5 In addition to the explanation given above, we have evidence on the application of these principles in a previous paper (Büchner *et al.*, 2005). In that paper, the same class of games was used and eight principles were defined and studied both evolutionarily and experimentally. The results of this previous study support the selection presented below. First, all the four principles used here are evolutionarily stable; second, among the eight principles suggested in that previous paper, more than 92% of the experimental subjects selected one of the four principles presented here.

6 We use the standard derivation of the risk dominance principle proposed by Harsanyi and Selten (1988).

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

Next, we will consider *Loss avoidance* (L), which excludes player A's choice of strategy X_A since it, together with Y_B, yields a loss for player A. Loss avoidance plus equilibrium behavior thus suggests the solution (Y_A, Y_B) always.[7]

Sum maximization (S) simply recommends the strategy constellation yielding the largest payoff sum, i.e.:

- (X_A, X_B) if $a > c$ and
- (Y_A, Y_B) if $a < c$,

where we again exclude the degenerate case $a = c$.

Finally, *Alternating* (A) recommends that both strategies alternate, e.g. choosing X_i in the even rounds and Y_i in the odd ones.[8]

This design clearly makes it possible to study the analytical difference and the empirical interaction between both notions of conventions presented above. In relation to the first question, it allows us to distinguish between specific actions and the general principles prescribing these actions using game theoretical concepts. Formalizing the notion of *specific convention* using game theoretical concepts is not problematic, since the three characteristics introduced in Section 2 – regularity, arbitrariness, and stability – are sufficient. On the contrary, the way we have formalized the notion of *general convention* requires a more explicit connection between the theoretical concepts introduced in the previous section and the technical definition presented here. The idea of *general convention* as a general principle prescribing specific behaviors profits from the contributions of Ullmann-Margalit (1977), Gilbert (1989), and Bicchieri (2006). From Ullmann-Margait (1977), we have taken the idea of conventions as cognitive devices to coordinate inter-individual behaviors. We have derived the notion of *general conventions* as shared principles from Gilbert (1989). Finally, we agree with Bicchieri that only plain (empirical) expectations have to be considered to explain how conventions are supported. As stated in Section 2, our notion of general principle is also consistent with the sociological definition of Biggart and Beamish (2003) of conventions as 'agreed-upon, if flexible, guides for economic interpretation and interaction'. The second advantage of our operationalization, which is directly derived from the first one, is that this framework can be easily implemented in the laboratory, as we explain next.

In the next section, the experimental design is described. The aim of the experiment is to study how the shared knowledge of these principles helps subjects to coordinate on *specific conventions*. Thus, in our particular class of games, these principles work as cognitive devices to coordinate behaviors.

7 For previous experimental results regarding *Loss avoidance* as an equilibrium selection principle, see Cachon and Camerer (1996), and Rydval and Ortmann (2005).

8 Using experimental data from some pilot sessions, Schotter and Sopher (2003, n5) argue that the emergence of alternating behavior in a battle-of-the-sexes type of game requires the game to be played repeatedly a sufficient number of times.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

4. Experimental design and procedure

For the experiment, 96 undergraduate students at Jena University (Germany) were recruited (from different disciplines) using ORSEE (Greiner, 2004). Three experimental sessions (each session corresponds to a different treatment)[9] were conducted using a different group of 32 participants in each one. The instructions for the experiment are given in Appendix A. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007) at the computer laboratory of the Max Planck Institute of Economics. Subjects received written instructions, which were also read aloud by the experimenter to guarantee their common understanding. No communication other than through the use of principles (as we will explain below) was permitted. Subjects could not identify which members of their matching group they were actually interacting with. At the end of every experimental session, subjects were paid in cash according to their payoff in the game (plus a show-up fee of 2.5 €).

Each subject was assigned a role (A or B), but they only learnt their own role at the end of the experiment after having made all the decisions. Note that this procedure created a kind of 'veil of ignorance' in which everyone knows the structure of the game, but no one knows the players' roles in one particular game. This feature of the experiment is crucial to interpret the results obtained. The veil of ignorance was purposely used to force the players to face the game from a collective point of view, neglecting the individual differences in the structure of the game. This procedure allows us to concentrate on how participants use general principles to coordinate on specific actions, without caring for the role asymmetries of the underlying games. In fact, the veil of ignorance turns the whole game symmetric. Despite this, two principles still rely on the asymmetric structure of the original game: *Alternating* suggests alternate between both strict equilibria (each of them preferred by each of the roles), and *Loss avoidance* suggests one of the two roles playing always the same strategy. Although the participants could not choose according to a role specific structure of incentives, they could still have preferences over the best collective solution to an asymmetric coordination problem, i.e. every participant could select the principle that, according to herself, establishes an appropriate pattern of recurrent behavior.

The exact procedure followed by subjects was as follows. In all three treatments, every participant interacted in a group of four subjects playing 20 single games. There were ten different games (see Appendix B), each of which were played twice. In the first ten games (1–10), each participant played with one partner, while they played with another partner during the second ten games (11–20). This procedure of changing partners is

9 Subjects were randomly allocated to the three sessions in order to avoid confounding treatment and session effects.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

designed to test participants' consistency in using principles. The treatments differed in terms of the information that the participants were given regarding principles.

