Laura Marcon, Pedro Francés-Gómez, and Marco Faillo*

Distributive Justice in the Lab: Testing the Binding Role of Agreement

https://doi.org/10.1515/auk-2020-0005

Abstract: Lorenzo Sacconi and his coauthors have put forward the hypothesis that impartial agreements on distributive rules may generate a conditional preference for conformity. The observable effect of this preference would be compliance with fair distributive rules chosen behind a veil of ignorance, even in the absence of external coercion. This paper uses a Dictator Game with production and taking option to compare two ways in which the device of the veil of ignorance may be thought to generate a motivation for, and compliance with a fair distributive rule: individually—as a thought experiment that should work as a moral cue—and collectively—as an actual process of agreement among subjects. The main result is that actual agreement proves to be necessary for agents to be led towards a fair distributive principle and to generate a significant amount of compliance in absence of external authority. This conclusion vindicates the role of actual agreements in generating motivational power in correspondence with fair distributive rules.

Keywords: distributive justice, veil of ignorance, social norms, compliance, experiments

1 Introduction

Distributive justice deals with allocation problems. A theory of distributive justice is "concerned with what rules, procedures, or mechanisms a society or group should use to allocate its scarce resources, commodities, and necessary burdens to individuals with competing needs and claims" (Oleson 2001, 13). From an empirical perspective, to understand how principles of distributive justice might provide moral guidance for social, economic and political structures, a possibility is

Laura Marcon, Department of Philosophy I, Campus de la Cartuja, Granada, Spain; Department of Economics and Management, University of Trento, Italy, e-mail: marcon4@ugr.es; laura.marcon@unitn.it
Pedro Francés-Gómez, Department of Philosophy I, Campus de la Cartuja, Granada, Spain, e-mail: pfg@ugr.es
***Corresponding author: Marco Faillo,** Department of Economics and Management, University of Trento, Italy, e-mail: marco.faillo@unitn.it

to focus on individuals' moral psychology; exploring whether and to what extent people are motivated by considerations of justice.[1] In this context, some interesting results have been obtained by introducing reasoning behind the 'veil of ignorance' (Voigt 2015; Huang et al. 2019). Some of these studies have shown that agreement in conditions of ignorance may play a role in determining both individual views about justice and individual motivation to act justly.

Drawing upon Degli Antoni et al.'s (2016)[2] work, the purpose of this paper is to test whether the collective unanimous choice behind a veil of ignorance (impartial agreement) can create the conditions to bring about a conception of distributive justice consistent with 'liberal egalitarianism' that is effective in motivating agents to actually behave as prescribed in absence of positive (rewards) or negative (punishment) incentives.

By 'liberal egalitarianism', we refer to theories of justice that combine two concerns: the *equal* distribution of basic resources and opportunities, and the possibly *unequal* distribution of resources derived from legitimate individual entitlements. We follow Degli Antoni et al. (2016, 8) and define a 'liberal egalitarian' rule that is made up of two normative demands: a principle of equality in resources, and an allocation criterion related to contribution.

In their experiment Degli Antoni et al. address two types of moral choice: the ex-ante collective choice of a rule for the distribution of a common output of a productive activity (Charness et al. 2018; Erkal et al. 2011), and the ex-post compliance with the agreed rule, in absence of a coercive authority. The main characteristic of their experimental design is that "subjects are assigned unequal endowments [working time] for which they are not responsible; the assignment is random. At the same time, their work naturally generates unequal levels of earnings" (Degli Antoni et al. 2016, 1). The aim of the study was to test a general hypothesis according to which the choice of a liberal egalitarian rule and its actual implementation is more likely when the rule is collectively chosen by means of an impartial unanimous agreement behind a veil of ignorance.

This general hypothesis is based on Sacconi (2011) and colleagues' (Grimalda/Sacconi 2005; Sacconi/Faillo 2008; 2010; Sacconi/Grimalda 2007; Sacconi et al. 2011) theory of conformist preferences. This theory can be characterized as a moral theory about the nature of fair distributive decisions affecting the agent and third parties. These decisions are supposed to be motivated by a preference for norm compliance when certain conditions obtain. The conditions include the

1 There is a very extensive literature on this. Cf. for example Frohlich/Oppenheimer 1992; Fehr/Schurtenberger 2018.

2 In this paper we refer both to Degli Antoni/Faillo/Francés-Gómez/Sacconi 2016, and to a revised version of it Degli Antoni et al. forthcoming.

institution of a fair distributive norm in an impartial situation (behind a veil of ignorance) in which the agent herself takes part. Once the norm is instituted this way, Sacconi argues, the subsequent interaction is not framed by the agents as a simple distributive decision, but as a moralized social interaction in which the choice the individual faces is not so much how to distribute a monetary payoff, but whether or not to comply with the agreed norm. In other words, agents in this condition stop acting instrumentally to maximize their self-interest and start acting on principle. Under this framing, compliance becomes the rational choice for most individuals, as long as they expect reciprocity. This hypothesis has already received some experimental support (Sacconi/Faillo/Ottone 2011).

The theory of conformist preferences focuses on the individual capacity of acting upon (moral) reasons. Individuals are supposed to possess what Rawls (1971, ch. 8) named 'a sense of justice': a practical disposition to abide by fair principles, provided they are shared by (most) other members of society. This theory can explain the emergence of moral norms in a way that is under-studied by conventionalist views on social norms (Cialdini et al. 1991; Young 2007), in particular by Bicchieri's approach (2006; 2008; 2016).

According to Bicchieri's theory, for example, distributive criteria would be a sub-set of social norms; they are adhered to by individuals depending on context and mutual expectations (both empirical and normative). Explicit agreements might work as part of a social framework complete with sanctions and mutual expectations for compliance; and explicit agreements can of course be the reason why a particular norm becomes salient. But in general, norms are supposed to emerge through habit and repeated observation of other's behaviour.

The theory of conformist preferences shows that fair agreements may institute norms with an immediate normative force that cannot be derived from their being pre-existing convention. This is relevant both for moral philosophy and for a theory of social norms. Showing that conformist preferences follow from an impartial agreement would add empirical support to the philosophical speculation that principles of justice may be stable given the capacities of ordinary people. Regarding the theory of social norms, the theory of conformist preferences highlights the role of explicit impartial agreements in generating norms that gather ample support, and sheds light on the nature of moral norms: they can be seen as joint commitments rather than purely personal preferences.

