EconomEtica

Centro interuniversitario per l'etica economica
e la responsabilità sociale di impresa
promosso dalla Fondazione Italiana Accenture

N.60 September 2016

Giacomo Degli Antoni,
Marco Faillo,
Pedro Francés-Gómez,
Lorenzo Sacconi

Distributive Justice with Production and the Social Contract. An Experimental study

Working papers

Distributive Justice with Production and the Social Contract.

An Experimental study

Giacomo DegliAntoni[1,4], Marco Faillo[2], Pedro Francés-Gómez[3], Lorenzo Sacconi[2,4]

Abstract

Drawing on the theoretical and experimental literature on distributive justice, we put some assumptions of the contractarian argument to an empirical test by means of an experiment which investigates the influence that explicit agreement under the veil of ignorance may have on individuals' conception of justice and its implementation in a context of the production and distribution of a common output. One crucial characteristic of our experiment is that subjects are assigned unequal endowments for which they are not responsible; the assignment is random. At the same time, their work naturally generates unequal levels of earnings.

Do the subjects involved in this interaction distinguish between the two types of inequality? Do they try to reduce the arbitrary one, while accepting the one generated through effort? Do they elaborate other distributive criteria? Does their choice ex-ante, when they are behind the veil, differ from their choice ex-post once the veil has been lifted and they know the outcome of the production phase? The main result is that the agreement under a veil of ignorance induces subjects to accept a liberal egalitarian division rule not only in the ex-ante agreement, but also in the actual implementation of the pie division, even if this contradicts their self-interest and some common economic assumptions about reciprocal expectations of rationality. In addition, our results show that deliberating through open discussion increases the level of ex-post compliance.

Keywords: Distributive justice, social contract, fairness, dictator game, contractarian business ethics.

JEL Classification:C72, C91, D02, D63

1. Department of Law, University of Parma, via Università 12, Parma, Italy
2. Department of Economics and Management, University of Trento, via Inama 5, 38122, Trento, Italy.
3. Department of Philosophy I, Faculty of Philosophy, Campus de la Cartuja 18071 Granada, Spain.
4. EconomEtica, c/o University Milano-Bicocca, via Bicocca degli Arcimboldi 8, 20126 Milan, Italy

Acknowledgement. The authors gratefully acknowledge the financial support from the General Directorate of Scientific and Technical Research, Spanish Ministry of Economy (Project BENEB FFI2011-29005).

Distributive Justice with Production and the Social Contract. An Experimental study

1. Introduction

The aim of this paper is to provide experimental evidence on the influence that explicit agreement under the veil of ignorance may have on individuals' conception of justice and its implementation in a context of production.

Drawing on the experimental literature on distributive justice, we put some theoretical assumptions of the contractarian argument to an empirical test by means of an experiment in which individuals play for real money in a laboratory situation that models the idea of an agreement behind a veil of ignorance, followed by implementation of the agreed distributive rule(s) in a context of decision in the absence of coercive authority. One crucial characteristic of our experiment is that subjects are assigned unequal endowments for which they are not responsible; the assignment is random. At the same time, their work naturally generates unequal levels of earnings.

Do the subjects involved in this interaction distinguish between the two types of inequality? Do they try to reduce the random one, while accepting the one generated through work? Do they elaborate other distributive criteria? Does their choice ex-ante, when they are ignorant of whether chance will favor them in the distribution of endowments, differ from their choice ex-post, once they know both their initial luck and their relative performance?

In fact, the main result that we obtain is that, in a context of division of the output of a real production, the experimental simulation of the agreement under a veil of ignorance on principles of justice induces agents to accept a *liberal egalitarian* division rule not only in the ex-ante agreement but also in the actual implementation of the pie division, after real production activity has been carried out. And they do so even if this contradicts their self-interest and some common economic assumptions about reciprocal expectations of rationality.

Our experiment is relevant to two strands of literature. First, it contrasts with the experimental literature on distributive justice as it has evolved to date. On the one hand, studies inspired by Rawls's principles of justice, in fact, focus almost exclusively either on the relevance of the difference principle (Brickman, 1977; Yaari and Bar-Hillel, 1984; Frohlich et al. 1987; Frohlich and Oppenheimer 1990 and 1992; Bond and Park, 1991; Lissowski, Tyszka, Okrasa, 1991; Jackson and

Hill; 1995; Michelbach et al. 2003; De la Cruz-Doña and Martina, 2000) or on the effects of the choice behind the veil of ignorance on stated preferences for redistribution (Anderson and Lyttkens 1999; Traub et al. 2005; Herne and Suojanen, 2004; Herne and Mard, 2008; Schildberg-Hörisch, 2010; Durante, Putterman and van der Weele, 2014). No study (except Faillo *et al.* 2015) has taken into consideration Rawls' concept of "sense of justice" and its role in the solution of the problem of ex-post compliance with the principle chosen behind the veil. In the typical experiment, once the principle has been chosen, it is automatically implemented. Considering the subset of studies that explicitly implement the choice among alternative principles, participants usually choose individually as they are confronted with hypothetical scenarios (Scott et al., 2001; Konow, 1996, 2001, 2003). To our knowledge, only Frohlich et al. (1987), and Faillo et al. (2015) have implemented agreement among participants as a way to choose the principles. In our experiment we take some steps forward with respect to this literature by: i) introducing an explicit (non-binding) agreement behind a veil of ignorance – in fact, we compare two alternative agreement procedures; ii) introducing a production phase in which the outcome depends on both luck and individual effort; and iii) studying the determinants of individuals' willingness to comply, ex-post, with the agreed principle.

Secondly, our study gives more substance to the 'social contract approach' to business ethics. With this expression we refer to the so-called Social Contract tradition in Business Ethics (Wempe 2005), as most notably represented by Donaldson (1982), Donaldson and Dunfee (1999), Phillips (2003), Evan and Freeman (1988), Freeman (1994), Bishop (2008), Lütge (2012, 2015). More specifically, however, we draw on Sacconi's view of the firm as an institution that can be normatively reconstructed as the result of a *constitutional* agreement among its essential stakeholders (Sacconi 2000, 2006, 2011a,b). This view is in line with Freeman's (1984) stakeholder perspective (see also Evan and Freeman 1993; Freeman *et al.*2011), and with Aoki's (1984, 2010) cooperative theory of the firm. For its theoretical-normative core, it rests on Binmore's (2005) re-casting of Rawlsian contractarianism (Rawls 1971), and Buchanan's (1975) logical reconstruction of constitutional and post-constitutional contracts. The basic idea is that the normative basis of the firm can be conceived as the result of a unanimous hypothetical agreement among all the firm's constitutive stakeholders. The content of the agreement is the set of principles and the governance structure that define a mutually beneficial cooperative venture under the rules of a market economy. In this experimental work, we inquire whether real-life subjects, who participate in a situation involving the relevant intuitions of justice concerning the division of the pooled output of a productive activity, actually

follow principles consistent with the social contract perspective in business ethics as outlined above. Obviously, we do not claim that our experiment represents all the complexities of the firm's productive activity, and its related governance structure. Only a stylized representation of some features of it can be subjected to laboratory experimentation. Nevertheless, with respect to the theoretical literature on the social contract of the firm, our experimental study adds an element of realism (Francés-Gómez *et al.* 2015). We inquire empirically whether real-life agents, when involved in an experimental simulation of the Rawlsian "veil of ignorance" ensuringthat they assess principles of division impartially, converge – not just in principle but also in practice –on the distributive principle recommended by a social contract. This is a principle of distribution that, according to the liberal egalitarian view, would rectify the effects of an initial arbitrary allocation of rights over production means, and would disentangle such effects from those attributable only to personal contribution and desert (see the next section for a theoretical discussion of this point). One may say that real economic environments or the governance structures of most firms seldom *de facto* satisfy (and certainly do not satisfy by any logical necessity) ideal conditions of justice. Hence our inquiry seems relevant to giving substance to contractarian business ethics by seeing whether real-life agents uphold contractarian principles of justice in such non-ideal productive contexts, so that they would be ready to rectify distribution outcomes after entering the experiment of the 'veil of ignorance'.

Before describing the experimental design and our empirical hypotheses, in Sections 2 and 3 we summarize the theoretical background of the research.

2. Theoretical background (I): the contractarian argument leading to liberal egalitarianism.

In our experiment, thesubjects were asked to agree on a distributive rule before they knew their own endowment, their own ability to contribute to production, and their actual effort. They were offered a number of possible rules, representing the most common intuitions about distributive justice. One of these rules –the liberal egalitarian rule– played a crucial role in our working hypotheses and consequently, in interpreting the experiment results.

Even if the rule itself does not exactly coincide with Rawls's second principle of justice, it will sound familiar to all acquainted with Rawls's theory, and it is virtually coincident with what Dworkin (2013) has suggested as a principle of social justice. In what follows we provide a detailed

specification of the theoretical reasoning behind our version of the liberal egalitarian rule and its connection with the contractarian approach. We propose to do so by means of a five-step argument:

1. First, let us accept two widely-held ideas about distributive justice: the egalitarian one and the desert/merit one. They are related to various philosophical traditions that need not be spelledout here. Egalitarianism claims that prospective participants in a joint venture deserve equal consideration and respect; while the conception of 'just desert' – or distribution proportional to contribution – focuses on the moral idea of responsible agency.
2. These two broad intuitions fit quite naturally into two contexts often found in economic approaches to theories of justices. These contexts can be dubbed 'manna from heaven' and 'non-manna from heaven' respectively. The first is the kind of situation in which a certain resource or good (sometime simplistically money, but also rights over resources or means of production) is given to the subjects to be distributed. In this case, insofar as there is no relevant difference amongst individuals, equality seems the obvious principle of distribution.[1] However, when subjects must earn the resources or goods in question, for example by working, keeping what one has earned by freely making a personal effort is seen as just.[2]
3. The social contract theory enables us to organize these two contexts hierarchically according to the two-tier model of constitutional and post-constitutional contracts (Buchanan 1975, Brock 1979, Sacconi 2000, 2006, 2011b). The agreement on the constitution is prior to any contribution. This agreement establishes the basic 'rules of the game'. This is the idea that a constitution allocates a set of rights to action. Essential among the latter are endowments of means of production and capabilities to be used in the next stage of social cooperation. Thereafter, specific cooperative ventures are established. These ventures are based on a post-constitutional agreement: subjects endowed with an amount of rights and capabilities must agree on how to distribute the surplus deriving from cooperation, in order precisely to set up mutually acceptable cooperative ventures. In the first stage, the egalitarian intuition seems to be the only focal point for agreement. But in the second stage, distribution proportional to contribution may be a rational agreement.

[1]A distribution proportional to relative needs – because needs are independent from contributions – may also fit the "manna from heaven" context. However, the constitutional choice is taken under a veil of ignorance. This prevents to have knowledge of the details of different individual life plans, and hence who needs what differential means in order to pursue her/his life-plan. Because there is no basis on which to distinguish among individual needs, at this level of analysis needs can be set as equal. Everybody equally needs some basic set of endowments in order to be able to pursue whatever life-plan.

[2]Illuminating on these distinctions are the arguments set out in Dworkin (1981a,b), Roemer (1986, 1996), Cohen (1989).

4. This point can be made analytically precise by a game-theoretical formulation of the two-tier constitutional/post-constitutional agreement model that defines it as a two-stage bargaining game (Brock 1979, Sacconi 2000, 2006, 2011b). At the initial stage bargaining focuses on the selection of a sub-set of strategies for each player from the overall set of all possible strategies - i.e. players select a constitution consisting of a limitation on any players' possibility of action, essentially curtailing the possibility of opportunistic behavior, and facilitating beneficial joint strategies. Then each player, being endowed with a particular sub-set of strategies, accedes to the second stage game – the sub-game resulting from taking as feasible only the strategy combinations allowed by the constitution. This is a cooperative game in which benefit (surplus) derives from the participation of players in several possible coalitions, and where no significant benefit can be secured by players acting on their own. One common solution concept for this kind of games is the 'Shapley value', whereby the cooperative benefit is allotted to players proportionally to the differential increase that each player contributes to the value of any possible coalition in the game. But of course contribution depends substantially on the constraints set by the initial selection of the constitution. In fact, assuming that agents are equally rational, their contribution is strictly conditioned by the initial restriction on strategy-sets decided through the first-stage bargaining solution. So, what about the initial-stage bargaining solution? From an abstract point of view, all kinds of constitutions could be selected at this stage. But then some possible coalition formation would reflect an antecedent asymmetry of power –as it actually happens in the real world. Coalitions of this sort could allot a minimum share of the total surplus to some parties simply because they had no bargaining power in the first place. But if we situate hypothetically behind a veil of ignorance, no rational bargainers would accept this. Rational players would restrict the set of feasible strategies to those allowed by a constitution that prevents opportunism and exploitation. In other words, they would agree to form mutually beneficial coalitions only after having secured basic rights/endowments for all. They would secure equal endowments first; and only afterwards they would bargain their way into coalitions that are mutually beneficial. To put it formally, in order to agree on a constitution at the initial stage, a rational agent must reason backwards: s/he must calculate, first, her/his expected payoff under any cooperative post-constitutional sub-game; then s/he must go back to the constitutional stage and choose among the cooperative sub-game solutions the one that most satisfies the conditions for an optimal agreement at the initial stage. At this stage, players play a symmetrical bargaining game. No differential contribution may be claimed at this point since neither different allocations of endowments nor their differential use may have still occurred - as

far as any possible allocation of endowments (strategy sub-sets) to players is still possible (i.e. for any endowment allocation assigning an advantage to a player also a symmetrical allocation, under which the players positions are symmetrically exchanged, is possible). The bargaining space is therefore symmetrical, and hence players must adopt the 'egalitarian solution': the Nash Bargaining Solution under a symmetrical payoff space(Nash 1950, Binmore 2005, Sacconi 2006, 2011b).