In Treatment B (baseline treatment), no information on principles was provided. Therefore, the participants played the 20 games solely with information about the game parameters (see Treatment B's computer screen in Appendix C).

In Treatment P (principles treatment), the participants had to fulfill two tasks. First, they had to rank the principles we introduced in Section 3 and, second, they had to play 20 successive games as in Treatment B. The first task involved filling in a table with the proposed principles (see Ranking screen in Appendix C). The participants did not have to suggest a complete order, but were only required to select at least one principle. Once all the subjects had given their ranking, they were matched with a partner with whom they played the ten rounds. After these ten rounds, they were re-matched with another partner from their matching group and played another ten rounds. During the 20 periods, the only information they knew, apart from the game parameters, was their own ranking.

Treatment D (dictator treatment) only differed from Treatment P in one respect. In this treatment, after everyone had made their ranking and had been matched with a partner, one of the two rankings was randomly selected by the computer (either A's ranking or B's ranking). After this ranking was shown to both players (see Treatment D's decision-making screen in Appendix C),[10] they played ten single-games. The players were subsequently re-matched with other partners, the random dictator mechanism was used again and another ten rounds were played. The same protocol was followed in every session. The treatments only differed in terms of the variable that was being manipulated: the possibility of using principles to coordinate expectations that prescribed how to coordinate behavior (see Table 1).

5. Experimental results

In this section we shall mainly attempt to answer three different questions. Firstly, do subjects actually follow general principles when deciding in coordination games? Secondly, is behavior consistent with first rank principles? Thirdly, does using principles improve coordination? To answer the first and the second questions we will analyze the treatments independently. To answer the third question we will compare results across treatments.

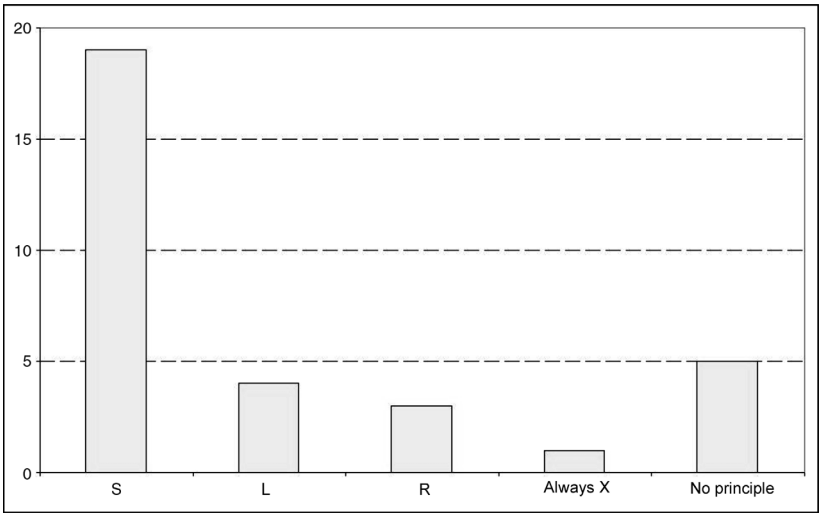
10 This is a so-called *random dictator mechanism*. One of the main advantages of using this mechanism is that it is incentive compatible, that is since there is a 50% probability that your ranking will be selected, you have a strong incentive to reveal your actual preferences about principles. However, the communication of the randomly selected ranking is *cheap talk*, in the sense that subjects may deviate from their ranking, although this implies going against their own interest.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

Table 1. Characteristics of the treatments

	Treatment B	Treatment P	Treatment D
Number of sessions	1	1	1
Periods per session	20	20	20
Matching groups	8	8	8
Participants	32	32	32
Treatment variable	None	Principles	Principles + Dictator
Coordination	75%	65.31%	80.94%
Average earnings	8.8€	8.1€	9.3€

Figure 1. Principles actually followed in Treatment B



Baseline treatment

In Treatment B, the baseline treatment, participants were not induced to use any general principle. However, as can be observed in Figure 1, the behavior of more than half of the participants was consistent with principle *S*, i.e. 19 out of 32 participants repeatedly chose the more efficient outcome. In addition, four individuals followed principle *L*, three individuals followed principle *R*, and one individual always chose strategy *X*. No individuals alternated between strategies (principle *A*) and the behavior of five subjects cannot be classified according to any principle.

Before continuing with the results, we should clarify how we define that a certain pattern of behavior is consistent with one or other of the principles. We assume that following a particular principle means that behavior is consistent with this principle at least 18 out of 20 times. According to our game parameters, if individual behavior is consistent with a principle 18 times, it cannot be consistent with another principle at the same time. This is the threshold that

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

we use to determine if one participant is consistent with the prescription derived from one of the principles.[11]

The predominance of efficiency-based behavior is consistent with the results of a previous paper (Büchner *et al.*, 2005) using the same class of games, and with previous experimental literature on the Stag Hunt Game (see, for instance, Rankin *et al.*, 2000). The use of the 'veil of ignorance' may help to explain why subjects, adopting a collective rather than individual point of view, prefer principle *S*, although this preference is not universal.