In this paper we start from the evidence collected by Degli Antoni et al., maintaining the overall experimental design. We assume accordingly that it is possible to argue that people do possess the capacity to converge on a distributive rule and that they would comply with it out of a conformist preference.

Now, the hypothesis of conformist preferences is linked to the idea that impartial agreements (brought about behind a veil of ignorance that hides whatever may

bias the bargaining process) yield fair norms. Once these norms are in place, they become salient and conditionally preferred. The conditionality refers to the mutual expectation of compliance; if agents have reasons to suspect that compliance is not to be reciprocated, they would stop seeing compliance as rational. However, the question remains whether the salience of fair norms requires an actual agreement, or may also be the result of individual reflection. After all, philosophical theories of justice are not drafted through collective bargain; they come out of arm-chair philosophy. Why would not ordinary folk's sense of justice be activated by the thought of an impartial agreement, rather than by the actual engagement in one?

Let us note that if this is the case, if impartial reflection alone (a sympathetic attitude towards peers) would generate a fairer attitude, as revealed by a disposition to accept egalitarian distributive rules even at a cost to oneself, then Bicchieri's view of norms might be vindicated. The reason is that this egalitarian disposition may be attributed to purely individual moral preferences that agents activate when prompted by the adequate script and context.

The experimental results reported in this article indicate that the veil of ignorance—the impartial attitude it induces—is a necessary condition for the choice and implementation of a liberal egalitarian rule, but it is not sufficient. It is effective only in combination with actual collective deliberation through impartial agreements. Our results reveal that, when reasoning behind the veil is applied individually, the liberal egalitarian rule is neither chosen nor implemented with the same frequency observed when agreement is activated. The context and the moral cue represented by the thought experiment of the veil of ignorance seem not to be enough to activate any social norm of fairness.

The remainder of this paper is structured into four sections: the theoretical background (2); the experimental design (3); the results (4) and the final discussion (5).

2 Theoretical Background

This section will lay out the research question this paper aims to answer, and add the theoretical framework in which it is set. We will be referring to four strands of literature that set the stage for the experiment: First, the experimental literature that explores individuals' ideals of fairness; second, the broadly Rawlsian argument that connects rational agreement with principles of fairness; third, the justification of the particular version of the liberal egalitarian ideal that we pick

as a model for fair distribution; fourth, the place that the hypothesis of conformist preferences occupies in relation to Bicchieri's account of social norms.

(1) First of all, let's say that we set out to explore whether fairness ideals provide causally efficient reasons to act. Let us suppose that principles of distributive justice are rationally chosen. Does that mean that individuals will comply, *de facto*, with the content of those principles? What else is necessary for people to abide by principles they acknowledge as justified when soberly reflecting on a given distributive problem? Our experiment focuses on just one of the possible determinants of fair distributive choices, namely, the commitment acquired by reaching an impartial agreement.

We will take up first the theory explaining the link between rationality and the ideal of fairness in situations similar to the one we design in this experiment. Defining what is meant by fairness in production contexts is one of the primary objectives of the different theories of distributive justice (Cappelen et al. 2007). Among those, "[…] equal opportunity theories of distributive justice (Rawls 1971; Dworkin 1981; Roemer 1998), that combine an egalitarian commitment with a concern for individual responsibility" (Cappelen et al. 2007, 2) remain at the centre of the contemporary debate of normative ethical theories. We certainly follow this tradition. These theories solve the normative question about what people deem fair in distributive contexts, but they do not touch the positive question of how people actually behave when faced with distributive dilemmas, that is, situations in which they can secure a higher payoff for themselves by deviating from the fair distributive principle. Rawls himself was aware that a theory of justice "must generate its own support" (1971, 261). He identifies this problem as the question of stability. Sacconi's theory of conformist preferences introduced above suggests how we can make sense of the stability of a theory of justice. Sacconi and colleagues suggest that impartial agreements create a mutual commitment (Gilbert 2014) by which a new rational framework arises. Under this new framework, deciding accordingly to the distributive rule adopted by agreement becomes the preferred option for the rational individual.

(2) Turning now to the purely normative literature on fairness, the method proposed by Rawls holds a characteristic role as a method capable to redress the initial unjustified inequalities. For our purpose, John Rawls is a reference in two ways: first, we adopt the difference principle as a criterion that "secures for all a guaranteed minimum of the all-purpose means (including income and wealth) that individuals need to pursue their interests and to maintain their self-respect as free and equal persons" (Freeman 2019, Original Position, 1). We do not pretend to be testing Rawls's principles of justice, or modelling the exact reasoning leading to these principles. But we assume that, while the background norms usually activated in production contexts do not usually have egalitarian results, impartial

agreement should make people aware of unjustified inequalities and wish to level them out. Social norms about the distribution of benefits in production contexts generally justify differences, so we will take it for granted that a concern for initial inequality is more difficult to justify in this context. However, we contend that the reasons Rawls and other liberal egalitarian scholars offer for equality in basic rights and liberties should be persuasive also in a production context created in the lab. In particular, we will show that parties to an impartial agreement do understand and accept a rule that involves redress as a form of levelling of what is perceived as an illegitimate inequality in resources.

Secondly, Rawls is relevant because we adopt, and try to operationalize, his contractualist procedure. We hypothesize that reaching an agreement behind the veil of ignorance would lead individuals to choose a rule—the 'liberal egalitarian' rule—that tracks Rawls's ideals of fairness. This rule would be chosen as the normative solution to a distributive problem *and* would play a role in inducing positive compliance with it (Sacconi 2006). It is suggested that reaching an agreement behind the veil of ignorance means reasoning from an impartial perspective, and this induces a normative perspective in subjects or reconstructs what subjects *ought* to do. This is related to the Rawlsian idea that people possess a sense of justice (Rawls 1971, ch. 8). Insofar as 'sense of justice' can be interpreted as the capacity to act on principle, it is an element of the above-mentioned theory of conformist preferences.

(3) Throughout the article we will be referring to a 'liberal-egalitarian rule' as the fairest one. By this it is meant that, while other rules will represent different ideals of fairness, the liberal-egalitarian ideal is assumed to represent the most rational and reflective approach to justice. We acknowledge that this may be controversial, but our results prove that it may be not. The structure of this rule as well as its code name, requires some explanation. This rule takes up Konow's accountability principle (Konow 2000; 2001; 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). This principle is applied to production contexts, where 'producers' are entitled to the share of the output that derives from the 'discretionary variables' but not to the variation in output derived from variables they do not control. Liberal egalitarian theories—exemplified here by Rawls (1971) or Dworkin (1981b)—consider individuals as "morally equal persons deserving equal consideration and respect [...] 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" (Degli Antoni et al. 2016, 7). We take Konow's accountability principle to be approximately equivalent—only for the purpose of this experiment—to other

formulations of the basic liberal egalitarian ideal that defends the equality of all as moral persons—which entails that no differences can be justified in the distribution of basic endowments—while allowing differences based on individual responsibility and, in some cases, subject to a number of conditions and restrictions.