5. Of course, in the real business world endowments are not distributed equally. This means that the constitution may have failed in distributing basic endowments equally, so that this is not an ideal situation of justice. Under this arbitrary condition, co-operative productive ventures yield arbitrarily unequal distributions of benefits – even if they have followed a fair proportional principle. Justice – according the view mentioned above –would require, first, making endowments equal, and only thereafter applying a proportional principle to the final distribution. Our question is therefore whether individuals situated behind a veil of ignorance would perceive that the arbitrary distribution of endowments is no ground for making a claim to a commonoutcome. If they do, they would choose a distribution rule whose effect is to redress that initial arbitrarinessas much as possible[3].

[3]The liberal egalitarian principle on which we focus in this work is akin to James Konow's accountability principle (Konow, 1996, 2000, 2003,2005) according to which "fair allocations are proportional to the contributions agents control (called 'discretionary' variables) but do not adjust for factors they cannot influence (called 'exogenous' variables)"(Konow, 2005: 378). If we apply this principle in a productive context, if a worker is twice as productive as her colleague, an allocation assigning twice as her colleague to the more productive worker would be deemed fair if the difference in productivity is due to different effort. But the same distribution would be judged as unfair if the difference is due to variables beyond the direct control of the two workers (for example, different working conditions). (For a more recent empirical study focused on the application of this principle see Cappelen et al., 2014). What our view of the liberal egalitarian principle adds to Konow's idea, is an explicit justification – missing in Konow's definition- of why equality is fair in absence of any discretionary factor that justifies unequal distributions. According to liberal egalitarians like Rawls (Rawls 1971) and Dworkin (Dworkin 1981b), individuals are morally equal persons deserving equal consideration and respect, and the introduction of any difference among them must be morally justified in terms of outcomes that they can be responsible for because of their agency and independently of the results of social and natural lottery.

Another benchmark to compare Konow's accountability principle with our results supporting the liberal egalitarian principle is how different the understanding of impartiality is in the two cases. Konow distinguishes between stakeholders and spectators, and he shows that stakeholders are affected by self-serving biases in allocations while the spectators are not. We also find this result in what we call the 'no-veil' treatment, where subjects make their decisions without being put behind a veil of ignorance. In fact, we don't see any convergence on the liberal egalitarian rule in such a context, whereas we find it substantially when subjects are put in a context of impartial deliberation (behind a veil of ignorance concerning their identity in the game). The two approaches have in common the role of impartiality in adopting fair distributive rules, but in addition we find that the stakeholders themselves, once they have the opportunity to deliberate under the veil don't show any self-serving bias (or at least, a weaker one).

This relates to a main difference in modelling impartiality. We enquire the explicit agreement among 'stakeholders' under the veil of ignorance, whereas Konow works with third-person impartial spectators – i.e. we work in the contractarian tradition whereas he is in the impartial spectator perspective. This allows us to say what the stakeholders themselves would do when positioned behind the veil, and moreover whether they comply ex post with their impartial agreement, which cannot be done by taking the third party spectator perspective.

The endowment that our experiment distributed unequally was time. Subjects had to perform a task in pairs, but one member was randomly assigned six minutes, and the other one ten minutes. The 'liberal egalitarian' rule that we tested with our experimental subjects established the following distribution: assign each member of the pair what she/he has produced in the first six minutes; then distribute what the member who had ten minutes produced in the last four minutes equally among both. As made clear above, according to the theory of the social contract the rule chosen under the veil of ignorance should be the liberal egalitarian one: inequalities should be caused only by the subjects' differential use of an *equal* endowment of time, whereas arbitrary inequalities of endowments should be neutralized – because, in fact, according to the rule its output is re-distributed equally. The question is whether ordinary experimental subjects, faced with random initial endowments allocation and the output of their actual work, would agree on such a rule or choose another, maybe simpler, one.

We added four more rules from which to choose. Two of them were obvious candidates for comparison with liberal egalitarianism: *pure egalitarianism* (the total product is distributed equally), and merit purely based on *contribution* (each subject gets what she/he has individually produced).The other two rules reflect views that are typical in economic contexts: *self-interest*(each subject claims the entire product of the pair), and distribution strictly proportional to *productivity* (so that the person with less endowment may actually get *more* if s/he has been more productiveper minute).

3. Theoretical background (II): ex-post choice and the "the sense of justice".

By including in our experiment an actual distributive decision after the veil has been removed and production has taken place, we sought to test the *motivational* force of the agreement itself.

Rawls's theory makes the crucial assumption that individuals possess a 'sense of justice'. They can be motivated to act in accordance with fair principles provided that there is common knowledge that they have been accepted through an impartial process, and others do the same. This assumption empirically implies that if a process is implemented to assure subjects that the distribution rule that they have adopted is fair, they should be motivated to comply with it, simply because of its fairness and the expected reciprocity of others in respecting the principles. This conjecture would have a great impact on applied ethics. It may imply that making participants (stakeholders) aware of the

justice of the rules would enhance compliance. And in particular, involving them in the process of agreement on distributive criteria would enhance compliance, with little or no need for control. And this would be a result useful for institutional design.

No matter which moral theory one favors, the 'veil of ignorance' can be regarded as capturing our considered view about fairness (Binmore 2005). It is not obvious that an agreement behind the veil would by itself make the subjects more prone to *be* fair when the veil is lifted. But this would be exactly the case if Rawls is right about individuals possessing a 'sense of justice'. If this sense does indeed work in ordinary people, then a distribution principle that is rationally agreed behind a veil of ignorance would be seen under a special light also ex-post, so that people would be inclined to comply with it, even in the face of strong incentives to defect. What really counts in eliciting a preference for compliance are (i)participation in the impartial agreement, and (ii) that the agreement elicits a mental model of an agent that, having agreed, also intends to carry out the agreed action. The agreement can thus be seen as a kind of moral learning device.

Focusing now on the liberal egalitarian rule, if the theory is sound we should then observe that those subjects who reach an agreement on this rule– according to the theory outlined in Section 3,this should be the majority of subjects who solve theirex ante bargaining game rationally – will easily find a motivation to comply. Because their agreement has been rational, they may align their actual free choice with the agreement that they willingly made. The hypothesis here is that those subjects who have ex-ante agreed on the liberal egalitarian rule develop the attitude to conform with that rule also when the veil has been removed, just because they want to act on the basis of their rational agreement; and they also expect that symmetrically situated others will do the same. Thus, the real choice, after the veil has been removed, does not test the impartial reasons to agree on a distributive principle chosen behind the veil of ignorance (as most experimental literature has tested), but the *motivational* force of the agreement in a compliance context.

According to standard economic rationality, individual decisions in the kind of game that we used (a 'dictator game') should be selfish: no matter what has been agreed, a dictatorial position (after the veil has been lifted) should bend decisions toward individual interests. This is all the more so if one thinks of the individual who gets lucky in the initial distribution of endowments. S/he has the chance to produce a larger part of the total output, so that s/he can easily rationalize a claim to all that s/he has produced, instead of sharing equally the production resulting from his/her excess endowment, as the liberal egalitarian principle requires. Since, according to traditional economic

theory of motivation, it is not rational to comply with the agreement, one should not expect other agents to willingly comply on their own. But the social contract perspective allows a reversal of this expectation by considering that an ex-ante agreement –if recognized as rational– may affect agents' preferences and beliefs.[4]

To investigate whether and how the agreement under the veil of ignorance may actually elicit such preferences, we used two different agreement procedures. The first one we call 'bargaining'. It consisted in a series of offers and counter-offers that a subject simply saw on his/her computer screen when they had been 'submitted' through his/her or the counter-party's vote - so that they could insist on their vote, make a new proposal, or accept the counter-party's proposal by voting for it in the following bargaining round. The second procedure we call 'chat'. In this case, the choice of rule was preceded by an anonymous on-line chat. The idea was to test whether the possibility of a discussion using natural language –even if through a chat interfacethat kept identities veiled– should enhance the possibility of agreement. This would be suggested by a deliberative approach to economic justice (Ulrich 2008, Baur & Palazzo 2011). Natural language may help subjects to discuss the merits of each rule, and argue for the rationality and fairness of each, so that their agreement may be consciously based on the moral merits of a certain rule. If we follow Habermas's case for discourse ethics, an agreement that is reached because reasons are offered –as opposed to an agreement reached simply to end a bargaining procedure– should be normatively superior and therefore more motivationally efficacious for individuals with a disposition toward justice. The chat procedure might enhance commitment to the agreed rule; but its real force would lie in the fact that the participants can offer each other *reasons* for choosing a rule.

4. Experimental design and procedures.

[4] A reformulation of economic rationality consistent with the sense of justice has been suggested – also with the support of early experimental tests – by the theory of *conformist preferences* (see Grimalda and Sacconi 2005, Sacconi and Faillo 2010, Sacconi 2011c, Faillo, Ottone and Sacconi 2011, 2015) in so called 'psychological games' (see Geanakoplos et al. 1989, Rabin 1993). After an ex ante agreement, agents develop ideal preferences favorable to implementation of the agreed principle which are intrinsically dependent on the formation of first-order and second-order beliefs predicting reciprocal compliance by the interacting agents with the principle and alignment of reciprocal beliefs with behaviors conforming (in turn) with the principle.

.

The experiment consisted of three treatments: Noveil, Bargaining and Chat. In all the treatments subjects were matched in pairs, and asked to perform a task. In each pair one subject was randomly given six minutes and the other ten minutes to perform the task. Subjects generated an amount of money that depended on the outcome of the task. At the end of the task each subject was asked how to divide the total amount of money produced by the pair through the task. They could answer either by reporting the percentage to be assigned to each member of the pair or by choosing a division corresponding to one of five rules proposed by the experimenter. Henceforth, we will call this latter decision the "ex-post choice" to distinguish it from the former or "ex-ante" choice, which is material only to the Bargaining and Chat treatments. In these treatments, in fact, before performing the task and before knowing who would have six minutes and who ten minutes to perform the task, the members of each pair had to agree on one of the five division rules mentioned below. We will refer to this phase also as the "ex-ante agreement". For the sake of comparability with the other treatments, we will refer to the division choice as the "ex-post choice" also in the NoVeil treatment, even if in this treatment there was no "ex-ante agreement" phase. Once the subjects had decided, one of the two members was randomly selected and his/her choice was implemented.

In all the treatments, the subjects knew about all the phases of the experiment from the beginning, before they made their very first choice.

A detailed presentation of the treatments is provided in what follows.

4.1 Noveil treatment

In the Noveil treatment subjects were randomly matched in pairs. The treatment consisted of three phases in the following sequence: a practice phase, a task phase, and a division phase. We describe the three phases following the order used in the instructions, in which subjects first learned about the task and the division phases, and then about the practice phase.

The task

The task consisted in encoding words. In each pair, one of the subjects was given a total time of ten minutes to perform the task, while the other was given only six minutes. The assignment was random. Information about the time limits was given just before the task. A sequence of words appeared on the subjects' screens, and using a conversion table they had to convert the words into sequences of numbers. A new word appeared only after a code (either correct or mistaken) was written for the current word. The remaining time was shown through a countdown on the computer

screen. The total production (i.e. the number of tokens- one token=0.15 euros) generated in the task corresponded to the number of words correctly encoded by the two subjects.
At the end of the task, the following data were provided to the subjects: the total production (total number of words correctly encoded) of the pair, individual productions of the two members of the pair, productivity (number of words/minute) of each member, production and productivity of the subject with the ten minutes both in the first six minutes and in the second four minutes.

The division phase and the rules.
In the division phase (or "ex-post choice") each member of the pair was asked to choose separately how to divide the total income generated by the pair in the task phase. S/he could do this either by choosing a percentage from 0 to 100%[5] of the total income to ask for him/herself or by choosing a division corresponding to the application of one of the five division rules. Subjects saw on their screens the final payoffs corresponding to the application of each of the five rules. No request to agree was involved at this stage.

The rules were the following:
1) *Rule 1 – Equal split*: each subject obtains exactly half of the total product generated through the activity performed by the two subjects.
Example: subject A encodes 60 words in 10minutes; subject B produces 40 words in 6 minutes. Both subject A and subject B obtain (60+40)/2=50 tokens.
2) Rule 2 – One gets all: one subject obtains all the total product generated through the activity performed by the two subjects. A random draw selects the subject who gets 100% of the total product. Both subjects have a 50% probability of being selected.
Example: subject A encodes 60 words in 10 minutes; subject B encodes 40words in 6 minutes. The subject who is randomly selected (50% of probability to be selected) obtains 60+40=100 tokens, the other subject obtains 0.
3) Rule 3 – One gets what one has produced: each subject obtains exactly what s/he has produced through his/her activity.
Example: subject A encodes 60 words in 10 minutes; subject B encodes 40 words in 6 minutes. Subject A obtains 60 tokens, subject B obtains 40 tokens.

[5] The option of free percentages characterized also the ex-post division choice in the two treatments with the agreement (see below). The possibility to choose a free percentage put compliance with the rule agreed behind the veil of ignorance in the worst condition to be realized. In fact, free percentages ensured that no subjects complied with the agreement because of the lack of alternatives.

4) *Rule 4 – Time independent division*: each subject obtains what s/he has produced through her activity during the first 6 minutes; what is produced by the subject who has 10 minutes of time in the last 4 minutes is divided at 50% between the two subjects.
Example: subject A encodes 40 words in the first six minutes and 20 words in the second four minutes; subject B encodes 40 words in 6 minutes. Subject A obtains 40+(20/2)=50 tokens and subjects B obtains 40+(20/2)=50 tokens.
5) *Rule 5 – Divide according to productivity*: if the ratio between the productivity (words per minute) of A and B is *x*, then A's payoff should be *x* times the payoff of B, subject to the constraint that the sum of the two payoffs is equal to the total income produced by the pair.
Example: subject A produces 60 words in 10 minutes, subject B produces 40 in 6 minutes. The ratio between A's and B's productivity is 6/6.66= 0.90. The payoff of A should be 0.90 times the payoff of B, and the sum of the two payoffs should be 60+40=100 tokens. A's payoff is 47.4 tokens and B's payoff is 52.6 tokens.