Result 1. When no coordination principle was suggested, 19 out of 32 subjects (59.4%) followed an efficiency criterion.

The incidence of *Risk-dominance* (as found by Van Huyck *et al.*, 1990, 1991; Van Huyck *et al.*, 1997) or *loss-avoidance* (as suggested by Cachon and Camerer, 1996 and Rydval and Ortmann, 2005) is minimal. There is no evidence of the use of any kind of alternating behavior.

Principles treatment

In Treatment P, subjects had to rank the principles but did not communicate their rankings to their partner. This treatment was designed to test if after ranking the principles, individuals felt obliged by their first-rank principles. Therefore, we will now consider the two following aspects: which principles were ranked first, and was behavior consistent with these principles?

The dark bars in Figure 2 represent the frequency of first-ranked principles. Of the four principles, the most frequent option was *Sum maximization* (*S*), followed by *Loss avoidance* (*L*), *Alternating* (*A*) and *Risk dominance* (*R*). On the contrary, the light bars plot the frequency of principles that were actually followed by the participants. As we can observe, not only was *S* the most frequently selected first-ranked principle, but was also the most widely followed principle. In fact, the results of Treatment P do not differ largely from the first treatment, as 17 instead of 19 out of 32 subjects followed the efficiency principle. Moreover, those who selected principle *S* showed a high level of consistency with the principle.

All but one of the subjects who ranked first principle *S* actually followed it. However, only four out 16 people who ranked any other principle first were consistent with their first-ranked principle. Consistency with principle *S* is statistically different from consistency with other principles (Mann–Whitney U-test: $z = -3.897$, $p = 0.001$).

Result 2. Consistency with principles critically depends on the principle that has been selected. Consistency with efficiency criteria is especially high.

11 The numerical parameter values actually used in the experimental sessions are presented in Appendix B. There were ten different games and participants had to play them twice. The definition of the principles takes into account the 20 games finally played.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

Figure 2. First-ranked principles and principles actually followed in Treatment P

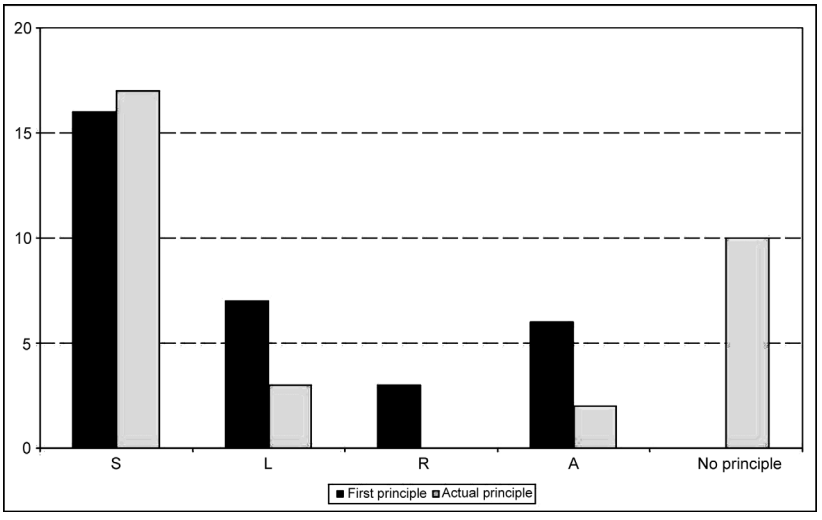
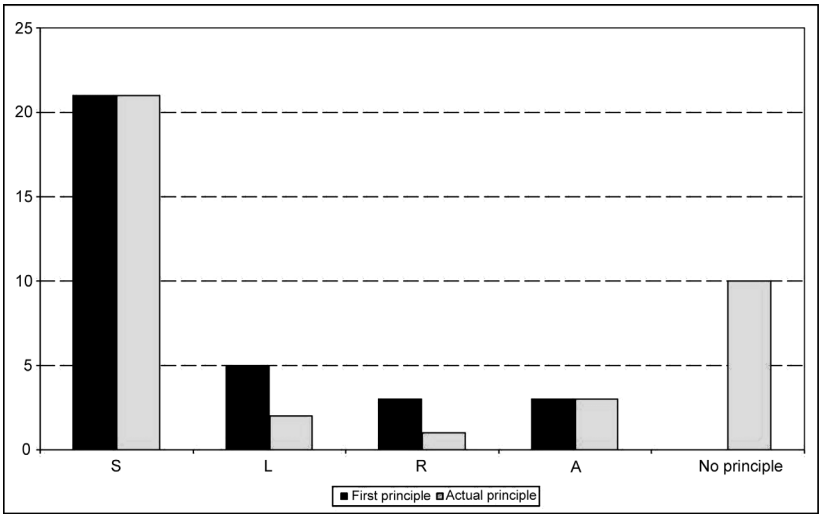