(4) Let us move now to the fourth set of theoretical considerations. They are related to the positive question of ex-post compliance. The question arises in our experiment because after choosing a distributive rule and performing the productive task, subjects are faced with a dictator game. In this game, their conduct would have been predicted to be self-interested—remember that there are no incentives or social bonds. However, Degli Antoni and colleagues found that the liberal egalitarian rule is not only preferred in the ex-ante agreement, but also followed ex-post. This result seems to support the theory of conformist preferences (Sacconi 2011; Grimalda/Sacconi 2005). However, the question remains whether the agreement itself has the double key role that Degli Antoni et al. suggest—as a deliberative procedure ex-ante *and* as a binding reason for parties ex-post, inducing high levels of unexpected compliance with the liberal egalitarian rule. The present experiment includes treatments where 'reasoning behind the veil of ignorance' is elicited in the absence of agreement. In this way we test the relative force of the actual agreement as mechanism for norm-elicitation in contrast to the (individually considered) idea of agreement behind a veil of ignorance.

The consideration underlying this exploration is the following. If the mere individual consideration of an agreement behind a veil of ignorance should make most people converge on the liberal egalitarian rule, the hypothesis of the conformist preferences—that argues that actual agreement is necessary—would be questioned. It would be apparent that there is a social norm of fairness that is conveniently activated in this sort of circumstances, and the procedure of actual agreement might be superfluous.

In our experiment we compare the treatment including agreement (fully based on the contractarian argument) with two treatments that substitute individual decisions ex-ante. In the individual treatments we study two situations that should (we hypothesize) influence compliance in different ways: in the first one—no common knowledge—the supposition is that the veil of ignorance may work as a moral cue. In the second one, the principle chosen ex-ante will be known by the subject's partner, and she will be able to know her partner's choice as well—and this is common knowledge from the beginning, when the experimenters read the instructions and check that they are properly understood. In this case the likely emergence of mutual normative and perhaps empirical expectations will add strength to the mere moral cue provided by the thought of an impartial agree-

ment. In our design, the code name of these treatments are 'Individual Choice' and 'Rule Other,' respectively.

3 Experimental Design and Hypotheses

In this study we compared three treatments, one called *Agreement*[3], which constitutes our baseline, a second one named *Individual Choice*,[4] and a third one called *Rule Other*. Common to all treatments is that subjects form anonymous pairs (they play in pairs, through personal computers, but they do not know who is their partner) and they perform a task for which one player will have ten minutes and the other player six minutes. Also, all the treatments consist of three stages: in the first stage, subjects had to choose how they would like to split a total income resulting from a task, selecting one rule from a menu of five rules (see below).

In the second stage, subjects performed the task. One member of the pair was assigned ten minutes to perform the task, while the other was assigned six minutes. Time limits assignment was random. The total income they have to split depends on performance doing the task. The task was coding words, by using a conversion table, and it was the same across all the treatments for all subjects. Before starting the task, participants saw on their monitors how much time (six or ten minutes) they would receive to complete it. At the end of the task, they were informed about each member's performance (total number of coded words and productivity measured as words *per minute*). Subjects were paid in experimental currency called token. At the end of the experiments tokens were converted in Euros at the exchange rate of 1 token = €0.15. They received one token for each word correctly coded.

In the third stage, subjects are asked to make a decision. They have three options: (i) confirm the distributive rule chosen in the first stage; (ii) change the previous choice by clicking on a different rule; or (iii) select a percentage corresponding to how much of the final amount they wanted to obtain themselves (the remaining amount would be left for their partner).

After subjects decide how to split the pair's total output, one member was randomly selected and her choice implemented. Each participant decided know-

3 Replication of 'Bargaining treatment' (Degli Antoni et al. 2016, 13–14).

4 Preliminary and less systematic evidence on individual commitment failure has been collected in a previous study (Marcon/Francés-Gómez/Faillo 2020) with distinct research questions, procedures and subjects.

ing that her choice had a 50% probability of being implemented as the real final payoff.

In stage 1 and 2 subjects are presented with a menu of five distributive rules that track different fairness ideals.

These are the five rules:

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 produces X in 10 minutes; subject Y in 6 minutes. Each one obtains [X+Y]/2.
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 produces X in 10 minutes; subject B produces Y in 6 minutes. The subject who is randomly selected (50% probability of being selected) obtains X+Y, 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 produces X in 10 minutes; subject B produces Y in 6 minutes. Subject A obtains X; subject B obtains Y.
4. Rule 4—*Time independent division*: each subject obtains what s/he has produced through her/his activity during the first 6 minutes; for the subject who has 10 minutes, the product of her last 4 minutes work is divided at 50% between the two subjects.
   Example: subject A produces X in the first 6 minutes and K in the last 4 minutes; subject B produces Y in 6 minutes. Subject A obtains X+(K/2) and subject B obtains Y+(K/2).
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.

These five distributive rules did not change across treatments and they try to follow the main moral intuitions (Haidt 2007; Sinnott-Armstrong et al. 2010)

about ideals of fairness. Rule 1 recalls 'pure egalitarianism (the total product is distributed equally)'. Rules 2 and 5 "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 productive per minute)" (Degli Antoni et al. 2016, 7). Rule 3 is an entitlement rule "based on contribution (each subject gets—is entitled to—what she/he has individually produced)". Rule 4 is proposed as the more obviously related to liberal egalitarianism.

Rule 5 may look counter-intuitive here. Even if it could be thought as another form of liberal egalitarian distributive solution, it would be more in line with Konow's accountability principle in that it not only cancels the arbitrary distribution of endowments but also rewards the ability or the hard-work of the more productive agents at the expense of the less productive. It implies that the distribution of resources is an illegitimate source of entitlements but that effort or natural luck (personal ability) are legitimate sources of differences.