When choosing the division, subjects could read the text of the rules, and they were also shown the payoff which they would obtain if that rule was applied, given the outcome of the task.
Once both the members of the pair had made their individual decisions, by opting for a division consistent with one out of the five rules or for a percentage, one of the members of the pair was randomly selected and his/her decision was implemented.

The practice phase.
Before starting the task, the subjects could practice with the rules, individually, by using a simulation platform that replicated the actual division choice screen of the third phase. They could read the five rules on their screen and choose one of them. They could also insert the number of words encoded by the person with six minutes and by the one with ten minutes both in the first six minutes and in the remaining four minutes, and they could decide the person (the one with six minutes or the one with ten minutes) whose final choice would be selected. They could play with the platform for five minutes, changing the parameters and checking the resulting outcomes. Figure 1a reports the exact sequence of the phases.

4.2 Bargaining treatment

In the Bargaining treatment, the practice phase, the task and the division phase (or "ex-post choice") were the same as in the NoVeil treatment, but the task and the division phases were preceded by a stage in which the members of the pair, before knowing the allocation of the time for the task, could reach an ex-ante agreement on one of the same five rules through a bargaining procedure – the agreement did not concern the choice of a percentage from 0 to 100% of the total production (see Figure 1b for the exact sequence of the phases). The procedure consisted of a maximum of thirteen rounds. In the first six rounds, subjects simultaneously chose one of the rules, proposing it for the final division of the total product generated through the task. They could choose the rule using a choice screen similar to the final division choice screen. At the end of each round, they were informed about the rule chosen by their partner, and if they had chosen the same rule, this was an agreement. Pairs unable to reach an agreement on one of the rules (by choosing the same rule) in the first six rounds accessed a second bargaining stage of four sequential choices. Each sequential choice consisted of an offer and, if the receiver refused it, of a counter-offer. At the beginning of each of the two sequential choices, one of the two members of the pair was randomly selected to make the first offer. The other member, once s/he had received the offer, decided whether to accept or refuse it. If s/he rejected the offer, then s/he had to make a counter-offer that might be accepted or refused by the counterpart. Pairs that failed toreachingan agreement also in this second stage moved to a final sequence of three further simultaneous choices.[6]The subjects knew that the rule was not going to be enforced, but they also knew that they could proceed to the experiment's next phase (the task) only if they reached an agreement. If they failed, they would be excluded from the experiment and they would be asked to fill in a questionnaire not related with the experiment. In this case, their earning would be equal to the initial 3 euros.

The agreement phase was preceded by the practice phase, which allowed subjects to become familiar with the choice interface and to the consequences, in terms of final payoffs, of their decisions.

[6]The first sequence of simultaneous division proposals was introduced in order to capture the simultaneous nature of the bargaining. The second sequential bargaining phase was introduced to help break possible non-coordination cycles in the simultaneous choices. Finally, the last simultaneous choices phase wasintended to prevent agreements reached in the sequential bargaining phase from suffering the typical hold-up problem that characterizes finite sequential bargaining, in which the second to last mover has an advantage over the last mover. The former, in fact, can put the latter in the condition to accept or refuse his/her preferred rule. By introducing a third simultaneous bargaining phase, we gave subjects the possibility to escape from the hold-up problem after having been involved in the first two phases. Note that only two pairs failed to reach an agreement within the first sequence of simultaneous choices. They reached the agreement in the first attempt pertaining to the sequential bargaining phase.

In the ex-post choice, subjects were reminded of the rule chosen by their pair (the rule appeared also with a different background color) in the ex-ante agreement, and they could choose either a percentage of the total product to ask for themselves, a division of the total product corresponding to application of the agreed rule, or a division corresponding to the application of a different rule. As in the NoVeil treatment, the final payoffs corresponding to the application of each of the five rules were reported on the subjects' computer screens.

4.3 Chat treatment

The Chat treatment was very similar to the Bargaining treatment. Subjects had to reach an agreement on one of the five division rules in order to access the task and the ex-post division phase. However in this treatment the ex-ante agreement procedure was based on a chat. Subjects were given five minutes for discussion. The chat was anonymous. Communication of personal information, PC number, threats, promises of side payments and the use of offensive language were prohibited. Once the two members of the pair had reached an agreement, they had to choose the same rule on a choice screen similar to the final division choice screen. The choice of the same rule could be done at any time within the limit of 5 minutes available to discuss through the chat function. Thus, selecting the same rule on the screen after having agreed to it by the chat was a way to make clear that the agreement had been actually reached and that there was no misunderstanding about it.[7]All the pairs succeeded in choosing the same rule (it took on average 3.75 minutes). As in the Bargaining treatment, they knew that the agreement was not going to be enforced in the later stage of the game, but if they had failed to reach the agreement they would have been excluded from the experiment and asked to fill in a general questionnaire not related with the experiment. See Figure 1c for the exact sequence of the phases.

As in the Bargaining treatment, in the ex-post choice subjects were reminded of the rule chosen by their pair and they could choose separately either a free percentage of the total product to ask for themselves, or a division corresponding to the application of the agreed rule, or a division corresponding to the application of a different rule.

4.4. ***Beliefs and questionnaire***

In all the treatments, at the end of the ex-post choice, before a subject knew if his/her choice was selected for payment, first- and second-order beliefs were elicited by asking what s/he believed the

[7]If a pair reached an agreement on a rule during the chat, but one of the members chose the wrong rule on the screen, a warning message about the "mistake" appeared and the subject could make another choice. Only one mistake was allowed.

other member of the pair had chosen (either one of the five rules or a percentage of total product) and what s/he believed the other member believed was her choice. Correct guesses were rewarded with one euro. Participants were also asked to fill in a questionnaire containing both socio-demographic questions and questions about trust, risk attitude and happiness.[8] In each treatment, in two sessions the questionnaire was administered at the beginning of the experiment, before the instructions about the phases of the experiment were read; and in two sessions it was administered at the very end of the experiment, just before the payment (note that our main empirical results are virtually unchanged when we consider this distinction).

4.5 Sessions and procedures

The experiment was programmed by using Z-tree (Fischbacher, 2007) and conducted at the EGEO laboratory of the University of Granada. Subjects were paid a show-up fee of three euros. We adopted a between-subject design. No individual participated in more than one session.

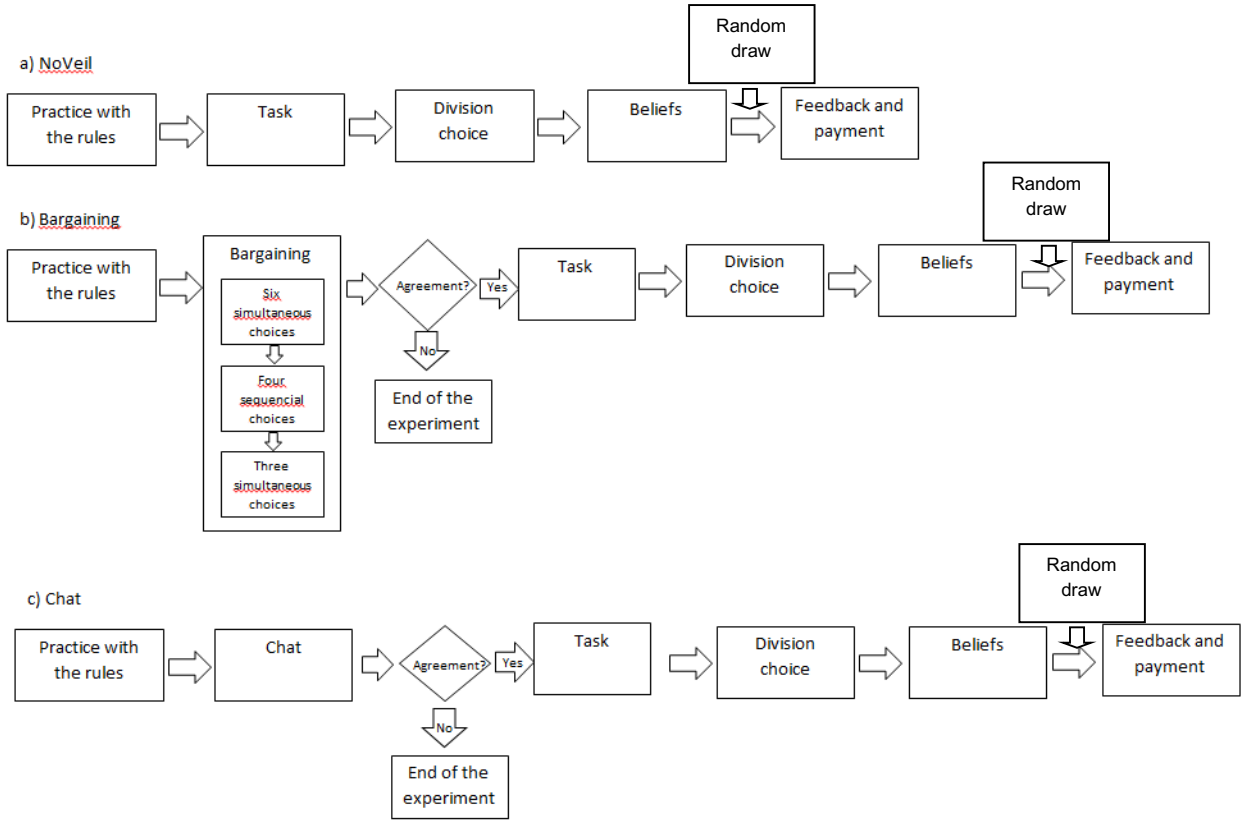
The average payment per participant was 9.80 € (including the show-up fee) and the sessions lasted approximately 1 hour and 15 minutes on average.

In all the treatments, at the beginning of each session, participants were welcomed and asked to draw lots. They were then randomly assigned to terminals. The instructions were handed to them in written form and were read aloud by the experimenter. The participants had to answer several control questions, and we did not proceed with the actual experiment until all participants had answered all questions correctly.

A total of 236 students participated in the experiment between May 2014 and March 2015. We ran four sessions of 20 subjects each for the Noveil and the Bargaining treatments, and four sessions, three with 20 participants and one with 16 participants, for the Chat treatment.

[8] The questionnaire is available from the authors upon request.

Figure 1. A synthesis of the structures of the three treatments



5. Empirical hypotheses and their justification.

Six main hypotheses, strictly connected with the theoretical framework presented in Sections 2 and 3, underlie our analysis. In this section they are summarized and theoretical support for each of them is provided.

In what follows, as we have done in Section 4, we will refer to the agreement characterizing both the Bargaining and the Chat treatments as the "ex-ante agreement". We will name "ex-post choice" the final choice made by subjects in the latter division phase of the Noveil, Bargaining and Chat treatments.

> *H1. In the ex-ante agreement (in both the Chat and the Bargaining treatments), subjects opt significantly more for the division of the total product consistent with the liberal egalitarian rule than for the other options of division.*

This is a hypothesis on the choice *across* principles (or rules) only *within* treatments characterized by an ex-ante agreement. Its justification is that we conjecture that, even if the cognitive task for the liberal egalitarian rule is relatively more cumbersome than those for the alternatives, the immediate reasoning accomplished by subjects under the 'veil of ignorance' favors the liberal egalitarian rule. In other words, the impartial perspective taken by subjects under the veil induces them to prefer rule 4.

> *H2. Subjects opt significantly more for the division of the total product consistent with the liberal egalitarian rule in the ex-ante agreement (in both the Chat and the Bargaining treatment) than in the ex-post choice in the Noveil treatment*.

This is a hypothesis on the choice of the liberal egalitarian rule *across* treatments and compares somewhat spurious choices (i.e. the number of ex-ante agreements over liberal egalitarianism, when such agreements are allowed) with the number of liberal egalitarian separate division choices when as in Noveil the only decision available to subjects is the ex-post choice. Its justification is that we conjecture that the preference for liberal egalitarianism is largely conveyed by the ex-ante agreement under the veil of ignorance. Hence a treatment in which no choice under the veil is allowed should exhibit preferences in favor of rule 4 expressed only by those subjects who - so to speak - by themselves carry out an impartial reasoning but are not aided in doing so by the actual simulation of 'veil of ignorance' through an ex-ante agreement stage. In other words, the ex-ante decision situation of "agreement under the veil of ignorance" elicits impartial reasoning (or impartial reasoning comes to the subjects' mind because of the ex-ante agreement situation). Consequently, we expect that liberal egalitarianism is chosen ex-post significantly less in the Noveil treatment than by decisions (agreements) under the veil in agreements treatments.

> *H3. In the ex-post choice in the Chat and Bargaining treatment we should observe more subjects choosing the division of the total product associated with the liberal egalitarian rule than in the Noveil treatment. This is the effect of the ex-ante agreement, reached through Bargaining or Chat, on the subjects' actual decision to select the division of the total product associated with the liberal egalitarian rule when they face the ex-post choice.*

Hypothesis H3 concerns the performance of the *liberal egalitarian* rule in the ex-post choice. It states that when ex-ante agreements are allowed, the number of ex-post choices for liberal egalitarianism should be higher, and what explains this increase is precisely that the ex-ante agreement affects the ex-post choices of the rule. This supports the theoretical hypothesis that the

ex-ante agreement may be understood as a player's 'reason to act' in favor of playing the liberal egalitarian rule ex-post, which is translated into an effective causal force for playing it. Observing that subjects chose rule 4 in the ex-post decision, and that this decision is a consequence of the ex-ante agreement, we may say that they are in fact carrying out a decision to *comply* intentionally and voluntarily with the agreed rule. Positive evidence with respect to these hypotheses would provide empirical support for the Rawlsian idea of an effective 'sense of justice' and the model of *conformity preference* (see Section 3 and note 3), even if complete coherence should consider preferences and beliefs simultaneously (see H6 hereafter).