Figure 3. First-ranked principles and principles actually followed in Treatment D



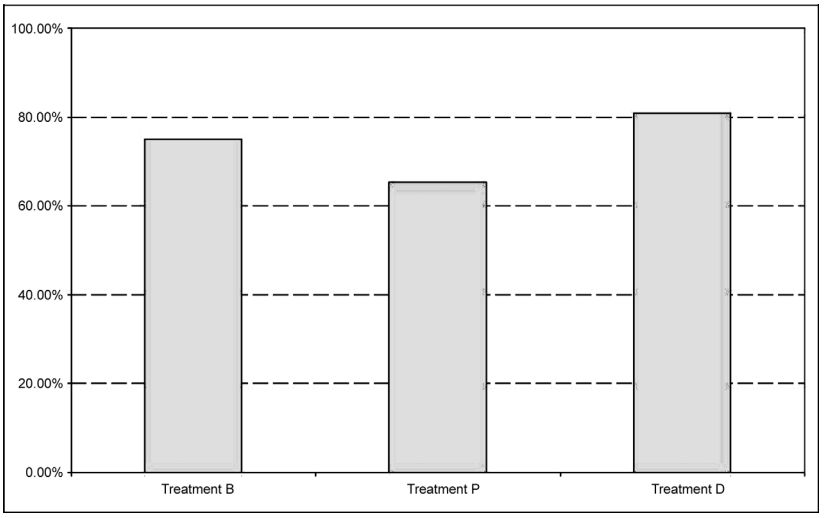
Dictator treatment

Figure 3 plots the first-ranked principles and the principles actually followed in Treatment D. Again, *Sum maximization* (21 subjects) was the most frequently first-ranked principle, followed by *Loss avoidance* (five subjects), *Risk dominance* (three subjects), and *Alternating* (three subjects).

Once we have verified that *Sum Maximization* plays a continuously predominant role even in a strategic environment for choosing principles, we question if participants' behavior was consistent with the random dictator's

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

Figure 4. Coordination by treatment



first-ranked principle. Here we have to consider 64 different behaviors because every subject made decisions in light of the two rankings chosen by the random dictator (one during the first ten rounds and one during the second ten rounds). Forty-four out 64 behaviors were consistent with the random dictator's first principle, eight out 64 deviate from the dictator's principle to follow principle *S*, and 12 behaviors cannot be classified. In general, participants were consistent with the dictator's prescription and the only reason for deviating from it was due to an efficiency criterion. The result is that when participants could rank principles strategically, they mostly ranked first principle *S* and they were mostly consistent with it.

Result 3. The possibility of ranking principles strategically does not alter the predominance of and the consistency with efficiency criteria.

Coordination

How did the behaviors described above affect coordination? Figure 4 plots the coordination levels in the three treatments. The mean coordination level in this experiment was 68.54%. There are significant differences between treatments (Kruskal–Wallis test; $X^2 = 12.667$, d.f. $= 2$, $p = 0.047$) regarding coordination levels.

Participants did relatively well in the baseline treatment where no principles were induced by the experimenter; achieving quite a high average coordination rate of 75%. However, the coordination rate dropped by more than ten points to 65.31% in Treatment *P* where the four principles were suggested. Furthermore,

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

the coordination level increased to 80.94% only when rankings were commonly known (Treatment *D*).[12]

According to our results, for a principle to work both players have to be commonly aware that this principle should be applied. Thus, whereas the introduction of principles in Treatment *D* improves the general coordination level, the introduction of principles in Treatment *P* does not.[13] This is clear and it is at the root of our proposal about how general principles facilitate coordination in *specific conventions*. However, this does not explain why coordination in Treatment *P* is significantly lower than in the baseline treatment. The reason for this is that the subjects in Treatment *B* mostly followed an efficiency criterion and, in so doing, they quite often coordinate on the efficient convention. Therefore, efficiency concerns were focal in this first treatment. However, when other principles were suggested, the general behavior was not as concentrated on an efficiency criterion and half of the subjects chose principles other than *S*. The implementation of these other principles was disastrous (only four out of 16 subjects implemented the principle correctly). As a result, in general, the subjects could not benefit from a highly spontaneous coordination on the efficiency criterion.

Result 4. In order for these principles to work, it is indispensable that they be commonly known and adopted.

6. Discussion

At least since the work by Lewis (1969), economic theorists have tried to give an adequate account of social conventions. However, this attempt has failed to incorporate one of the most important features of this concept: its normative character. In this paper we have tried to build a game theory framework approach to this normative character. This framework consists of two notions of conventions, namely *specific conventions*, which are arbitrary but stable social regularities, and *general conventions*, which are principles of action prescribing how to behave in certain types of situations. Following the economic literature on conventions, *specific conventions* have been represented in this paper as stable solutions to a particular type of coordination game. *General conventions* have been represented as a set of *normative sentences* indicating which *specific convention* to choose. This framework has already been analyzed evolutionarily

12 A potential interpretation of this result is that participants use the revealed first-principle as an advice. In doing so, they create the kind of expectations that lead to coordination in the set of actions prescribed by the principle.