About the purported meaning of the five rules, it must be acknowledged that the fit between common intuitions about fairness and the description of these rules is not perfect. It is important to focus on the distributive examples that follow the rules and recall that the subjects were able to practice and learn about the effect of each rule beforehand. And when they decided, they were informed about the exact distributive implication of the rule chosen; so in the end, subjects were able to choose the rule that fitted the material distribution that they wished. Thus, if both rules 3 and 5 may be conceived as rewarding effort, they are very different in their implications. Rule 5 cancels the initial inequality while Rule 3 does not. Also, while both rules 4 and 5 cancel the initial inequality, Rule 4 is more egalitarian and less demanding about accountability in its implications since the extra production of the person with ten minutes is divided equally, which gives this person the chance to decide how hard she want to work in what may seem an 'extra' time. Rule 5, on the contrary, would greatly penalize a diminishing productivity in the extra time: it may imply that the lucky person loses out if she is not as productive as possible during her whole ten minutes.

As said above, in all treatments, (i) participants are grouped in pairs; (ii) the endowment was earned (by the pair) through a task; (iii) each member of the pair was randomly assigned different time limits (ten or six minutes) to perform their task; (iv) in the third stage, participants played a dictator game, in which the software randomly assigned the dictator and responder roles. The underlying assumption behind condition (iii) is that the person with more available minutes

has an advantage, *and* a corresponding responsibility;[5] the foreseeable larger contribution of the person with ten minutes would not simply be the effect of chance, but the combined effect of chance *and* additional work on her part. This situation purports to represent the most common social distributive problems –those that are solved through liberal-egalitarian principles.

The three treatments differed in what follows:

Agreement

It was the baseline treatment and it reproduced the *Bargaining treatment* by Degli Antoni et al. (2016).[6] In the first phase of the game, subjects had 13 total rounds available in order to reach an agreement on which principle to choose. Subjects were informed that agreement was the condition to go ahead to the next stage of the game. The first 6 bargaining rounds were simultaneous, then there were 4 sequential offer and counter-offer rounds. The sequential process stopped when the rule proposed by one player was accepted by the other. For example, suppose that player A proposed the Rule 4: player B could accept it, so the agreement was reached and they could start the task phase. If player B did not, she could make a counter-offer by proposing a different rule. Player A could accept and they had an agreement, otherwise they had another sequential round as such. If subjects failed to agree, they had 3 additional simultaneous rounds. If agreement was not reached within the 13 rounds, the subjects were excluded from the experiment, but they had to remain in the lab—filling a questionnaire unrelated to the experiment, until the end of the session, at which time they would be paid the show-up fee.

Individual Choice and Rule Other

In both individual treatments, subjects did not have to reach an agreement. In the first stage, they were asked to individually choose, behind the veil of ignorance, one rule. The task (second stage) remained the same. The third stage, however, differed in the two treatments. In Individual Choice, participants were asked to confirm the previously chosen rule, choose a different rule or select a percentage.

5 Giving the nature of the task, extra work time can be conceived not as an additional effort but as an opportunity to earn more. This institution is confirmed by the data on productivity (words/minute) of the subjects with ten minutes, which does not change moving from the first six minutes to the second four.

6 "[…] the task and the division phases were preceded by a stage in which the members of the pairs, 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" (Degli Antoni et al. 2016, 14).

In Rule Other, subjects were informed about their partner's choice ex-ante.[7] Notice that this fact is known by subjects from the beginning through an instruction by the experimenter. That is, subjects in this treatment make a choice in the first stage knowing that their choice will be public knowledge afterwards. After they are informed about each other's choice ex-ante, the third stage proceeds as in the other treatments: subjects can decide whether to confirm their ex-ante choice, change it, or ask for a percentage of the product.

Given the theoretical background explained above, we propose two hypotheses, one related to the difference we expect to observe between agreement *vs*. individual treatments; and the other one derived from the difference between individual treatments.

> **H1**. The procedure via agreement behind the veil of ignorance (treatment 1) is more effective than procedures without agreement (treatments 2 and 3) in leading subjects to: (i) adopt a distributive rule in line with liberal egalitarianism (Rule 4 in our experimental design); or (ii) comply with Rule 4.

This hypothesis concords with Degli Antoni et al.'s hypotheses and evidence on the effectiveness of the agreement in inducing the convergence, both ex-ante and ex-post, on the liberal egalitarian principle of distributive justice.

> **H2**. Information about the rule chosen by the other person (and common knowledge about this fact) induces both a convergence ex-ante on the liberal egalitarian rule and ex-post compliance with it.

Knowing that the other participant will be informed about her ex-ante choice (and vice versa) could create a condition, proper of common knowledge, for the elicitation of mutual expectations of compliance: hence, the Rule Other treatment may provide justification for compliance. Within the experimental literature dealing with conformity, Bicchieri's (2006) theory locates the necessary conditions for compliance with a social norm in mutual expectations. In fact, Bicchieri identifies three reasons for an agent to decide to comply: to avoid a negative social sanction; to promote one's own desire to please others; to accept others' normative expectations as well founded. She says: "If I recognize your expectations as reasonable, I have reason to fulfil them. I may still be tempted to do something contrary to your expectations, but then I would have to justify (if only to myself) my choice by offer-

7 By using the expression ex-ante choice regarding to individual treatments, we want to establish a parallelism with the baseline. However, given the absence of the agreement, it must be remembered that ex-ante choice is to be understood as a choice made behind a veil of ignorance.

ing alternative good reasons and show how they trump your reasons." (Bicchieri 2006, 23–24)

What Bicchieri means by 'reasonable' is not so clear. However, it is useful to underline what are the three reasons, in her theory, for observing conformist behaviour. Two of them are incentives that can promote conformist behaviour, i.e. negative sanctions and social rewards, regardless of the value given to the norm by the individual who conforms. These two elements, however, do not explain those situations in which there is compliance within a reference group without either punishment or social rewards, in conditions of anonymity and non-observability. If the individual who conforms, given the conditions described above, assigns a moral content to the norm itself, then her preference to follow it will be unconditional (no matter what other members of the relevant group do or think she should do).

This underexplored issue in Bicchieri's theory is one of the objects of Sacconi and colleagues' theory of conformist preferences. This theory draws directly upon Bicchieri's theory of social norms, but it tries to complete the account of norms by exploring the process by which normative expectations that may became unconditional moral norms for the agents are elicited by descriptive elements, such as, for example, the particular past experiences that each of us collects. Categorization would seem to be more a descriptive process, which occurs by learning from the situations in which we find ourselves. So, the crucial point would be the following: the categorization process would allow people to infer that, under specific circumstances, they might expect a certain kind of behaviour (observed in the past). But why should that behaviour be the one to be followed? Why, from pure descriptive evidence categorized in the past, should people infer that the expected behaviour is the one that they *ought* to conform to?