> *H4. The Chat treatment is more effective in inducing compliance than the Bargaining treatment. Chat is also more effective in affecting beliefs in the sense that first-order and second-order beliefs concerning reciprocal compliance with the agreement are mostly aligned in the Chat treatment and this explains why subjects comply with the agreement more in Chat than in Bargaining*

This hypothesis rests on the conjecture that if the veil of ignorance activates an impartial viewpoint that affects beliefs, ex-post preferences and choice, then this effect should be stronger in the case of the Chat treatment. The reason is that, in the Chat treatment, an ex-ante agreement is reached not only under the conditions of a veil of ignorance but also through deliberation, and with the possibility of putting forward impartial arguments during the agreement stage to convince the counterparties. We therefore conjecture that the impartiality of the agreementshouldbe clearer in the Chat than the Bargaining treatment. One tenet of deliberative democracy theory (Habermas 1996) is that impartial deliberation may affect the initial preferences whereby subjects enter the deliberative process,thus facilitating them in reaching impartial agreements. If "a common ground on questions of right and wrong, or fair and unfair can only be found through joint communicative processes between different actors" (Scherer and Palazzo 2007: 4), we should observe a closer alignment of first-order and second-order reciprocal beliefs –beliefs consistent with mutual compliance with the agreement– in the case of natural-language discussion[9]. In other words, we will observe higher ex-

[9]Full transcriptions of the chats (in Spanish) are available upon request. The transcription of a typical exchange may illustrate and help the reader to see the point (for anonymity, in all pairs subjects get the nickname "A" or "B"): "B: Hello; A: Hello. I suggest rule 4. What do you think?; B: Rule 4? I thought of going discarding [different rules]; A: Discarding? B: but yes, 4 is appropriate. I meant, saying which rules we do not want. And which one we do want; A: So we agree on 4? - ? – If you want? B: What about rule 5? A: I prefer rule 4, because the final difference is not so big, and it [payoff] depends more on what each one does; B: well, okay. Rule 4 then? A: okay, perfect. We agree on 4; B: Do we click already? A: Yes; B: okay.

post compliance in Chat than in Bargaining, and the difference between the two should be attributable to the observation that Chat mostly induces aligned beliefs.

> *H5.Compliance with the ex-ante agreement in which subjects choose the liberal egalitarian rule characterizes also subjects endowed with ten minutes of time, i.e. those who give up part of the output from their production when choosing that rule.*

This hypothesis is based on the theoretical conjecture that the motivational force behind the decision to carry out rule 4 in the ex-post choice –if the subjects have ex-ante chosen the liberal egalitarian rule – is truly a preference for compliance with the liberal egalitarian rule based on reasons of principle: the sense of justice expressed as adherence to a commonly agreed conception of fair distribution. The 'crucial experiment' with which to test the force of this motivation would be one able to disentangle it from the concurrence of any other motivations. The obvious candidate for such concurrence is the self-interest of those agents who could gain personally from implementation of this rule. Hence we must ascertain that ex-post adhesion to rule 4 is not solely by those who gain personally from it because they having been randomly assigned only six minutes to perform the task. That also the 'lucky' subjects (the ones endowed with ten minutes to perform the task) decide to carry out liberal egalitarianism in the ex post decision, after they have learned and exploited the advantage of the additional 4 minutes, would clearly show a counter-interested behavior (note, moreover, that by sharing the output of the additional four minutes they also set aside principles of remuneration according to work or productivity).[10] All these explanations are compatible with the idea of a 'sense of justice' and the conformity preference theory (see again note 3).

> *H6 (a) Most of those who comply with the agreement to which they have subscribed also hold beliefs aligned with compliance. (b) Beliefs about the counterpart's compliance are engendered by the ex-ante agreement.*

H6 concerns first-order and second-order beliefs about conformity, and it is split into two independent but complementary sub-hypotheses. The justification for H6 (a) is that taking it together with H3 we have the empirical hypothesis directly deriving from both the idea of a sense of justice and the equivalent theory of preferences for conformity with principles. The latter are based on two premises. First, an agreement under a veil of ignorance must be reached on a distributive

[10]Only in two pairs involved in the Chat and Bargaining treatments did the subject endowed with 4 minutes of time produce more than his/her counterpart. See note 12 for a robustness check when these pairs are excluded from the empirical analysis.

principle. Second, agents compliant with the agreement hold beliefs aligned with mutual conformity – i.e. each agent believes that the other participant will do his/her best in terms of the principle given what s/he believe about the agent's own choice. When these two hypothesis are satisfied, the theory concludes that a psychological component positively enters the agent's preference system (i.e. a "sense of justice" motivates compliance).

In order to predict evidence consistent with the theory, however, H6(a) and H3 are necessary but not sufficient. Besides the evidence on those who comply with the agreement, it is necessary that a sufficiently high number of subjects who have ex-ante subscribed to an agreement hold beliefs of mutual compliance. In order to make sense of this conjecture, we hypothesize with H6(b) that those who make an agreement will also hold mutually aligned beliefs consistent with reciprocal conformity, and this will happen because the agreement induces such beliefs. Thus, we expect that if a subject has agreed ex-ante on a rule, then s/he will hold the first-order belief that the counterparty in the agreement will comply with the rule, and the second-order belief that the counterparty also expects him/herself to comply with the same rule. Even if this hypothesis may seem psychologically natural, we admit that it is by no means trivial. In fact, in general respecting an agreement on a fairness rule like *liberal egalitarianism* or others considered here may contradict standard economic rationality (see appendix 1 for theoretical views that underpinthis hypothesis) .

6. Hypotheses testing

The presentation of the empirical results, both descriptive evidence and econometric estimates, is organized by hypotheses, each of the following sub-section being devoted to a single hypothesis, except for section 6.4, which investigates H4 and H5.

In what follows we consider the ex-post choice of opting for a percentage equal to 50% or to 100% as equivalent to the ex-post choice of Rule 1 or 2, respectively. In fact, in terms of the ex-post division of the total production, opting for Rule 1 (Rule 2) in the ex-post choice is equivalent to the choice of opting for the 50% (100%). Results are virtually unchanged if we do not merge subjects who opted for the previous percentages in their division choices with subjects who opted for Rule 1 or 2. When differences emerge in the econometric estimates, they are reported in the text or footnotes.

6.1. H1: The choice in the ex-ante agreement.

Table 1 shows the percentages of subjects who chose the various rules across treatments. At a first glance, it is evident that Rule 4 was chosen by the great majority of subjects in the ex-ante agreement – 57.5% of subjects in the Bargaining treatment and 57.89% in the Chat treatment. The other rules chosen with greater frequency in the ex-ante agreement were Rule 3 and Rule 5 in the Bargaining and in the Chat treatment, respectively. In both cases, these Rules were chosen by less than half of the subjects who opted for Rule 4. A test of proportions revealed that both in the Chat and in the Bargaining treatment Rule 4 was chosen by a proportion of subjects significantly greater than 40%.[11] Conversely, the same test revealed that the other rules were agreed by proportions of subjects equal to or lower than 20%.[12]. Therefore, the subjects' choices seem to support *H1*, according to which the division of the total product consistent with Rule 4 is expected to be the most agreed rule in the ex-ante agreement.

Table 1. The rule chosen across treatments (percentage values)

Rule	**Noveil**	**Bargaining**		**Chat**	
		Ex-ante agreement	**Ex-post choice**	**Ex-ante agreement**	**Ex-post choice**
Rule 1. Pure Equal Split	22.5	12.5	17.5	15.79	21.05
Rule 2. One gets all	13.75	2.5	3.75	0	7.89
Rule 3. One gets what one has produced	22.5	15	22.5	5.26	6.58
Rule 4. Time independent division	16.25	57.5	40	57.89	40.79
Rule 5. Divide according to productivity	10	12.5	10	21.05	17.11
Rule 6. Percentage	15	*Option not available*	6.25	*Option not available*	6.58

[11] One-sample proportion test, p= proportion of subjects who agreed on Rule 4; H_0 =0.4: Bargaining treatment, Ha: p != 0.4, Pr(|Z| > |z|) = 0.0014; Ha: p > 0.4, Pr(Z > z) = 0.0007; Ha: p < 0.4, Pr(Z > z) = 0.9993; Chat treatment, Ha: p != 0.4, Pr(|Z| > |z|) = 0.0015; Ha: p > 0.4, Pr(Z > z) = 0.0007; Ha: p < 0.4, Pr(Z > z) = 0.9993.

[12] One-sample proportion test, p= proportion of subjects who agreed on the different rules; H_0 =0.2: Rule 1: Bargaining treatment, Ha: p != 0.2, Pr(|Z| > |z|) =0.0935; Ha: p < 0.2, Pr(Z > z) =0.0468; Ha: p > 0.2, Pr(Z > z) = 0.9532; Chat treatment, Ha: p != 0.2, Pr(|Z| > |z|) =0.3588; Ha: p < 0.2, Pr(Z > z) =0.1794; Ha: p > 0.2, Pr(Z > z) = 0.8206; Rule 2: Bargaining treatment, Ha: p != 0.2, Pr(|Z| > |z|) = 0.0001; Ha: p < 0.2, Pr(Z > z) =0.0000; Ha: p > 0.2, Pr(Z > z) = 1.000; Chat treatment: no observations; Rule 3: Bargaining treatment, Ha: p != 0.2, Pr(|Z| > |z|) =0.2636; Ha: p < 0.2, Pr(Z > z) =0.1318; Ha: p > 0.2, Pr(Z > z) = 0.8682; Chat treatment, Ha: p != 0.2, Pr(|Z| > |z|) =0.0013; Ha: p < 0.2, Pr(Z > z) =0.0007; Ha: p > 0.2, Pr(Z > z) = 0.9993; Rule 5: Bargaining treatment, Ha: p != 0.2, Pr(|Z| > |z|) =0.0935; Ha: p < 0.2, Pr(Z > z) =0.0468; Ha: p > 0.2, Pr(Z > z) = 0.9532; Chat treatment, Ha: p != 0.2, Pr(|Z| > |z|) = 0.8185; Ha: p < 0.2, Pr(Z > z) =0.5907; Ha: p > 0.2, Pr(Z > z) =0.4093.

6.2. H2: A comparison between subjects' decisions in the ex-ante agreement and in the ex-post choice.

When we compare ex-post choices made by subjects in the Noveil treatment and choices concerning the ex-ante agreement in the Chat and Bargaining treatments, we find that the percentage of subjects who chose the division consistent with Rule 4 was lower in the Noveil treatment (ex-post choice) than in the ex-ante agreement concerning the Bargaining (Pearson $chi^2(1)$, Pr=0.000) and the Chat (Pearson chi2(1), Pr=0.000) treatment.

In order to check for the significance of differences between subjects' decision in the ex-post choice in the Noveil treatment and their decision in the ex-ante agreement in the Chat and Bargaining treatments, we ran a Logit regression (Table 2). The dependent variable was a binary indicator (*Rule_4_ex_ante-post*) which identified subjects who opted for the division consistent with Rule 4 in the Noveil treatment or in the ex-ante agreement in the Chat or Bargaining treatment. The independent variables of main interest were the dummies identifying the treatment in which subjects were involved, i.e. *Chat* (equal to one if subjects took part in the Chat treatment) and *Bargaining* (equal to one if subjects took part in the Bargaining treatment). Estimates included socio-demographic characteristics - i.e. age, sex, income, the propensity to take financial risk, religious orientation, the propensity to trust unknown others - controls connected with the experimental conditions - i.e. the number of words encrypted in the task, the number of words encrypted per minute - and the fact of having already taken part in Lab experiment[13] (see Appendix 2 for a description and descriptive statistics of all variables included in the estimates). These control variables have been excluded from the Tables for reasons of space. Full estimates results are included as *Appendix 3*.

The significance of the two *Chat* and *Bargaining* variables clearly reveals that subjects opted significantly more for the division of the total product associated with Rule 4 in the ex-ante agreement than in the ex-post choice in the Noveil treatment.[14] Moreover, the Wald test reported in the last line of Table 2 shows that no differences characterize the decision to opt for Rule 4 in the

[13] Two tailed Kruskal-Wallis tests run for gender (p=0.0067), age (p=0.0026) and income (p=0.0698) revealed that the three sub-samples of subjects involved in the different treatments were not perfectly balanced with respect to these variables. We replicated all the estimates reported in the following tables by controlling for these differences when significant. In particular, we included in our regressions interaction terms (when statistically significant) between the two treatment variables *Chat* and *Bargaining* and the three variables *Female*, *Age* and *Income*. We report in the footnotes the main differences emerging when interaction terms are considered.

[14] When we consider interaction terms (see footnote 12), we find that the significance of the Bargaining variable slightly decreases (p=0.047).

ex-ante agreement when the Chat and the Bargaining treatment are compared. We conclude that our data give support to *H2*.

Table 2.
Ex-post choices in the Noveil treatment vs. ex-ante agreement in the Bargaining and Chat treatment

	(1)
	Logit
Dependent variable: *Rule_4_ex_ante-post* - DV=1 if the division consistent with Rule 4 is chosen in the Noveil treatment or in the ex-ante agreement in the Chat and Bargaining treatment	
Chat	1.886***
	(0.397)
Bargaining	1.847***
	(0.406)
Constant	4.311
	(7.180)
Control variables	YES
Observations	236
Pseudo R^2	0.1386
Chat-Bargaining	0.038
	(0.343)

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

6.3. H3: The treatment effect on the ex-post division choice

When we focus on the ex-post choices, we observe that the division of the total production consistent with Rule 4 is most frequently chosen both in the Bargaining and in the Chat treatment, with percentage values equal to 40% and 40.79% respectively (Table 1). By contrast, in the Noveil treatment the majority of subjects opted for divisions associated with Rule 3 (22.5%). Moreover, the percentage of subjects who chose divisions consistent with Rule 4 was significantly lower in the Noveil treatment than in the Bargaining (Pearson $chi^2(1)$, Pr=0.001) and the Chat (Pearson $chi^2(1)$, Pr=0.001) treatment; conversely, no difference emerges between the Bargaining and Chat treatments (Pearson $chi^2(1)$, Pr=0.920).