13 Coordination in Treatment D is significantly higher than in Treatment P (Mann–Whitney U-test $z = -2.265$, $p = 0.024$).

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

and experimentally in a previous paper (Büchner *et al.*, 2005). In light of these previous results, this new paper has tried to improve upon our understanding of the relation between *specific* and *general conventions*. The main results of the experiment are:

- Even when principles are not suggested, participants mainly follow an efficiency behavior in the class of games considered here.
- This result is robust to the suggestion of other principles as well as the possibility of using principles strategically.
- Consistency with principles depends critically on the principle chosen. Participants are more consistent with the efficiency principle.
- For principles to work, they need to be commonly known and adopted.

The experiment reveals that *specific conventions* are established quite well when subjects play a sufficiently large number of rounds in the same class of games. And, in this sense, the paper is in line with previous experimental literature on the emergence of conventions in two-person coordination games (see, for instance, Rankin *et al.*, 2000). However, *general conventions* are more difficult to achieve, and they are only well established when the general convention is commonly known and adopted. Furthermore, general conventions only help when they lead to the choice of the specific convention that would prevail in the absence of any general convention. Specifically, some participants tried to establish a different convention to the predominant one (efficiency), but this behavior leads to disastrous results in terms of coordination.

These results suggest that the sociological notion of convention is more difficult to observe and apply for different reasons. First, different principles are salient to different individuals, preventing the spontaneous emergence of a univocal principle of behavior. Second, principles may imply actions that people do not understand. In fact, simplicity may be a crucial characteristic to evaluate the popularity of a principle. Third, when general principles that contradict the underlying pattern of behavior – in this experiment efficiency – are suggested, they may indeed prevent coordination. In this sense, individuals have to be aware of the common principle to implement it successfully.

This is the second paper of a research program that focuses on the use of general principles when deciding in asymmetric coordination games. Since the results presented above are quite consistent with the ones reported by Büchner *et al.* (2005), we consider that analyzing the use of principles, testing the consistency of behaviors according to every single principle, and evaluating the results of applying these principles are crucial to our understanding of how people actually play coordination games. In doing so, not only should we understand the behavioral patterns we actually observe, but also the general criteria that support these patterns.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

References

Batifoulier, Philippe (2001), *Théorie des conventions*, Paris: Economica.
Bicchieri, Cristina (2006), *The Grammar of Society*, Cambridge: Cambridge University Press.
Biggart, Nicole W. and Beamish Thomas D. (2003), 'The Economic Sociology of Conventions: Habit, Custom, Practice, and Routine in Market Order', *Annual Review of Sociology*, **29**: 443–464.
Binmore, Ken (1992), *Fun and Games: A Text on Game Theory*, Lexington: D.C. Health.
Binmore, Ken (1998), *Just Playing. Game Theory and the Social Contract*, Cambridge: The MIT Press.
Binmore, Ken (2005), *Natural Justice*, Oxford: Oxford University Press.
Büchner, Susane, Werner Güth and Luis M. Miller (2005), 'Conventions for Selecting among Conventions: An Evolutionary and Experimental Analysis', *Discussion Papers on Strategic Interaction*, Jena: Max Planck Institute of Economics.
Cachon, Gerard P. and Colin Camerer (1996), 'Loss-avoidance and Forward Induction in Experimental Coordination Games', *Quarterly Journal of Economics*, **111**: 165–194.
Clark, Kenneth, Stephen Kay, and Martin Sefton (2000), 'When are Nash Equilibria Self-Enforcing? An Experimental Analysis', *International Journal of Game Theory*, **29**: 495–515.
Coleman, James (1990), *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.
Cubitt, Robin P. and Robert Sugden (2003), 'Common Knowledge, Salience and Convention: A reconstruction of David Lewis' Game Theory', *Economics and Philosophy*, **19**: 175–210.
Dequech, David (2005), 'Cognition and Valuation: Some Similarities and Contrasts between Institutional Economics and the Economics of Convention', *Journal of Economic Issues*, **39**: 465–473.
Harsanyi, John C. and Reinhard Selten (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge, MA: The MIT Press.
Hume, David (1975 [1777]), *An Enquiry Concerning the Principles of Morals*, Oxford: Oxford University Press.
Hume, D. (1978 [1740]), *A Treatise on Human Nature*, Oxford: Oxford University Press.
Favereau, Olivier (1986), 'La formalisation du rôle des conventions dans l'allocation des ressources', in Robert Salais y. Laurent Thévenot (eds), *Le Travail Marchés, règles, conventions*, Paris : Economica.
Favereau, Olivier and Emmanuel Lazega (eds) (2002), *Conventions and Structures in Economic Organization*, Cheltenham: Edward Elgar.
Fischbacher, Urs (2007), 'z-Tree: Zurich Toolbox for Ready-made Economic Experiments', *Experimental Economics*, **10**: 171–178.
Gilbert, Margaret (1989), *On Social Facts*, London: Routledge.
Greiner, Ben (2004), 'An Online Recruitment System for Economic Experiments', in K. Kremer and V. Macho (eds), *Forschung und wissenschaftliches Rechnen 2003: GWDG Bericht 63*, Göttingen: Ges. für Wiss.
Lewis, David (1969), *Conventions: A Philosophical Study*, Oxford: Basil Blackwell.
Orléan, André (1999), *Le pouvoir de la finance*, Paris: Odile Jacob.
Orléan, André (2004), *Analyse économique des conventions*, Paris: Quadrige/Presses Universitaires de France.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