Upon listening to the instructions, participants in the Rule Other treatment became aware of the initial unjustified inequality in the assignment of endowment (as were subjects in the other two treatments), *and* they understood also that the ex-ante choice of each player would be communicated to the other player before stage three. By these external cues, subjects might activate a script for which, not knowing who would have a larger endowment (and therefore an opportunity to produce a larger contribution), Rule 4 turned out to be the focal ex-ante distributive rule. Given empirical expectations and conditional preferences, if the context was properly understood, participant should be sensitive to Rule 4 even without an explicit agreement (H2).

If H2 is supported, then the impartial agreement behind the veil of ignorance loses its central role as a determinant for compliance with the liberal egalitarian rule. It could mean that both the effect of individual reasoning behind the veil of ignorance and the common knowledge of the others' choice selected behind the

veil make Rule 4 focal behind the veil of ignorance. In addition, these elements could also determine the actual compliance with this rule, because the information given to the subjects and the situation in which they find themselves interacting would legitimize those 'reasonable expectations' which Bicchieri refers to as the third reason for conformity. If this were observed, then the impartial agreement would not be so fundamental in making Rule 4 salient and ex-post complied with. In other words, it would mean that the process of eliciting fairness norms and the one of categorization, as held by Bicchieri, could be sufficient features to explain conformity with a rule, such as the liberal egalitarian, that prescribes what it should be done ex-post.

Let's add, finally that in all treatments, instructions were read aloud by one of the experimenters, a set of control questions were proposed to make sure that participants understood the instructions and they were paid a fixed show-up fee of €3. The experiment was programmed by using zTree (Fischbacher 2007) and conducted at the Cognitive and Experimental Economics Laboratory (CEEL) at the University of Trento. A total of 176 students participated in the experiment between March 2018 and March 2019. Two sessions of 18 subjects and one with 20 participants were run for the Agreement treatment, three sessions of 20 subjects each for the Individual Choice treatment, two sessions of 20 subjects and one with 18 participants for the Rule Other Information treatment.

4 Results

Task

Looking at subjects' performance in the task, we do not observe significant differences in productivity, measured as words per minute, within the pairs. The median difference between the productivity of the two members of the pair is very close to zero in all the treatment (Figure 1). This supports the idea, at the basis of the original design by Degli Antoni et al. (2016), that different abilities have only a marginal role in explaining differences in performance in the task, and the main source of difference are the different time limits assigned to the two categories of subjects. It is important to stress this point, since in this design, the veil of ignorance is applied only to the amount of time subjects will have for the task. If different abilities had a great impact on production, they might impact the choice of rule ex post[8]—Rule 5 might have been more frequently selected by more produc-

8 Any differences in ability cannot be known ex-ante, for the experiment is anonymous.

tive individuals. This is not the case, and this confirms that in this design the veil of ignorance hides the only element that is relevantly unequal in the situation.

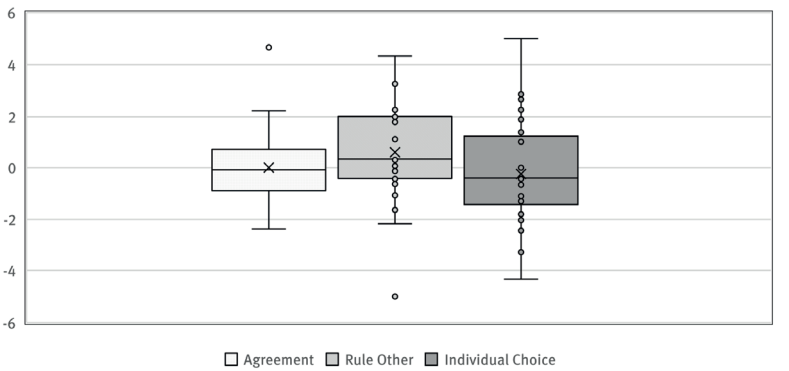

Fig. 1: Difference in productivity (words per minute) between the subject with six minutes and the subject with ten minute

Figure 2 reports the choices made by participants ex-ante (before knowing the time they would be assigned).

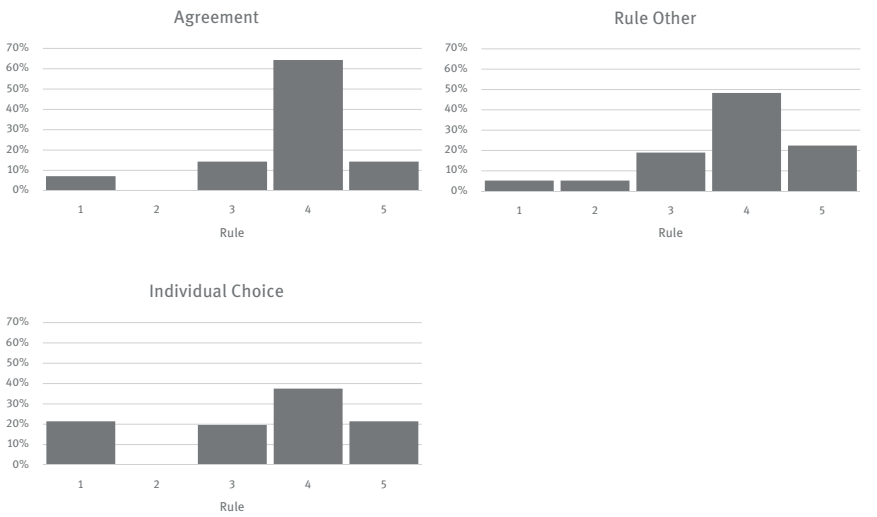

Fig. 2: Choice ex-ante

In all the treatments Rule 4 is the most frequently chosen rule. A proportion test reveals that Rule 4 is chosen more frequently in the Agreement than both in the Individual Choice ($z = 2.97$, $p = 0.003$) and in the Rule Other treatments, even if in the latter case the difference is significant only at the 10% ($z = 1.72$, $p = 0.08$).[9] There is not a significant difference in the frequencies of Rule 4 choices in the two individual treatments. The number of subjects choosing other rules is too small to perform a detailed analysis. Notice however that the frequency of choice of Rule 1 is significantly higher in the Individual treatment than both in Rule Other and in the Agreement treatment (Individual Choice vs. Rule Other: $z = -2.61$, $p = 0.01$; Individual Choice vs. Agreement $z = 2.21$, $p = 0.02$).

We can then put forward the following results.