In order to check for the significance of the treatment effect on the ex-post division choice consistent with Rule 4, we ran Logit regressions (Table 3). The dependent variable was a binary indicator (*Rule_4_ex-post*) whichtook value 1 if subjects opted for a division consistent with Rule 4 in their ex-post choice. The independent variables of main interest were the two dummies

identifying the treatment in which subjects were involved, i.e. *Chat* and *Bargaining*. Estimates included the same control variables as those included in Table 2 (see Appendix 2). Moreover, we considered a dummy variable (*Rule_agr_4*) equal to 1 if subjects involved in the Chat or in the Bargaining treatment opted for Rule 4 in the ex-ante agreement. This variable, included in Columns 2, captured the role of the agreement in affecting subjects' ex-post choice.

Table 3 shows the estimate results. The last line of the Table reports Wald-tests useful for comparing subjects' behavior in the Chat and the Bargaining treatments.

The division consistent with Rule 4 was more likely to be chosen in the ex-post choice both in the Chat and the Bargaining treatment than in the Noveil treatment,[15] while no differences emerge between Chat and Bargaining (column 1).[16] These results support the hypothesis that Rule 4 is chosen significantly more in the treatment characterized by the agreement than in the Noveil treatment.

Being involved in the Chat (Bargaining) treatment increases by 27.2% (26.4%) the probability of opting for the division associated with Rule 4 in the ex-post choice with respect to the Noveil treatment.

As expected from *H3*, the dummy identifying subjects who opted for Rule 4 in the ex-ante agreement (*Rule_agr_4*) significantly affected the decision to select a division consistent with that rule in the ex-post choice (column 2). Moreover, it entirely explains the propensity to opt in the ex-post choice for a division consistent with Rule 4 more frequently in the Chat and in the Bargaining treatment than in the Noveil one.[17]

[15]When we consider interaction terms (see footnote 12), we find that: a) the level of significance disappears for Men with respect to the Bargaing treatment; b) the level of significance decreases for Women, even though it remains within the 10% level (7.4%).

[16] When the interaction terms are considered, the difference between Chat and Bargaining emerges for Men, who opt for Rule 4 more in the Chat than in the bargaining treatment.

[17] When we consider interaction terms, this result is in general confirmed. Moreover, with specific respect to the Bargaining treatment, when controlling for *Rule_agr_4* it turns out that Men choose Rule 4 less than in the Noveil. This further highlights the importance of the agreement in favouring the decision to opt for Rule 4 in the ex-post choice.

Table 3. Determinants of choice of the rule

	(1)	(2)
	Logit	Logit
Dependent variable:	*Rule_4_ex-post* - DV=1 if a division consistent with Rule 4 is selected in the ex-post choice	
	Whole sample	
Chat	1.231***	-0.344
	(0.408)	(0.545)
Bargaining	1.201***	-0.326
	(0.417)	(0.553)
Rule_agr_4		2.485***
		(0.448)
Constant	13.12*	14.80*
	(7.184)	(8.684)
Control variables	YES	YES
Observations	236	236
Pseudo R^2	0.0904	0.2262
Chat-Bargaining	0.030	-0.018
	(0.352)	(0.405)

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

6.4. H4, H5: Compliance across treatments, rules and differences in time endowments

In this section we consider our fourth and fifth hypothesis (*H4* and *H5*) and investigate the level of compliance characterizing subjects who opted for Rule 4 in the ex-ante agreement.

We find that 59.62% of subjects complied with the rule chosen in the ex-ante agreement. The percentage increases in the case of the Chat treatment (71.05%) and decreases in the Bargaining treatment (48.75%). This difference is statistically significant (Pearson chi2(1), Pr=0.005). Moreover, compliance varies across rules and treatment (Table 4). The percentage of subjects who complied with the rule chosen in the ex-ante agreement is highest for subjects who agreed on Rule 1 in the Chat treatment (91.67%). The lowest percentage concerns Rule 5 in the Bargaining treatment (20%). In the Bargaining and the Chat treatment, the percentage of subjects who opted for Rule 4 in the ex-ante agreement and complied with the agreement is equal to 54.35% and 68.18% respectively, even though this difference is not statistically significant (Pearson chi2(1), Pr=0.178).

Table 4. Subjects who complied with the rule chosen in the ex-ante agreement – percentage values (absolute values in parenthesis).

	Rule 1	**Rule 2**	**Rule 3**	**Rule 4**	**Rule 5**
Bargaining	50	50	50	54.35	20
	(5)	(1)	(6)	(25)	(2)
Chat	91.67	*No observations*	50	68.18	68.75
	(11)		(2)	(30)	(11)

Table 5 shows the econometric analysis related to the determinants of compliance. Estimates consider only subjects involved in the Chat and Bargaining treatment. With respect to the estimates presented in Table 2, we added the following to the control variables: the rule chosen in the ex-ante agreement (*Rule_agr_1, Rule_agr_2, Rule_agr_3, Rule_agr_5*) - the residual category is represented by subjects who agreed on Rule 4; the payoff associated with the rule agreed in the ex-ante agreement (*Payment_agreement*). In column 2 we include two dummy variables aimed at capturing the role of first-order and second-order belief, *Belief_first=agr* (DV=1 if the subject believes that the other player in the pair is going to comply), *Belief_second=agr* (DV=1 if the subject believes that the other player believes that s/he is going to comply). In column 3 we include a dummy identifying subjects with 10 time minutes (*10_minutes*). In column 4 we focus on subjects who chose Rule 4in the ex-ante agreement. By including the interaction term between *10_minutes* and *Chat*, the estimate in column 5 checks for possible interaction effects concerning the endowment of time and the treatment. Our particular interest concerns subjects with ten minutes of time, since their incentives clashed with the adoption of this rule in the ex-post choice, meaning that they gave up part of the output stemming from their production when choosing the division related to Rule 4.[18]

Table 5 shows that:

a) the level of compliance is higher in the Chat than in the Bargaining treatment (column 1).[19] However, this effect is entirely explained by the role of beliefs. In fact, when the latter are included in the estimate (column 2), the difference in the level of compliance between subjects involved in the Chat and in the Bargaining treatment is no longer significant.[20] Since the chat procedure is more effective than the bargaining procedure also in inducing beliefs about the counterpart's compliance (see next section), this effect of beliefs is not surprising, and it is perfectly in line with *H4*, according to which the Chat treatment is more effective in inducing compliance (affecting subjects' beliefs) than the Bargaining treatment. Moreover, second-order beliefs positively affect the decision to comply, while no effect emerges for first-order beliefs once both first-order and second-order beliefs are included in the regressions

[18]The results presented below hold also when we exclude from the analysis the two pairs of subjects involved in Chat and Bargaining treatment in which the subject endowed with 4 minutes of time produced more than his/her counterpart (see footnote 9).

[19] When we consider possible differences of behavior between Men and Women (see footnote 12), we find that this result holds only for Men.

[20] In this case, when the analysis takes specific account of differences between Men and Women, we find that Men still comply more in the Chat than in the Bargaining treatment.

b) no differences in the level of compliance emerge between subjects endowed with 10 and 6 minutes of time (column 3). This result holds also with respect to the sub-sample of subjects who opted for Rule 4 in the ex-ante agreement (column 4), either *within* treatments (column 5, Wald test on the null that the sum of the coefficient of *10_minutes* and *10_minute_chat* is equal to 0: Pr=0.822) or *across* treatments (column 5, Wald test on the null that the sum of the coefficient of *Chat, 10_minutes, 10_minute_chat* is equal to 0: Pr=0.544).

To sum up, empirical evidence seems to support both *H4* and *H5*.

The estimates in Table 5also show that the level of compliance is higher for Rule 1 than for Rule 4 (Pr=0.015 –Table 5, column 2), Rule 3 and Rule 5 (column 2, probabilities concerning Wald tests on the null that the difference between the coefficients of *Rule_agr_1* and *Rule_agr_3, Rule_agr_5* are: Pr=0.019, and Pr=0.028, respectively);[21]

[21] When we do not merge subjects who opted for percentages equal to 50% and 100% with subjects who opted for Rule 1 or Rule 2, respectively, the differences between Rule 1 and Rule 3 and Rule 5 are no longer significant at 10% level, while the difference between Rule 1 and Rule 4 is still significant but at a lower level (10% instead of 5%).

Table 5. The determinants of compliance

	(1)	(2)	(3)	(4)	(5)
Method	Logit	Logit	Logit	Logit	Logit
Dependent variable	*Compliance*				
	All subjects involved in the Chat and Bargaining treatment			Sub-sample of subjects who opted for Rule 4 in the ex-ante agreement	
Chat	0.973**	0.553	0.553	-0.194	-1.490
	(0.407)	(0.470)	(0.470)	(0.655)	(0.986)
10_minutes			0.693	-3.174	-3.380
			(2.389)	(3.804)	(3.843)
10_minutes_chat					2.487*
					(1.283)
Rule_agr_1	1.720**	2.045**	2.113**		
	(0.731)	(0.837)	(0.872)		
Rule_agr_2	-2.393	-1.463	-1.588		
	(2.045)	(2.430)	(2.501)		
Rule_agr_3	0.280	-0.124	-0.0865		
	(0.670)	(0.730)	(0.741)		
Rule_agr_5	0.0659	0.0575	0.113		
	(0.638)	(0.718)	(0.748)		
Payment_agreement	0.0648*	0.0567	0.0595	-0.0503	0.00188
	(0.0354)	(0.0389)	(0.0401)	(0.151)	(0.160)
Belief_first=agr		0.597	0.572	1.116	0.978
		(0.492)	(0.499)	(0.691)	(0.721)
Belief_second=agr		2.310***	2.310***	2.926***	2.878***
		(0.530)	(0.532)	(0.831)	(0.854)
Constant	1.145	0.847	1.037	23.37	26.07
	(14.66)	(17.08)	(17.12)	(30.80)	(31.94)
Control variables	YES	YES	YES	YES	YES
Observations	156	156	156	90	90
Pseudo R^2	0.1596	0.2930	0.2934	0.3263	0.3596

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

6.5. H6: Ex-ante agreement, beliefs and compliance

In this section we analyze the correlation among the choice of a certain rule in the ex-ante agreement, the belief that the other player is going to comply with the agreement, and the decision to comply (*H6*).

70.51% of subjects believed that the other player in the pair was going to comply. The percentage increases to 78.95% when we consider the Chat treatment and decreases to 62.50% in the Bargaining treatment. This difference is statistically significant (Pearson chi2(1), Pr=0.024) . When we consider second-order beliefs, we find that 73.72% of subjects believed that the other player in

their pair believed that they were going to comply. The percentage increases to 86.84% when the Chat treatment is considered and decreases to 61.25% in the Bargaining treatment. This difference is statistically significant (Pearson chi2(1), Pr=0.000).

Among those who complied, 77.42% believed that the counterpart would comply as well. This percentage increases when we look at the Chat treatment (87.04%) and decreases for the Bargaining treatment (64.10%), generating a statistically significant difference between the two treatments (Pearson $chi^2(1)$, Pr=0.009). As regards second-order beliefs, 90.31% of subjects who complied believed that the counterpart believed that they were going to comply. Also in this case the percentage is significantly larger (Fisher's exact=0.032) in the Chat (96.30%) than in the Bargaining (82.05%).

Finally, 73.12% of subjects who complied hadaligned first-order and second-order beliefs: that is, they believed that the counterpart would comply and believed that the counterpart believed that they would do the same. Also in this case, the percentage is significantly larger (Pearson chi2(1), Pr=0.000) in the Chat (87.04%) than in the Bargaining (53.85%) treatment. It should be noted that, in the Chat, all subjects who complied and believed that the counterpart was going to comply, also had the second-order belief aligned with compliance. Note that the differences in first- and second-order beliefs between the Chat and the Bargaining treatment further confirm the higher effectiveness of the former in favoring the creation of beliefs coherent with compliance, as stated in H4 and commented on in the previous section.

In Table 6 we analyze the relation between the decision to comply with the agreement and the reciprocal alignment of beliefs. Variable *Belief_aligned_compliance* takes the value of 1 for subjects who believed that the counterpart was going to comply (first-order belief) and, at the same time, believed that the counterpart believed that they would comply (second-order belief). The significance of this variable (at 1% level) in the regression presented in Table 6 – column 1, in which the dependent variable is the dummy taking the value of 1 for subjects who complied with the agreement, shows a strict connection between compliance and first-order and second-order beliefs concerning compliance. Moreover, we find that the alignment of belief, despite the differences characterizing the Chat and the Bargaining treatments, is correlated with compliance also when we consider separately the sub-sample of subjects involved in each of these two treatments (Table 6, columns 2 and 3).[22] This evidence supports *H6(*a*).*

[22] When we do not merge subjects who opted for percentages equal to 50% and 100% with subjects who opted for Rule 1 or Rule 2, respectively, the *Belief_aligned_compliance* variable becomes significant at 10% in column 3.

Table 6. Compliance and Beliefs

	(1)	(2)	(3)
Method	Logit	Logit	Logit
Dependent variable	*Compliance*		
	All subjects involved in the Chat and Bargaining treatment	Sub-sample of subjects involved in the Chat	Sub-sample of subjects involved in the Bargaining
Chat	0.601		
	(0.421)		
Belief_aligned_compliance	1.894***	4.449***	1.130**
	(0.431)	(1.169)	(0.558)
Payment_agreement	0.0434*	0.0438	0.0184
	(0.0244)	(0.0508)	(0.0285)
Constant	8.253	4.454	16.04
	(16.15)	(33.82)	(28.41)
Control variables	YES	YES	YES
Observations	156	76	80
Pseudo R^2	0.2260	0.4651	0.2054

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

To investigate *H6(b)*, we compare first-order and second-order beliefs of subjects involved in the two Chat and Bargaining treatments with those of subjects involved in the Noveil treatment. We expect first-order and second-order beliefs to be aligned with the division rule chosen in the ex-post choice significantly more in the two treatments with the agreement than in the Noveil. That is, we expect to observe significantly more in the Chat and Bargaining than in the Noveil treatment subjects who believed: a) that their counterpart was going to choose the same rule that they would choose and b) that their counterpart believed that they were going to choose that rule. Moreover, we expect to find that this evidence is explained by the role of the agreement.