Rankin, Frederick W., John B. Van Huyck, and Raymond C. Battalio (2000), 'Strategic Similarity and Emergent Conventions: Evidence from Similar Stag Hunt Games', *Games and Economic Behavior*, **32**: 315–337.
Rydval, Ondrej and Andreas Ortmann (2005), 'Loss Avoidance as Selection Principle: Evidence from Simple Stag-hunt Games', *Economics Letters*, **88**: 101–107.
Schotter, Andrew (1981), *The Economic Theory of Social Institutions*, Cambridge: Cambridge University Press.
Schotter, Andrew and Barry Sopher (2003), 'Social Learning and Coordination Conventions in Intergenerational Games: An Experimental Study', *Journal of Political Economy*, **111**: 498–529.
Sillari, Giacomo (2005), 'A Logical Framework for Convention', *Synthese*, **147**: 379–400.
Skyrms, Brian (1996), *The Evolution of the Social Contract*, Cambridge: Cambridge University Press.
Sugden, Robert (1986), *The Economics of Welfare, Rights and Co-operation*, Oxford: Basil Blackwell.
Sugden, Robert (1989), 'Spontaneous Order', *Journal of Economic Perspectives*, **3**: 85–97.
Sugden, Robert (1998), 'Conventions', in P. Newman (eds), *The New Palgrave Dictionary of Economics and the Law*, New York: Palgrave.
Ullmann-Margalit, Edna (1977), *The Emergence of Norms*, Oxford: Oxford University Press.
Van Damme, Eric (1987), *Stability and Perfection of Nash Equilibria*, London: Springer.
Van Huyck, J., R. C. Battalio, and R. O. Beil (1990), 'Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure', *American Economic Review*, **80**(1), 234–248.
Van Huyck, John, Raymond C. Battalio, and Richard O. Beil (1991), 'Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games', *Quarterly Journal of Economics*, **106**: 885–911.
Van Huyck, John, Raymond C. Battalio, and Frederick W. Rankin (1997), 'On the Origin of Conventions: Evidence from Coordination Games', *The Economic Journal*, **107**: 576–596.
Vanderschraaf, Peter (1998), 'Knowledge, Equilibrium and Convention', *Erkenntnis*, **49**: 337–369.
Weber, Max (1978 [1922]), *Economy and Society*, Berkeley: University of California Press.
Young, H. Peyton (1996), 'The Economics of Convention', *Journal of Economic Perspectives*, **10**: 105–122.

Appendix A: Experimental instructions[14]

General instructions (common to all treatments)

Welcome and thank you for participating in this experiment. Please read the instructions carefully. From now on, any communication with the other participants is forbidden. Otherwise, we will have to exclude you from the experiment and you will receive no payment for your participation. If you have any questions, please raise your hand. We will come to you and answer your questions individually.

During the experiment, we use ECUs (experimental currency units) instead of Euros. Your experimental income will be calculated in ECUs. At the end of the

14 This is a translation of the original instructions in German.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

experiment, the ECUs you have earned will be converted to Euros (1 ECU = 0.5 Euro) and the amount you obtain will be paid to you in cash.

In this experiment, you will interact with one other participant. Each of you will play a different role which, to simplify matters, is either A or B. Both participants can choose between two strategies: X_A or Y_A (for participant A) and X_B or Y_B (for participant B). However, you will only know which role has been assigned to you at the end of the experiment after making all your decisions.

These strategies lead to the following payments:

Payment for participant A	Payment for participant B	If participant	
		A selects	B selects
1	A	X_A	X_B
D	0	X_A	Y_B
0	B	Y_A	X_B
C	1	Y_A	Y_B

Values a, b, c and d will fulfill the following conditions

$$1 > a > b > 0 \quad \text{and} \quad 1 > c > 0 > d > -1$$

Whether X_i or Y_i is better for you depends on your partner's decision, e.g. which alternative she selects. You should therefore be interested in collectively arranging a strategy constellation (X_A, X_B), (X_A, Y_B), (Y_A, X_B), or (Y_A, Y_B).

Specific instructions for Treatment B

You can choose either strategy X or strategy Y. In each round, you will learn the values of the parameters a, b, c and d. Then, you will play another nine rounds of the game (i.e. you play a total of ten rounds). In each round, the values of parameters a, b, c and d will change.