Result 1.
In the ex-ante choice of Agreement treatment, the choice of Rule 4 is more frequent than in the Individual choice treatments.

Result 2.
In the ex-ante choice, choice of Rule 4 in the Rule Other treatment is more frequent than in Individual Choice treatment but this difference is not statistically significant.

These two results support H1 with regard to the ex-ante choice. As for H2, using the Individual Choice treatment as a benchmark, the convergence of Rule Other results towards the results obtained in the Agreement treatment is only partial.

Choice ex-post

After the task, subjects can choose either one of the five rules or a free percentage. Subjects opting for the percentage were six in the Agreement treatment, four in Individual Choice and two in the Rule Other treatment. We decided to consider the two choice of a percentage of 50% as equivalent to the choice of Rule 1, and two choices of a percentage of 100% as equivalent to the choice of Rule 2.[10]

As for the choice ex-post (*figure 3*), Rule 4 is the most chosen in the Agreement and in the Rule Other treatments. Rule 3 prevails in the ex-post decision of Individual choice treatment. Rule 4 is chosen more frequently in the Agreement than in the Individual Choice treatment ($z = 2.41$, $p = 0.01$), Rule 3 is chosen more frequently in the Individual Choice treatment than in the Agreement treatment ($z = 2.14$, $p = 0.031$), and Rule 2 is chosen more frequently in the Individual Choice

9 The first result is confirmed by a probit estimation in which we control for subject's age, gender and experience with the experiments (Table 1A in the *appendix*).

10 Two subjects choose a percentage of 1% and they have been removed from the sample.

treatment than both in the Baseline (z = 2.15, p = 0.03) and in the Rule Other treatments (z = 1.69, p = 0.09), even if the latter difference is significant at only the 10%.[11]

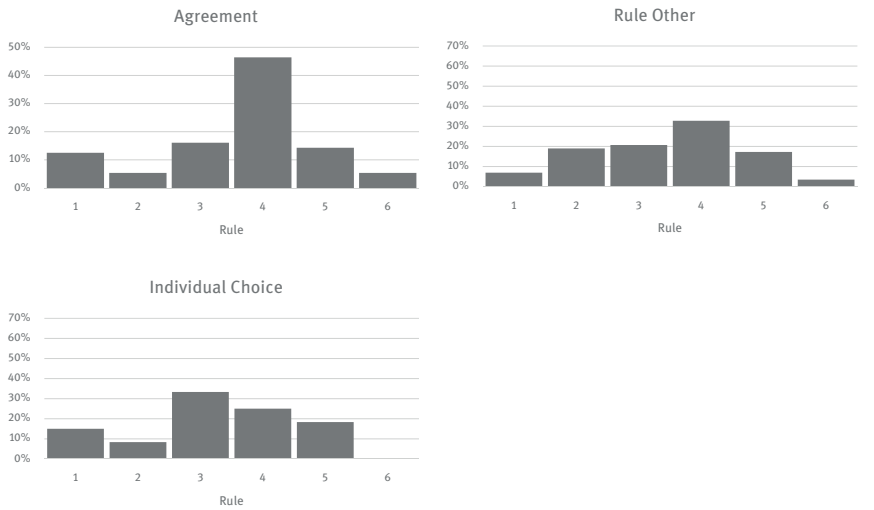

Fig. 3: Choice ex-post

Independently of the rule chosen, the frequency of compliance (ex-post choice confirming the choice ex-ante) is significantly higher in the Agreement than in the Individual Choice treatment (z = 2.48, p = 0.013) (*figure 4*).

11 This evidence is confirmed by probit estimations (*table 2A* in the Appendix, columns 1 and 2). *Table 2A* show also that subjects with ten minutes are less likely to choose Rule 4 ex-post than subjects with six minutes. We checked for difference between treatments by using interactions between the dummy variable Ten and treatment dummies Agreement and Rule Other. None of the coefficients was different from zero. The third column of *table 2A* reports the results of a probit estimation limited to the choice made by the subjects in the Rule Other treatment. We observe that knowing that the other subject in the pair chose Rule 4 ex-ante (Rule other 4) has no effect on the choice of Rule 4 ex-post.

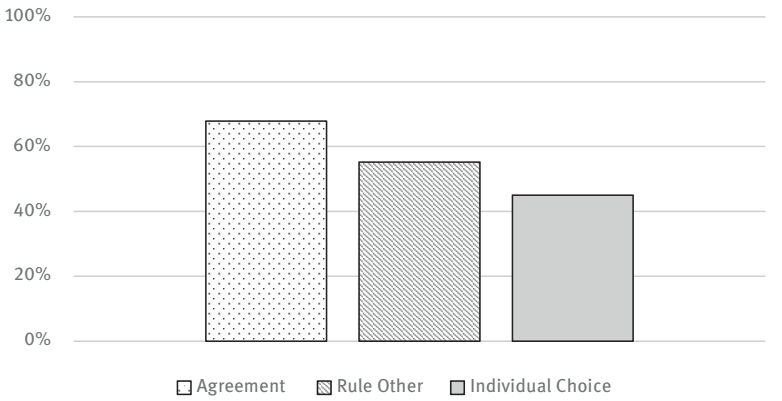

Fig. 4: Compliance across treatments

Looking at the frequency of compliance across rules and treatments (*table 3*), we also see that compliance with Rule 4 is significantly higher in the Agreement than both in the Individual (proportion test: z = 3.66, p= 0.002) and in the Rule Other treatments (z = 2.10, p = 0.03).[12]

Tab. 3: Proportion of compliant subjects across treatments and rules

	Treatment		
Rule	Agreement	Individual Choice	Rule Other
1	4/4	1/3	5/13
2	0/0	2/3	0/0
3	5/8	8/11	5/11
4	25/36	15/28	11/22
5	4/8	6/13	6/14
	38/56	32/58	27/60

We summarize the evidence on ex-post choice with an additional set of results.

12 This result is confirmed by the probit estimation of *table 4A* in the *appendix*, in which we also observe that the likelihood of compliance with Rule 4 for subjects with ten minutes is lower than that for subjects with six minutes.

Result 3.
The choice of Rule 4 ex-post is more frequent in the Agreement treatment than in the Individual Choice treatment.

Result 4.
The frequency of Rule 4 ex-post choices in the Rule Other treatment is not statistically smaller than that observed in the Agreement treatment and it is not statistically greater than in the Individual Choice treatment.

Result 5.
Compliance with Rule 4 in the Agreement treatment is higher than that observed in both Individual Choice and Rule Other treatments.