We will test this hypothesis in three different steps. First, we will show that the alignment of belief with the rule actually chosen in the ex-post choice is more likely to be observed in the two treatments characterized by an ex-ante agreement than in the Noveil treatment. Second, we will show that this is due to the subjects whose first-order and second-order beliefs were aligned with the rule chosen in the ex-ante agreement. Therefore, we conclude that the higher probability of observing beliefs reciprocally aligned (and consistent with compliance) in the two treatments with the ex-ante agreement stems exactly from the agreement itself, which generates aligned beliefs concerning compliance with the ex-ante agreement.

As regards the first step, we note that only 17.5% of subjects had first-order and second-order beliefs aligned with the choice in the Noveil treatment. The percentage increases to 27.5% and to 63.16% in the Bargaining and Chat treatment respectively. Overall, the difference in the alignment of beliefs between the treatments with the agreement and the Noveil treatment is statistically significant (Pearson chi2(1), Pr=0.000). However, note that the significance is mainly due to the subjects involved in the Chat treatment. In fact, when we compare the Bargaining and the Noveil treatment we do not find a statistically significance difference characterizing the alignment of belief (Pearson chi2(1), Pr=0.130). Table 7, column 1, analyses the determinants of a dummy variable (*Belief_aligned_division*) capturing the alignment of belief with the ex-post choice (i.e. this variable assumes the value of 1 when first-order belief, second-order belief and ex-post division choice indicate the same rule).[23] The negative and statistically significant coefficient of the *NoVeil* variable in column 1 - Table 7 confirms that being involved in the NoVeil treatment implies a lower alignment of belief with the ex-post choice than being involved in the other two treatments. Column 2 confirms that this result is mainly due to subjects involved in the Chat treatment. In fact, we find that beliefs of subjects involved in the Bargaining treatment are significantly less aligned than beliefs of subjects involved in the Chat, and no differences characterize the alignment of beliefs of subjects in the Bargaining and NoVeil treatments (Wald test on the null hypothesis that the coefficient of *NoVeil* and *Bargaining* is equal to zero: p=0.103)[24].

In order to investigate if the alignment of belief in the Chat and Bargaining treatment is due to the role of the agreement in inducing first-order and second-order beliefs concerning compliance, we include in the estimates the variable *Belief_aligned_compliance*: that is, the previously described variable which identifies subjects who believed that the counterpart was going to comply with the agreement and, at the same time, believed that the counterpart believed that they would comply.[25]Column 3 shows that, when we consider this variable, the relation between the different treatments and the alignment of beliefs completely changes. In particular, when *Belief_aligned_compliance* is included in the analysis, the coefficient of *No_veil* (column 3, Table 7) becomes positive and significant.[26] It reveals that the higher probability of observing the alignment of beliefs with the ex-post choice in the two treatments characterized by the agreement

[23] Note that no subjects who chose the percentage in the ex-post division had first-order and second-order beliefs aligned with their choice.

[24] The difference becomes significant at 10% level when we do not merge subjects who opted for percentages equal to 50% and 100% with subjects who opted for Rule 1 or Rule 2, respectively.

[25] Note that in Table 6 *Belief_aligned_compliance* only concerns subjects involved in the two treatments with the agreement. In Table 7, this variable takes the value of zero for all subjects involved in the Noveil treatment.

[26] The significance of *NoVeil* is slightly lower (6.9%) when we do not merge subjects who opted for percentages equal to 50% and 100% with subjects who opted for Rule 1 or Rule 2, respectively

was entirely due to subjects who had beliefs aligned with compliance with the agreement. When we control for the effect of the compliance with the agreement on the alignment of belief with the ex-post choice through a specific dummy variable (i.e., *Belief_aligned_compliance*), it turns out that the subjects were less likely to have beliefs aligned with the rule chosen in the ex-post choice in the Chat and Bargaining treatments than in the Noveil. Indeed, we conclude that the difference in the alignment of beliefs which emerges between the treatments with the agreement and the NoVeil is due to the ex-ante agreement and the beliefs coherent with compliance with the agreement itself.

Finally, when we distinguish between the Chat and Bargaining treatments (Column 4), we find that the decisive role of the agreement in generating the closer alignment of beliefs (captured by including the variable *Belief_aligned_compliance*in the regression) is confirmed for both the treatments: a) beliefs are more aligned in the Noveil treatment than in the Bargaining treatment (Wald test on the null hypothesis that the coefficient of *NoVeil* and *Bargaining* is equal to zero: p=0.003)[27]; b) no differences in the alignment of belief emerge between subjects involved in the NoVeil and in the Chat.

Table 7. Ex post division choice and Beliefs

	(1)	(2)	(3)	(4)
Dependent variable		*Belief_aligned_division*		
No_veil	-1.495***	-2.100***	1.910**	1.306
	(0.370)	(0.401)	(0.800)	(0.822)
Bargaining		-1.408***		-1.289**
		(0.363)		(0.516)
Belief_aligned_compliance			4.857***	4.863***
			(0.787)	(0.812)
Constant	6.176	6.412	6.401	5.368
	(7.318)	(7.185)	(9.582)	(8.979)
Control variables	YES	YES	YES	YES
Observations	236	236	236	236
Pseudo R^2	0.0969	0.1487	0.4116	0.4328

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

[27] P=0.011 if subjects who opted for percentages equal to 50% and 100% are not merged with subjects who opted for Rule 1 or Rule 2, respectively.

7. Final remarks: the significance of this study for the social contract theory of justice and business ethics.

In this section we consider the meaning and implications of the results of this study in relation to our theoretical background on the social contract theory of justice and the advancement of contractarian business ethics.

7.1. Ex-ante preference for liberal egalitarianism and ex-post compliance

The experiment's main result is confirmation that a significant proportion of standard experimental subjects situated in a context of agreement 'behind a veil of ignorance' reach an agreed solution by opting for a distributive rule based on liberal egalitarianism. This result is remarkable insofar as the agreed solution in this case –as in most real cases in everyday economic life – requires a normatively complex argument regarding arbitrary initial endowments, individual effort, and dictatorial distributive decision.

This seems to confirm the two-tier constitutional/post-constitutional approach to justice (see Section2 and Sacconi 2006, 2011b). It entails a descending hierarchy between the constitutional principle of distribution according to equality of production endowments and the post-constitutional distribution principle according to contribution. Even if subjects must choose a single principle of distributive justice that will regulate the outcome division after production has already taken place, they do not forget the requirement concerning the constitutional stage. Subjects claim redress for the initial injustice of the endowment allocation, asking for the fruit of unequal endowments to be redistributed equally. And they agree that only equal endowments may be used as the basis for a legitimate distribution according to contribution.

It is intuitive that the veil of ignorance, by covering any individual bias, facilitates agreement on egalitarian distributive rules. In the absence of any reason for establishing personal differences, equal distribution is the salient solution in this situation. However, an egalitarian solution would not have been expected in our case, since the experiment involved a task, and the knowledge of each person's contribution, by his/her own effort, to the common output. In this situation, equal distribution is by no means the most salient; the individual right to own output equally is. It could be argued that this is in fact the 'social norm' in most production contexts where individual production can be easily isolated or calculated: each gets what s/he has produced. The device of the veil of ignorance in the experiment made subjects aware of the arbitrary character of one crucial factor in determining individual production, i.e. the endowment of time; and they easily agreed to redress for that arbitrariness, thus restoring the egalitarian intuition, even if the productive nature of

the experiment seemed to cue for a solution based exclusively on individual production. Hence our result supports Rawls's view that pure initial luck is not a rational basis for special claims on the total collective output.

Our second main result is that subjects behave according to agreed principles –in particular, according to the liberal egalitarian principle. This is a crucial contribution to experimental ethics, since the experiment was designed, so to speak, *against* compliance: in the ex-post choice, subjects decided in the role of dictator and there was no 'second round' in which reputation effect could have an impact. Virtually any instrumentally and self-interested rational motivation to comply was removed. According to our hypotheses, compliance can only be explained because the subjects possessed a 'sense of justice' that was activated by the agreement behind the veil. The *ex-ante* agreement was taken as a rational commitment that held *ex-post*. Subjects who believed that their counterpart would act on the agreement were motivated to do so, contrary to the predictions of rational choice theory. And most of the subjects who especially agreed after the chat procedure believed that their counterpart would comply – even if this belief was not in accordance with the hypothesis of rational self-interest. Apparently, mutual trust emerges from deliberative agreement, which somehow elicits both beliefs of counterparty's compliance and a (non-self-interested) preference for compliance, which we identify with the attitude that Rawls called the "sense of justice".

7.2.New insights into contractarian business ethics

The function of the social contract as a philosophical device for the foundation of business ethics is to provide *normative reasons* supportive of certain governance rules. In this case, we focus on the distribution principle of the joint surplus generated by the productive activity in which the corporation consists. Admittedly, we do not account for all the details of joint production in the firm. Nevertheless, the basic issue of how different stakeholders (their social difference being represented by different 'luck' in the initial appropriation of endowments) may agree on a principle for the distribution of a surplus in order to cooperate in its production is represented in its essential features. This principle can be taken as representing the one basic distributive rule of the corporation, or else a component of a wider governance structure. In the contractarian view of the interaction amongst essential stakeholders of the firm, in fact, cooperation is voluntary and must at least in principle be acceptable by all of them. However it is interpreted, our experiment sought to

test whether real subjects, involved in the production of an outcome to be shared, would behave as mainstream twentieth-century contractarianism has predicted.

The experiment makes several simplifying assumptions: we implicitly take the individual cost of effort to be nil or compensable; we also assume, with Rawls, individual talent to be a common asset; while acquired skills are treated as somehow individual. Assumptions of this kind may be objectionable if business ethics is understood as dealing with the conduct of businesses *as they are*. But the social contract tradition in business ethics adopts a critical stance. The contractarian argument is a counter-factual one, of course; and certain simplifications are required to put it to a test. Simplifications allow us to test the 'empirical' force of an agreement behind the veil of ignorance in a way that it could not be tested in real life.

But our conclusions do have a bearing on the moral legitimacy of business. They speak to the many situations in business life where ideal conditions of justice –equality of endowment; equality of basic rights– do not hold. In these situations, distribution principles including redress are conceptually required, and, as our experiment shows, they are supported by laboratory evidence: subjects agree with them and behave in accordance with them when they are playing for money.

If the social contract approach could only serve to identify the choice that would be taken in an ideally fair original position, it would have little relevance to the real world, and to business ethics as well. In the real world, stakeholders face given endowments of rights and production means, and they interact across the corporation, each claiming a remuneration that fairly reflects his/her contribution to the corporate surplus. But what we have in fact shown is that, even if the context is not that of an initial fair allocation of endowments, stakeholders will understand their remuneration – taking fair account of their contribution – in a clearly corrective way. Thus, stakeholders participating in an already-structured social production situation will decide to pay attention to the social contract perspective that requires equality in basic endowments. Hence they will agree to compensate the disadvantaged stakeholders through redistribution of part of the outcome, neutralizing the effects of the initial unfairness.

Apart from the implications entailed by confirmation of the social-contract general hypothesis concerning just distribution in contexts of production, the experiment offers some suggestions for future research in business ethics. First, it recommends the device of the veil of ignorance as a way to elicit fair agreements. Obviously, the veil of ignorance cannot be implemented in reality (in collective bargaining, for instance), but it could be used as a learning device –for example, in

training– to help negotiators gain better understanding of the demands of impartiality. Second, the higher mutual trust (first-order and second-order beliefs on compliance) elicited by the Chat treatment supports deliberative procedures to deal with conflicting distributive issues. Even in a deliberative situation, we claim that anonymity plays the key role in eliciting the agreement and compliance. Face-to-face discussion may hinder impartial agreement by turning the focus to particular interests. This should be ascertained in the future by means of laboratory or field experiments.

References

Anderson, F., & Lyttkens, C.H. 1999.Preferences for equity in health behind a veil of ignorance.***Health Economics,***8(5): 369-78.

Aoki, M. 1984.***The Co-operative Game Theory of the Firm*.** Oxford: Oxford University Press.

Aoki, M. 2001.***Comparative Institutional Analysis*,** Cambridge Mass: The MIT press.

Aoki, M. 2010.***Corporations in Evolving Diversity, Cognition, Governance, and Institutions*.** Oxford: Oxford University Press.

Bacharach, M. 1994.The Epistemic Structure of a Game.***Theory and Decisions*,** 37: 7-48.

Bacharach, M., Gold N., & Sugden, R. 2006.Beyond Individual Choice: Teams and Frames in Game Theory**.** Princeton: Princeton University Press.

Baur, D., & Palazzo, G. 2011. The Moral Legitimacy of NGOs as Partners of Corporations.***Business Ethics Quarterly*,** 21(4): 579-604.

Binmore, K. 1991. Game Theory and the Social Contract. In R. Selten (Ed.), ***Game Equilibrium Models II, Methods, Morals, Markets*.** Berlin: Springer Verlag.

Binmore, K. 2005. ***Natural Justice*.** Oxford: Oxford University Press.

Bishop, J.D. 2008. For-profit corporations in a just society: a social contract argument concerning the rights and responsibilities of corporations. ***Business Ethics Quarterly*,**18(2): 191–212.

Bond, D., & Park, J. 1991. An Empirical Test of Rawls's Theory of Justice: A Second Approach, in Korea and the United States. ***Simulation Gaming*,** 22(4): 443-462.

Brickman, P. 1977. Preference for Inequality.***Sociometry*,** 40(4): 303-310.

Brock, H.W. 1979.A Game Theoretical Account of Social Justice.***Theory and Decision*,** 11: 239-265.