After these ten rounds, you will be randomly matched up with a new partner with whom you will interact for another ten rounds. In each round, you will again learn the values of parameters a, b, c and d. At the end of these 20 rounds, we will inform you about the values of parameters a, b, c and d in the individual rounds, whether you have been participant A or B in the individual rounds and about your payoff in each round.

To compensate for possible losses, each participant will receive a basic fee of 5 ECUs (2.5 Euros). Should you finish the experiment with a negative balance, we will ask you to compensate for this by doing some odd jobs (for 32 ECUs per hour) at our institute. If you don't agree, please raise your hand. You may then leave the experiment.

Before the experiment begins, we will ask you to correctly answer a number of questions aimed at improving your understanding of the experiment. Please fill in the questionnaire thoroughly. Once you have finished, an experimenter will come round to collect the questionnaire.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

Specific instructions for Treatment P

To reach such an agreement, we propose principles by which you can select one of the four strategy constellations. (These principles will be explained to you on the next page.)

First, you have to rank the principles. It is not necessary (but you may) to rank all the principles. It is sufficient enough to indicate your priority. Then, knowing your ranking, you can choose either strategy *X* or strategy *Y*. In each round, you will learn the values of parameters *a*, *b*, *c* and *d*. Then, you will play another nine rounds of the game (i.e. you play a total of ten rounds). In each round, the values of parameters *a*, *b*, *c* and *d* will change. Your selected strategy rank order will remain the same.

After these ten rounds, you will be randomly matched up with a new partner with whom you will interact for another ten rounds. You will keep your strategy rank order. In each round, you will again learn the values of parameters *a*, *b*, *c* and *d*. At the end of these 20 rounds, we will inform you about the values of parameters *a*, *b*, *c* and *d* in the individual rounds, whether you have been participant *A* or *B* in the individual rounds and about your payoff in each round.

To compensate for possible losses, each participant will receive a basic fee of 5 ECUs (2.5 Euros). Should you finish the experiment with a negative balance, we will ask you to compensate for this by doing some odd jobs (for 32 ECUs per hour) at our institute. If you don't agree, please raise your hand. You may then leave the experiment.

We will now present to you the principles for selecting one of the four strategy constellations:

A

In even rounds, we play *X* and in uneven rounds *Y* so that both participants rely either on their *X*, resp. *Y* choice.

L

To avoid losses, we play *Y*, i.e. Y_A, Y_B.

R

We use the strategy constellation where each participant reacts optimally to the choice of the other and the product of the corresponding payoff increases resulting from a strategy change is maximal.

For participant A, the payoff increases by $1 - 0 = 1$ for X_B and $c - d$ for Y_B. For participant B the increase is $a - 0 = a$ for X_A and $1 - b$ for Y_A. As only for (X_A, X_B) and (Y_A, Y_B) both react optimally to the choice of the other, participants should use (X_A, X_B) if $a > (c - d)(1 - b)$ and (Y_A, Y_B) if $a < (c - d)(1 - b)$.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

S

We use the strategy constellation for which the payoff sum is maximal, i.e. participants play (X_A, X_B) if $a > c$ and (Y_A, Y_B) if $a < c$.

Before the experiment begins, we will ask you to correctly answer a number of questions aimed at improving your understanding of the experiment. Please fill in that questionnaire thoroughly. Once you have finished, an experimenter will come round to collect the questionnaire.

Specific instructions for Treatment D

To reach such an agreement, we propose principles by which you can select one of the four strategy constellations. (These principles will be explained to you on the next page.)

First, you have to rank the principles. It is not necessary (but you may) to rank all the principles. It is sufficient enough to indicate your priority. Then, one ranking will be randomly selected by the computer (either your ranking or your partner's ranking) and this ranking will be shown to your partner and to you. After that, you can choose either strategy *X* or strategy *Y*. In each round, you will learn the values of the parameters *a*, *b*, *c* and *d*. Then, you will play another nine rounds of the game (i.e. you play a total of ten rounds). In each round, the values of parameters *a*, *b*, *c* and *d* will change. The selected ranking will remain the same.

After these ten rounds, you will be randomly matched up with a new partner with whom you will interact for another ten rounds. One of the two rankings will be randomly selected again (either your ranking or your new partner's ranking). In each round, you will again learn the values of parameters *a*, *b*, *c*, and *d*. At the end of these 20 rounds, we will inform you about the values of parameters *a*, *b*, *c*, and *d* in the individual rounds, whether you have been participant A or B in the individual rounds and about your payoff in each round.

To compensate for possible losses, each participant will receive a basic fee of 5 ECUs (2.5 Euros). Should you finish the experiment with a negative balance, we will ask you to compensate for this by doing some odd jobs (for 32 ECUs per hour) at our institute. If you don't agree, please raise your hand. You may then leave the experiment.

We will now present to you the principles for selecting one of the four strategy constellations:

(The principles description, which is similar to the one presented in Treatment P, is introduced here again.)