These results support hypothesis H1. H2 is only partially supported.

5 Discussion and Conclusions

In this final section we would like to explore the lessons from this experiment. This is an experiment designed to test the robustness of the conclusions of a previous experiment by Degli Antoni et al. That experiment, along with Sacconi et al. (2011) and Faillo et al. (2015) was intended to support the theory of conformist preferences. In these three antecedents, the experiment involved an agreement behind a veil of ignorance, and then a production and distribution phase; and in all three the claim was that impartial agreement leads to distribution rules consistent with liberal egalitarianism and that agreement itself transforms the parties' interpretation of the distributive choice; parties see it as a normative situation in which agents tend to abide by the norm previously selected by agreement. We designed two individual treatments in order to test whether the effect previously observed was due to agreement itself, or to some contextual characteristic of the interaction that may be influencing the subjects.

In these treatments we used the agreement behind a veil of ignorance as a moral cue with different levels of strength (with or without common knowledge). We would like to argue that lessons can be drawn from this experiment both for moral philosophy and for a theory of social norms. We collected data about the conditions under which ordinary individuals reach a normative conclusion that is in line with liberal egalitarian theories of distributive justice. Besides, we found that the context and cues provided in the experiment were not enough by themselves, to recall the social norm of fairness that seems to be easily recognized through bargaining and agreement.

This experiment compares the baseline treatment—involving actual agreement through a 'bargaining' over rules—with two individual treatments. Given that the situation was the same, participants very similar, and that an impartial reflection over the distributive problem was induced in all three treatments, we contend that the presence of a 'social norm of fairness' should have been detected across treatments. From our theoretical framework, we suggested that the rule closer to the ideals of liberal egalitarianism (Rule 4) might be the focal point. Both normative beliefs and actual behaviour should have been aligned, perhaps with slight variations, with this rule in all treatments. *However, this was not observed*.

Our experimental results support the role of agreement in distributive contexts, based on the theory of conformist preferences put forward by Sacconi and colleagues. Let's elaborate this conclusion and its eventual implications.

The support found for hypotheses 1 and 2 has underlined how an explicit prior agreement between the parties is a fundamental condition for making the liberal egalitarian rule (Rule 4) salient behind the veil of ignorance and ensuring ex-post compliance. The veil of ignorance alone—adopted as a thought experiment—seems not to be strong enough to create conditions for (most) individuals to adhere to a liberal egalitarian rule. This result aligns with other data on classic bargaining games, in conditions of anonymity, non-iteration and absence of external negative sanction and/or positive reward (Hoffman et al. 1996; Rodriguez-Lara/Moreno-Garrido 2012). However, our result is still surprising, because the device of the veil of ignorance was purposively introduced in our design to make subjects fully aware that the situation of production was going to include an unjustified inequality and that they would not know in advance whether they would be on the advantaged or disadvantaged end.

Our conjecture about introducing the veil of ignorance applied to individual reasoning wanted to test, indirectly, to what extent the veil of ignorance had an effect as a moral device. The effect of the veil of ignorance, as envisioned by Rawls, is to put subjects in the condition to reason as if they were the disadvantaged party. This should have led subjects towards Rule 4 (at least ex-ante). However, this type of reasoning proved to be ineffective: it did not constitute a motivational source strong enough to single out that distributive rule as the one to apply, not even as a purely normative belief in *foro interno*. We use here the distinction introduced by Hobbes about the binding force of the Laws of Nature in the state of nature as opposed to its binding force in the civil state. Even Hobbes (1991, ch. 14) accepts that, while one is not bound to follow the laws of nature in her external behaviour before there is a common power capable to enforce them, they are still authoritative as counsels of reason. We expected that our subjects would detect the nature of the distributive problem at hand, and they would correctly infer that, from the point of view of the least advantaged—the person who got six minutes to work—,

a distributive rule including re-dress was the rational rule to choose, even if afterwards they would to choose selfishly. This expectation is based on accepting the tenets of Rawlsian moral philosophy, combined with a disingenuous view on social norms. Subjects aware of the unjustified differences should acknowledge that the fairest norm in our case was Rule 4; but given the absence of social incentives, they would disregard the norm in practice.

Notice that merit is hardly elicited in our experimental design, since the advantage position (having more time to perform the activity) is not the result of a tournament/competition/skill test. This is important, for it means that there was no way in which the random assignment of time could be seen as 'deserved.' This should make one expect that people would see their being lucky as a reason to be generous, so to say, with their partners. However, in the Individual Choice treatment, the probability of being more or less favoured seemed to have no effect whatsoever, as if this feature of the design was a fact wholly irrelevant for the choice of distributive rule. Subjects disregarded the effect of luck and most accepted the distribution rule that assigns each what she has produced in the available time. Actually, we observe that the likelihood of compliance with Rule 4 for subjects with ten minutes is lower than that for subjects with six minutes: the more advantaged subjects (let's say the richest ones) do not care much about helping the most disadvantaged fellows. This effect has been observed before (Cabrales et al. 2012).

Results concerning H1 a and b strengthen the 'contractarian argument' that supports Sacconi and colleagues' theory of conformist preference according to which the deliberative process behind the veil of ignorance ensures to reach an agreement on Rule 4. The conditions of impartiality and impersonality, guaranteed by the procedure itself, make Rule 4 uniquely salient for both players, and it becomes the solution of the bargaining process and the agreement's content. Once an explicit prior agreement has been achieved, conformism—ex-post behaviour aligned with the agreed rule—is generally observed. Again, this is surprising from the perspective of standard economic rationality. The contractarian argument assumes conformism because it assumes that individuals do possess a sense of justice: a conditional disposition to follow fair norms. This result points to the conclusion that the script for liberal egalitarian justice, which is so obviously a distributive ideal under the condition of agreement, is not elicited by clear moral cues and common knowledge about other's choice.

As Bicchieri argues, compliance with a norm depends on the salience stemming from it and on the expectations that emerge among those who must decide whether or not to conform. By comparing the two individual treatments, in Rule Other, Rule 4 is chosen more and the degree of compliance with what was decided behind the veil is greater than the pure individual decision-action. The fact

of knowing that the choice made by the other player behind the veil of ignorance could have affected subjects by creating mutual expectations: "According to this approach, agreement on the norm is not a necessary condition for compliance, and it is replaced by a general idea of awareness of the existence of the norm (its salience) in the community." (Faillo et al. 2015, 230)

The treatment 'Rule Other' was designed to capture this feature of social norms. Even in the absence of explicit agreement, the conditions of this treatment should have induced a focus on Rule 4—recall that in this treatment the idea of a veil of ignorance is not simply a moral cue, because common knowledge introduces the other in each subject's reasoning, therefore enhancing the focus on the common acceptability of rules. But again, this treatment did not increase ex-ante choice or ex-post compliance with Rule 4 to a significant level.[13] It seems that in contexts where neither pure salience nor external cues are sufficient conditions for conforming to a liberal egalitarian distributive rule, the procedure by which the rule is chosen seems to be the key. The explicit prior agreement might guarantee collective deliberation, acceptance, and willingness to conform accordingly.