Buchanan, J.M. 1975. ***The Limits of Liberty, Between Anarchy and Leviathan*.** Chicago: The University of Chicago Press.

Cappelen, A. W., Moene, K. O., Sørensen, E. Ø., & Tungodden, B. (2014). Just Luck : An Experimental Study of Risk Taking and Fairness. ***American Economic Review***, *124*(4), 1398–1413.

Choen, G.A. 1989. On the Currency of Egalitarian Justice.***Ethics,*** 99: 906-944.

De La Cruz-Dona, R., & Martina, A. 2000. Diverse groups agreeing on a system of justice in distribution: Evidence from the Philippines. ***Journal of Interdisciplinary Economics*,** 11: 35-76.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G.G. 2011. Individual risk attitudes: measurement, determinants and behavioral consequences. ***Journal of the European Economic Association,*** 9(3): 522–550.

Donaldson, T. 1982.***Corporations and Morality*.** Englewood Cliffs, N.J.: Prentice Hall.

Donaldson, T., & Dunfee, T.W. 1999.***Ties that Bind: A Social Contracts Approach to Business Ethics*.** Cambridge, MA: Harvard University Business School Press.

Donaldson T., & Preston, L. 1995, The stakeholder Theory of the Corporation: Concepts, Evidence, and Implications. ***Academy of Management Review*,** 20(1): 65-91.

Dunfee, T.W. 2006. A Critical Perspective of Integrative Social Contracts Theory: Recurring Criticisms and Next Generation Research Topics. ***Journal of Business Ethics*,**68(3): 303–328.

Durante, R., Putterman, L., & Van der Weele, J. 2014. Preferences for Redistribution and Perceptiono Fairness: An Experimental Study**.** ***Journal of the European Economic Association,***12(4): 1059-1086.

Dworkin, R. 1981a. What is Equality? Part 1: Equality of Welfare. ***Philosophy & Public Affairs,***10(3): 185-246.

Dworkin, R. 1981b. What is Equality? Part 2: Equality of Resources. ***Philosophy & Public Affairs,***10(4): 283-345.

Dworkin, R. 2013.***Justice for Hedgehogs*.** Harvard (Mass.): Harvard University Press.

Faillo, M., Ottone, S. & Sacconi, L. 2015. The social contract in the lab.An experimental analysis of self-enforcing impartial agreements.***Public Choice*,** 163 (3-4): 225-246.

Francés-Gómez, P., Sacconi, L. & Faillo, M. 2015. Experimental economics as a method for normative busniess ethics.***Business Ethics: A European Review,*** 24 (S1): 41-53.

Freeman, R.E. 1994. The politics of stakeholder theory: Some future directions. ***Business Ethics Quarterly***, 4: 409–421.

Freeman, R.E., & Evan, W.M. 1990. Corporate governance: A stakeholder interpretation. ***Journal of Behavioral Economics,***I/19: 337–359.

Freeman, R.E. 1984. ***Strategic Management: A Stakeholder Approach*,** Boston: Pitman.

Freeman, R.E., & Evan, P. 1988. Stakeholder Management and the Modern Corpora-tion: Kantian Capitalism. In T. Beuchamp & N. Bowie (Eds.), ***Ethical Theory and Business*.**(3rd ed.). Englewood Cliffs N.J.: Prentice Hall.

Freeman, R. E., Wicks, A. C., & Parmar, B. 2011.Stakeholder Theory as a Basis for Capitalism. In L. Sacconi, M. Blair, R.E. Freeman, & A. Vercelli (Eds.), ***Corporate Social Responsibility and***

Corporate Governance. The Contribution of Economic Theory and Related Disciplines. (pp. 52–72). New York: Palgrave MacMillan.

Frohlich, N., & Oppenheimer, J.A. 1990. Choosing Justice in Experimental Democracies with Production. ***American Political Science Review*****,** 84(2): 461-477.

Frohlich, N., & Oppenheimer, J.A. 1992. *Choosing Justice: An Experimental Approach to Ethical Theory.* Berkeley: University of California Press.

Frohlich, N., Oppenheimer, J.A., & Eavey, C.L. 1987. Choices of Principles of Distributive Justice in Experimental Groups.***American Journal of Political Science*****,** 31(3): 606-636.

Geanakoplos, J., Pearce, D., & Stacchetti, E. 1989. Psychological Games and Sequential for Non-Cooperative Games.***International Journal of Game Theory,*** 5: 61–94.

Grimalda G.L., & Sacconi, L. 2005. The Constitution of the Not-For-Profit Organisation: Reciprocal Conformity to Morality. ***Constitutional Political Economy*****,** 16(3): 249-276.

Habermas, J. 1996. ***Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*****.** Cambridge, MA: MIT Press.

Herne, K., & Mard, T. 2008. Three Versions of Impartiality: An Experimental Investigation. ***Homo Oeconomicus*****,** 25(1): 27-53.

Herne, K., & Suojanen, M. 2004.The Role of Information in Choices Over Income Distributions.***Journal of Conflict Resolution*****,** 48(2): 173-193.

Holyoak, K.J., & Spellman, B.A. 1993. Thinking. In L. W. Porter & M.R. Rosenzweig (Eds.), ***Annual review of psychology,*** 44: 265-315. San Diego, CA: Academic Press.

Jackson, M. & Hill, P. 1995. A Fair Share, *Journal of Theoretical Politics,* 7 (2), 169-180.

Johnson-Laird, P.N., 1983.*Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*, Harvard University Press, 1983

Johnson-Laird, P.N., & Byrne, R.M.J. 1991.*Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Keeley, M. 1988.***A Social-Contract Theory of Organizations*****.** Notre Dame (Indiana): University of Notre Dame Press.

Konow, J. 1996. A Positive Theory of Economic Fairness.***Journal of Economic Behavior and Organization*****,** 31(1): 13-35.

Konow, J. 2000. Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions.*American Economic Review*, vol. 90(4), pages 1072-1091.

Konow, J. 2001. Fair and Square: The Four Sides of Distributive Justice. ***Journal of Economic Behavior and Organization*****,** 46(2): 137-164.

Konow, J. 2003. Which Is the Fairest One of All? A Positive Analysis of Justice Theories.*Journal of Economic Literature, American Economic Association*, vol. 41(4), pages 1188-1239.

Konow, J. 2005. Blind spots: The effects of information and stakes on fairness bias and dispersion. ***Social Justice Research***, *18*(4), 349–390.

Legrenzi P., Girotto V., & Johnson-Laird, P.N. 1993. Focussing in Reasoning and Decision Making. ***Cognition*****,** 49: 37-66.

Lissowski, G., Tyszka, T., & Okrasa, W. 1991.Principles of Distributive Justice.***Journal of Conflict Resolution*****,** 35(1): 98-119.

Lütge, C. 2012.The Idea of a Contractarian Business Ethics. In *Handbook of the Philosophical Foundations of Business Ethics*: 647–658. Springer Netherlands.

Lütge, C. 2015.*Order Ethics or Moral Surplus. What Holds a Society Together?* Lanham (Maryland): Lexington Books.

Michelbach, P.A., Scott, J.T., Matland, R.E., & Bornstein, B.H. 2003. Doing Rawls Justice: An Experimental Study of Income Distribution Norms. *American Journal of Political Science***,** 47(3): 523-539.

Nash, J. 1950. The Bargaining Problem.***Econometrica*****,** 18: 155-162.

Phillips, R., 2003.***Stakeholder Theory and Organizational Ethics*****.** San Francisco: Berrett-Koehler Publishers.

Rabin, M. 1993. Incorporating Fairness into Game Theory and Economics. ***American Economic Review*****,** 83: 1281-1302.

Rawls, J. 1971. ***A Theory of Justice*****.** Cambridge Mass.: Harvard University Press.

Reiter, R. 1980.A Logic for Default Reasoning. ***Artificial Intelligence*****,** 13: 81-132.

Roemer, J. 1996. ***Theories of Distributive Justice*****.**Cambridge Mass: Harvard University Press.

Sacconi, L. 2000. *The Social Contract of the Firm. Economics, Ethics and Organisation*. Berlin: Springer Verlag.

Sacconi, L. 2006b. A Social Contract Account For CSR as Extended Model of Corporate Governance (Part I): Rational Bargaining and Justification. ***Journal of Business Ethics*****,** 68(3): 259-281

Sacconi, L. 2011a. A Rawlsian view of CSR and the Game Theory of its Implementation (Part I): The Multistakeholder Model of Corporate Governance. In L. Sacconi, M. Blair, E. Freeman & A. Vercelli (Eds.), ***Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*****.** Basingstoke: Palgrave Macmillan.

Sacconi, L. 2011b. A Rawlsian View of CSR and the Game Theory of Its Implementation (Part II): Fairness and Equilibrium. In L. Sacconi, M. Blair, E. Freeman & A. Vercelli (Eds.), ***Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*.** Basingstoke: Palgrave Macmillan.

Sacconi, L. 2011c. A Rawlsian View of CRS and the Game of its Implementation (Part III): Conformism and Equilibrium Selection. In L. Sacconi and G. DegliAntoni (Eds.), ***Social Capital, Corporate Social Responsibility, Economic Behavior and Performance.*** Basingstoke: Palgrave Macmillan.

Sacconi, L., & Moretti, S. 2008. A Fuzzy Logic and Default Reasoning Model of Social Norms and Equilibrium Selection in Games under Unforeseen Contingencies. ***International Journal of Uncertainty, Fuzziness and Knowledge Based Systems,***16 (1): 59–81.

Sacconi, L., & Faillo, M. 2010.Conformity, reciprocity and the sense of justice. How social contract-based preferences and beliefs explain norm compliance: the experimental evidence. ***Constitutional Political Economy*,** 21(2): 171-201.

Sacconi, L., Faillo, M., & Ottone, S. 2011. Contractarian compliance and the 'sense of justice': a behavioral conformity model and its experimental support. ***Analyse&Kritik*,**33(1): 273-310.

Scherer, A.G., & Palazzo, G. 2007. Toward a Political Conception of Corporate Responsibility: Business and Society Seen from a Habermasian Perspective. ***The Academy of Management Review*,***32*(4): 1096–1120.

Schildberg-Hörisch, H. 2010. Is the veil of ignorance only a concept about risk? An experiment. ***Journal of Public Economics*,** 94(11-12): 1062-1066.

Scott, J.T., Matland, R.E., Michelbach, P.A., & Bornstein, B.H. 2001. Just Deserts: An Experimental Study of Distributive Justice Norms. ***American Journal of Political Science*,** 45(4): 749-767

Searle, J.R. 2005. ***Rationality in Action*.** Cambridge Mass: The MIT press.

Traub, S., Seidl, C., Schmidt, U., Levati, M. 2005. Friedman, Harsanyi, Rawls, Boulding – or Somebody Else?An Experimental Investigation of Distributive Justice.***Social Choice and Welfare,*** 24(2): 283-309.

Tversky, A. & Kahneman, D. 1981. The framing of decisions and the psychology of choice.***Science,*** 211.4481: 453-458.

Ulrich, P. 2008. ***Integrative Economic Ethics. Foundations of a civilized market economy*.**Cambridge University Press.

Wempe, B., 2005.In defense of a self-disciplineddomain-specific social contracttheory of business ethics.***Business Ethics Quarterly,*** 15:113–135.

Yaari, M.E., & Bar-Hillel, M. 1984. On Dividing Justly. ***Social Choice and Welfare,*** 1, 1-24.

Appendix 1. Theoretical views that underpin hypothesis H6(b)

As noted in section 5, hypothesis H6(b) may conflict with standard economic rationality. An ex-ante unenforced agreement would not be complied with unless it is consistent with self-interested ex-post incentives – which is not the case in this experiment. Moreover, under the assumption of symmetrical economic rationality, a subject would not expect others to conform with an agreement on rule 4, and would not believe that s/he is expected to comply with it. Thus, what justifies observations concerning H6(b)?

First, the intentional explanation of action (Searle 2005) suggests in our case that the content of the ex-ante agreement, i.e. dividing an outcome according to a rule, is a commitment to dividing the outcome according to the rule later on. Hence it is an intention to act. The agreement means undertaking a commitment. Having accepted (for some reason, such as impartiality, fairness, etc.) to subscribe to an agreement amounts to having a reason to act upon the corresponding commitment. A commitment is an intentional state for an action, which is not a desire-based, but a commitment-based reason to act (Searle 2005). By no means is this reason to act the only logically possible intention explaining action. Nevertheless, such an intentional state may translate into a preference to act upon the commitment and hence may be effective - among other intentions - in causing action. Only a free deliberation may pick this out of the admissible set of reasons to act, thus 'filling the gap' (Searle 2005) and letting it produce actual conduct.

Second, this intentional explanation constitutes a mental model (Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991,Holoyak and Spellman, 1993, Legrenzi et.al, 1993). Subjects, having agreed ex ante, may hold this model as the basis for interpreting and predicting ex-post actions of other subjects, consistently with the evidence that they have agreed on a rule. By no means is this the only logically possible interpretation of the situation. But agents do not have enough thinking resources to consider all the logically possible state of affairs; and as a matter of fact this is the mostly immediate intentional interpretation of their behavior elicited by the content of the agreement to which they have subscribed. Moreover, it fits the situation: it makes sense of the behavior of both the self and other agents by giving an intentional interpretation of their action.

Third, the cognitive mechanism at work is *framing* (Bacharach 2006): because of the agreement, it comes to the subject's mind the *frame* of an agent who acts upon a commitment, and hence has the intention to carry out the commitment. A frame delimits the ways in which a subject may 'see' or understand a given situation. In this case the frame coming to the agent's mind is that subjects are

intentional agents acting upon the undertaken commitment, and as long as the situation is framed this way, there is no room for explaining subjects as agents pursuing their self-interest in the ex-post decision.