Before the experiment begins, we will ask you to correctly answer a number of questions aimed at improving your understanding of the experiment. Please fill in the questionnaire thoroughly. Once you have finished, an experimenter will come round to collect the questionnaire.

https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

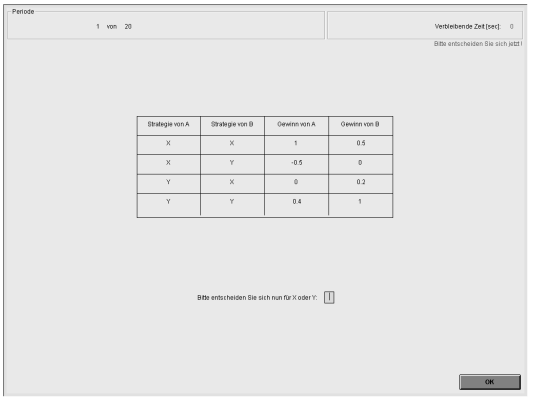
Appendix B: Games actually played

		Payoffs				Payoffs	
Rounds	Behavior	Payoffs A	Payoffs B	Rounds	Behavior	Payoffs A	Payoffs B
Period 1	X_A, X_B	1	0.5	Period 2	X_A, X_B	1	0.8
	X_A, Y_B	−0.5	0		X_A, Y_B	−0.1	0
	Y_A, X_B	0	0.2		Y_A, X_B	0	0.1
	Y_A, Y_B	0.4	1		Y_A, Y_B	0.7	1
		Payoffs				Payoffs	
Rounds	Behavior	Payoffs A	Payoffs B	Rounds	Behavior	Payoffs A	Payoffs B
Period 3	X_A, X_B	1	0.9	Period 4	X_A, X_B	1	0.7
	X_A, Y_B	−0.3	0		X_A, Y_B	−0.2	0
	Y_A, X_B	0	0.8		Y_A, X_B	0	0.6
	Y_A, Y_B	0.6	1		Y_A, Y_B	0.5	1
		Payoffs				Payoffs	
Rounds	Behavior	Payoffs A	Payoffs B	Rounds	Behavior	Payoffs A	Payoffs B
Period 5	X_A, X_B	1	0.6	Period 6	X_A, X_B	1	0.3
	X_A, Y_B	−0.4	0		X_A, Y_B	−0.8	0
	Y_A, X_B	0	0.5		Y_A, X_B	0	0.2
	Y_A, Y_B	0.7	1		Y_A, Y_B	0.6	1
		Payoffs				Payoffs	
Rounds	Behavior	Payoffs A	Payoffs B	Rounds	Behavior	Payoffs A	Payoffs B
Period 7	X_A, X_B	1	0.7	Period 8	X_A, X_B	1	0.5
	X_A, Y_B	−0.6	0		X_A, Y_B	−0.1	0
	Y_A, X_B	0	0.2		Y_A, X_B	0	0.2
	Y_A, Y_B	0.9	1		Y_A, Y_B	0.5	1
		Payoffs				Payoffs	
Rounds	Behavior	Payoffs A	Payoffs B	Rounds	Behavior	Payoffs A	Payoffs B
Period 9	X_A, X_B	1	0.2	Period 10	X_A, X_B	1	0.3
	X_A, Y_B	−0.2	0		X_A, Y_B	−0.2	0
	Y_A, X_B	0	0.1		Y_A, X_B	0	0.2
	Y_A, Y_B	0.5	1		Y_A, Y_B	0.6	1

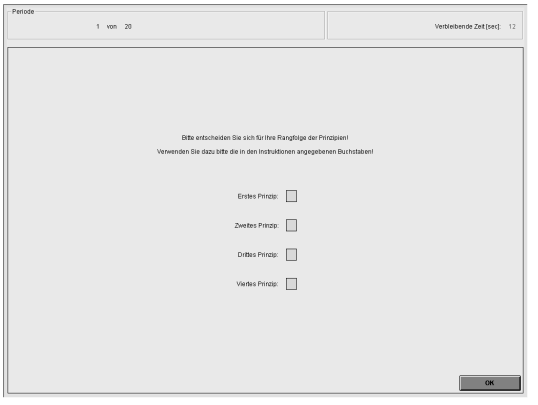
https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press

Appendix C: Computer screens

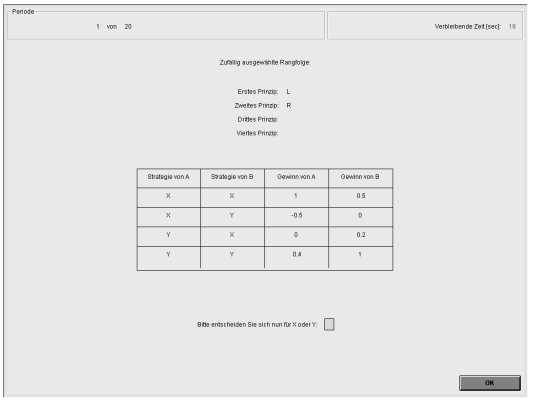
Treatment B decision-making screen



Ranking screen



Treatment D decision-making screen



https://doi.org/10.1017/S1744137408001112 Published online by Cambridge University Press