Fourth, a frame defines the (necessarily incomplete) delimited base of knowledge whereby any default, fallible but nevertheless reasonable, prediction of the subjects' behavior must be drawn by inference (Reiter 1980, Bacharach 1994, Sacconi 2000, Sacconi and Moretti 2008). A default inference works as follows: as long as there is no evidence contrary to the assumption that subjects satisfy the model of an intentional agent acting upon commitments, nothing contradicts that, if an agent has the commitment-based intention to act according an agreed rule, s/he will in fact carryout the rule. Whence a default reasoner derives the prediction that subjects will act according to their commitment. It may be wrong, of course. But this is the simplest intentional explanation and the only one consistent with the framed mental model of an intentional agent that delimits the 'base of knowledge' held by subjects.

Summing up, moral reasoning behind the veil of ignorance leads mostly to an agreement on the liberal egalitarian rule, which is a commitment to redressingan unequal allocation of endowments ex post. Such a commitment provides a basis for a reasonto act that may translate into a preference, so that the agent acts not on a desire-based intention but on a commitment-based intention that may engender a preference(a reason-to-act-based preference). At the same time the commitment-based intentional explanation constitutes a model for understanding other agents' behavior. Well-known cognitive constraints on reasoning, however, explain why this model rules out other in principle possible predictions of other agents' behavior. Thus, as long as no contradictory evidence unfolds – i.e. by default – subjects expect that because they have agreed on a rule (and hence committed themselves to carrying out the rule) they will act accordingly. But this completes the picture about the emergence of the conditional desire to comply. Since the agent expects mutual conformity, the commitment-based intention is selected as the one effectively determining his/her choice, and hence how *de facto*s/he desires to behave. This is not only consistent with the 'sense of justice' idea but also explains why so many subjects behaved consistently with this idea in our experiment.

Appendix 2 – Variable legend and descriptive statistics

	Legend	*Obs.*	*Mean*	*St. dev.*	*Min.*	*Max.*
Bargaining	DV=1 if subjects took part in the Bargaining treatment	236	0.339	0.474	0	1
Chat	DV=1 if subjects took part in the Chat treatment	236	0.322	0.468	0	1
No_veil	DV=1 if subjects took part in the Noveil treatment	236	0.339	0.474	0	1
Belief_first=agr	DV=1 if the subject believes that the other player in the pair is going to comply	156	0.705	0.457	0	1
Belief_second=agr	DV=1 if the subject believes that the other player believes that s/he is going to comply	156	0.737	0.442	0	1
Belief_aligned_compliance	DV=1 if *Belief_first=agr*=1 and *Belief_second=agr*=1	156	0.571	0.497	0	1
Belief_aligned_division	DV=1 when first-order belief, second-order belief and ex-post division choice indicate the same rule	236	0.356	0.480	0	1
Compliance	DV=1 if subjects comply with the rule they agreed on in the ex-ante agreement	156	0.596	0.492	0	1
Payment_agreement	the payoff – in experimental tokens - associated with the rule agreed in the ex-ante agreement	156	43.291	12.985	16.47	87
Rule_agr_1	Dummy variables equal to 1 if subjects involved in the Chat or in the Bargaining treatment opt for one of the five different rules in the ex-ante agreement	156	0.141	0.349	0	1
Rule_agr_2		156	0.013	0.113	0	1
Rule_agr_3		156	0.103	0.304	0	1
Rule_agr_4		156	0.577	0.496	0	1
Rule_agr_5		156	0.167	0.374	0	1
Rule_4_ex_ante-post	DV=1 if the division consistent with Rule 4 is chosen in the Noveil treatment or in the ex-ante agreement in the Chat and Bargaining treatment	236	0.432	0.496	0	1
Rule_4_ex-post	DV=1 if a division consistent with Rule 4 is selected in the ex-post choice	236	0.318	0.467	0	1

10_minutes	DV=1 if the subject is endowed with 10 minutes of time	236	0.500	0.501	0	1
10_minutes_chat	DV=1 if the subject is endowed with 10 minutes of time and is involved in the Chat treatment	236	0.161	0.368	0	1
	Control variables included in all the estimates					
Age	Subject's age in years	236	20.682	2.488	18	33
Experiment	DV=1 if the subject has already taken part in Lab experiment	236	0.242	0.429	0	1
Female	DV=1 if the subject is a female	236	0.521	0.501	0	1
Income	Income level of the subject's household, on a 5-level scale between 1 (less than 17,000€) and 5 (more than 120,000€)	236	1.911	0.925	1	5
No_religious	DV=1 if the subject is not a believer	236	0.487	0.501	0	1
Productivity	Encrypted words per minute	236	4.586	0.958	1.33	6.83
Risk	Risk aversion measure based on the following question: "Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on the scale, where the value 0 means: 'unwilling to take risks' and 10: 'fully prepared to take risk"(see Dohmen et al., 2011).	236	6.458	1.846	0	10
Trust	DV=1 if the subject declares that, generally speaking, most people can be trusted	236	0.233	0.424	0	1
Words	Number of words encrypted in the task	236	36.788	12.451	8	64

Appendix3. Full estimates results

Table 2.
Ex-post choices in the Noveil treatment vs. ex-ante agreement in the Bargaining and Chat treatments

	(1)
	Logit
Dependent variable: *Rule_4_ex_ante-post* - DV=1 if the division consistent with Rule 4 is chosen in the Noveil treatment or in the ex-ante agreement in the Chat and Bargaining treatments	
Chat	1.886***
	(0.397)
Bargaining	1.847***
	(0.406)
Female	0.495
	(0.309)
Age	-0.635
	(0.649)
Age2	0.0137
	(0.0146)
Income	0.0531
	(0.167)
No_religious	-0.0160
	(0.297)
Trust	0.211
	(0.360)
Risk	0.0428
	(0.0820)
Experiment	-0.0218
	(0.383)
Words	-0.00600
	(0.0158)
Productivity	0.178
	(0.211)
Constant	4.311
	(7.180)
Observations	236
Pseudo R^2	0.1386
Chat-Bargaining	0.038
	(0.343)

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 3. Determinants of choice of the rule

	(1)	(2)
	Logit	Logit
Dependent variable:	*Rule_4_ex-post* - DV=1 if a division consistent with Rule 4 is selected in the ex-post choice	
Chat	1.231***	-0.344
	(0.408)	(0.545)
Bargaining	1.201***	-0.326
	(0.417)	(0.553)
Female	0.469	0.347
	(0.319)	(0.356)
Age	-1.365**	-1.552*
	(0.650)	(0.792)
Age2	0.0295**	0.0345*
	(0.0146)	(0.0180)
Income	0.114	0.124
	(0.170)	(0.189)
No_religious	-0.0833	-0.0747
	(0.304)	(0.338)
Trust	0.581	0.696*
	(0.360)	(0.402)
Risk	0.0333	0.0383
	(0.0843)	(0.0962)
Experiment	-0.234	-0.115
	(0.407)	(0.446)
Words	-0.0169	-0.0172
	(0.0161)	(0.0180)
Productivity	0.122	0.106
	(0.213)	(0.239)
Rule_agr_4		2.485***
		(0.448)
Constant	13.12*	14.80*
	(7.184)	(8.684)
Observations	236	236
Pseudo R^2	0.0904	0.2262
Chat-Bargaining	0.030	-0.018
	(0.352)	(0.405)

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 5. The determinants of compliance

	(1)	(2)	(3)	(4)	(5)
Method	Logit	Logit	Logit	Logit	Logit
Dependent variable	*Compliance*				
	All subjects involved in the Chat and Bargaining treatment			Sub-sample of subjects who opted for Rule 4 in the ex-ante agreement	
Chat	0.973**	0.553	0.553	-0.194	-1.490
	(0.407)	(0.470)	(0.470)	(0.655)	(0.986)
10_minutes			0.693	-3.174	-3.380
			(2.389)	(3.804)	(3.843)
10_minutes_chat					2.487*
					(1.283)
Rule_agr_1	1.720**	2.045**	2.113**		
	(0.731)	(0.837)	(0.872)		
Rule_agr_2	-2.393	-1.463	-1.588		
	(2.045)	(2.430)	(2.501)		
Rule_agr_3	0.280	-0.124	-0.0865		
	(0.670)	(0.730)	(0.741)		
Rule_agr_5	0.0659	0.0575	0.113		
	(0.638)	(0.718)	(0.748)		
Female	0.471	0.949**	0.952**	0.502	0.779
	(0.411)	(0.472)	(0.473)	(0.642)	(0.678)
Age	-0.245	-0.415	-0.466	-2.242	-2.539
	(1.367)	(1.589)	(1.600)	(2.939)	(3.053)
Age2	0.00347	0.00755	0.00872	0.0490	0.0574
	(0.0317)	(0.0368)	(0.0371)	(0.0693)	(0.0719)
Income	0.350	0.438	0.422	0.423	0.392
	(0.241)	(0.275)	(0.281)	(0.415)	(0.422)
No_religious	-0.706*	-0.746	-0.742	-0.642	-0.372
	(0.389)	(0.464)	(0.463)	(0.665)	(0.702)
Trust	0.840*	0.819	0.826	1.315	1.243
	(0.491)	(0.551)	(0.551)	(0.860)	(0.871)
Risk	-0.0644	-0.0240	-0.0323	0.0559	0.101
	(0.109)	(0.122)	(0.125)	(0.201)	(0.208)
Experiment	-0.773	-0.781	-0.779	-0.749	-0.947
	(0.514)	(0.589)	(0.589)	(0.897)	(0.935)
Words	-0.0725**	-0.0770**	-0.115	0.188	0.0815
	(0.0329)	(0.0363)	(0.136)	(0.292)	(0.301)
Productivity	0.411	0.407	0.703	-1.060	-0.799
	(0.288)	(0.318)	(1.069)	(1.624)	(1.653)
Payment_agreement	0.0648*	0.0567	0.0595	-0.0503	0.00188
	(0.0354)	(0.0389)	(0.0401)	(0.151)	(0.160)
Belief_first=agr		0.597	0.572	1.116	0.978
		(0.492)	(0.499)	(0.691)	(0.721)
Belief_second=agr		2.310***	2.310***	2.926***	2.878***
		(0.530)	(0.532)	(0.831)	(0.854)
Constant	1.145	0.847	1.037	23.37	26.07
	(14.66)	(17.08)	(17.12)	(30.80)	(31.94)
Observations		156	156	90	90
Pseudo R^2	0.1596	0.2930	0.2934	0.3263	0.3596

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 6. Compliance and Beliefs

	(1)	(2)	(3)
Method	Logit	Logit	Logit
Dependent variable	*Compliance*		
	All subjects involved in the Chat and Bargaining treatment	Sub-sample of subjects involved in the Chat	Sub-sample of subjects involved in the Bargaining
Chat	0.601		
	(0.421)		
Belief_aligned_compliance	1.894***	4.449***	1.130**
	(0.431)	(1.169)	(0.558)
Female	0.673	1.019	1.091*
	(0.429)	(0.886)	(0.617)
Age	-0.974	-1.024	-1.490
	(1.506)	(3.137)	(2.676)
Age2	0.0231	0.0294	0.0334
	(0.0350)	(0.0733)	(0.0630)
Income	0.390	1.351**	-0.0237
	(0.254)	(0.580)	(0.378)
No_religious	-0.961**	-0.844	-0.867
	(0.420)	(0.831)	(0.567)
Trust	0.575	0.372	0.679
	(0.511)	(0.940)	(0.790)
Risk	-0.0364	0.598**	-0.321*
	(0.114)	(0.293)	(0.176)
Experiment	-0.875	-2.806**	-0.460
	(0.545)	(1.147)	(0.769)
Words	-0.0580**	-0.0555	-0.0500
	(0.0274)	(0.0518)	(0.0382)
Productivity	0.214	-0.706	0.578
	(0.290)	(0.672)	(0.395)
Payment_agreement	0.0434*	0.0438	0.0184
	(0.0244)	(0.0508)	(0.0285)
Constant	8.253	4.454	16.04
	(16.15)	(33.82)	(28.41)
Observations	156	76	80
Pseudo R^2	0.2260	0.4651	0.2054

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 7. Ex post division choice and Beliefs

Dependent variable	(1)	(2)	(3)	(4)
	Belief_aligned_division			
No_veil	-1.495***	-2.100***	1.910**	1.306
	(0.370)	(0.401)	(0.800)	(0.822)
Bargaining		-1.408***		-1.289**
		(0.363)		(0.516)
Belief_aligned_compliance			4.857***	4.863***
			(0.787)	(0.812)
Female	-0.330	-0.149	0.179	0.293
	(0.306)	(0.323)	(0.403)	(0.416)
Age	-0.644	-0.563	-1.015	-0.852
	(0.664)	(0.649)	(0.871)	(0.812)
Age2	0.0145	0.0120	0.0235	0.0200
	(0.0150)	(0.0145)	(0.0197)	(0.0181)
Income	0.216	0.164	0.397*	0.371*
	(0.162)	(0.168)	(0.206)	(0.210)
No_religious	0.0418	-0.0884	-0.218	-0.306
	(0.295)	(0.310)	(0.389)	(0.399)
Trust	0.578*	0.582	0.288	0.273
	(0.349)	(0.362)	(0.451)	(0.466)
Risk	-0.0744	-0.0526	-0.00575	0.0146
	(0.0813)	(0.0844)	(0.105)	(0.108)
Experiment	0.0297	0.177	0.503	0.563
	(0.382)	(0.393)	(0.501)	(0.500)
Words	-0.00220	0.000331	-0.00336	-0.00591
	(0.0157)	(0.0164)	(0.0205)	(0.0213)
Productivity	0.182	0.0981	0.00224	-0.0425
	(0.211)	(0.221)	(0.274)	(0.280)
Constant	6.176	6.412	6.401	5.368
	(7.318)	(7.185)	(9.582)	(8.979)
Observations	236	236	236	236
Pseudo R^2	0.0969	0.1487	0.4116	0.4328

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

The instructions and the questionnaire actually used in the experiment are available on a pdf file here:

https://goo.gl/v9SIIV