In addition, Rule 4, as it is jointly chosen, creates a motivational source for compliance, even if there are no rewards for pro-social or social sanctions for self-interested behaviour. Therefore, data seem to support a theory of conformity that has highlighted the role of impartial agreements in help us generate fair distributive norms. Compliance with a fairness norm would depend on the emergence of expectations, normative and empirical, but also crucially on the fact that following a specific norm is the result of a process that involves the players: when the norms of fairness are collectively constructed, via agreement, they could constitute not only normative reasons—what should be done—but also motivational factors that ensure compliance (Grimalda/Sacconi 2005; Sacconi/Grimalda 2007). These findings would pinpoint that, in production contexts, the possibility for individuals to express their consent on one norm rather than another would guarantee the emergence of a liberal egalitarian view and a corresponding conformist behaviour.

One possible criticism to this conclusion is that it seems to imply that social life would require constant explicit contracts in order to secure distributive fairness. This is not a consequence we would like to commit to: we claim to have established that in order to secure voluntary compliance with a well-founded norm of fairness, explicit impartial agreement is more effective than

13 In the Rule Other treatment, the correlation between subject's ex-post choice and the rule chosen by the other before the task is very weak (Spearman's rho = 0.15, p = 0.24).

other mechanisms. This is a conclusion about finding fair norms and inducing voluntary fair behaviour. Of course, social norms are seldom entirely voluntary. Once the fair norm is chosen by agreement, coercive and socialising mechanisms are implemented—and legitimately so. What this experiment show is that in absence of social coercion we must rely on what we called the 'contractarian argument.' The results of individual treatments show that the context created in the lab did not include a social norm of fairness. However, the agreement treatment succeeded in finding one rule that is very plausibly fair, and motivates subjects to comply with it. This is remarkable.

When we see social norms of fairness in action they are backed by social sanctions. This experiment is not about social norms in action, but about the discovery, or institution, of the social norm (the distributive rule) that is appropriate for a specific productive situation described as neutrally as possible. In this very particular case, principled action ex-post seems to require a personal engagement that is not easily attained by mere individual reflection. This does not pretend to be a conclusion about social life, but about the way distributive norms must be established.

By confirming Degli Antoni et al. findings, we contribute to moral philosophy. The relationship between an impartial agreement and a rule coherent with liberal egalitarianism, which is a key feature of most liberal theories of justice, is experimentally established. In addition, the peculiar nature of the moral norm is empirically observed, since the level of compliance is high only when a rational commitment has been elicited through an impartial agreement, while the analogous individual elicitation of the presumed social norm induces a much lower level of awareness and compliant behaviour.

Let us comment, finally, that this study is limited in its scope. First of all, it would be necessary to explore in depth the reasons why agreement may be so motivationally effective in absence of external incentives. One obvious possibility points to the idea that agreement, but not other forms of normative practical reasoning, generates a structure of 'joint commitment and action' (Gilbert 2014). However, our data prevent us to get to definitive conclusions about this. Further research should explore also to what extent the results can be interpreted as a contribution to the problem of stability in Rawlsian theory. Our subjects showed a stable disposition to abide by a rule that was selected from a purely normative perspective—under a veil of ignorance—and this is promising. But our sequential game may be too simple: the whole experiment takes place in less than one hour. Subjects may be influenced by the shadow of the agreement—the impact of the agreement just reached. Proving the stability of principles in general would require showing that the rule chosen is interiorized as the 'rational thing to do'

not only in an interaction with partners to the agreement, but also in other future interactions of the same form.

Despite the limited scope of this contribution, the clear negative answer to our research question does contribute to the study of the emergence of social norms of fairness, a topic understudied by conventionalist views on social norms. From this and other related experimental results, it is becoming clear that impartial agreements do uniquely lead to liberal egalitarian distributive rules, and this is a contribution to moral philosophy that may have ample practical implications.

Acknowledgment: Funding for the research reported in this paper was provided by the Spanish Ministry of Economy through Research Grants BES-20 FFI2017-87953-R.

References

Bicchieri, C. (2006), *The Grammar of Society: The Nature and Dynamics of Social Norms*, New York

— (2008), The Fragility of Fairness: An Experimental Investigation on the Conditional Status of Pro-Social Norms, in: *Philosophical Issues* 18, 229–248

— (2016), *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*, Oxford

Cabrales, A./R. Nagel/R. Rodríguez/J. V. Mora (2012), It is Hobbes, Not Rousseau: An Experiment on Voting and Redistribution, in: *Exp Econ* 15, 278–308

Cappelen, A. W./A. D. Hole/E. O. Sørensen/B. Tungodden (2007), The Pluralism of Fairness Ideals: An Experimental Approach, in: *The American Economic Review* 97, 818–827

Charness, G./U. Gneezy/A. Henderson (2018), Experimental Methods: Measuring Effort in Economics Experiments, in: *Journal of Economic Behavior and Organization* 149, 74–87

Cialdini, R. B./C. A. Kallgren/R. R. Reno (1991), A Focus Theory of Normative Conduct: A Theoretical Refinement and Re-evaluation of the Role of Norms in Human Behavior, in: *Advances in Experimental Social Psychology* 24, 201–234

Degli Antoni, G./M. Faillo/P. Francés-Gómez/L. Sacconi (2016), Distributive Justice with Production and the Social Contract. An Experimental Study, in: *Econom Etica* 60, 1–51

Erkal, N./L. Gangadharan/N. Nikiforakis (2011), Relative Earnings and Giving in a Real-Effort Experiment, in: *The American Economic Review* 101, 3330–3334

Faillo, M./L. Sacconi (2007), Norm Compliance: The Contribution of Behavioral Economics Models, *Discussion Paper* 4, University of Trento, 1–36

—/S. Ottone/L. Sacconi (2015), The Social Contract in the Laboratory. An Experimental Analysis of Self-enforcing Impartial Agreements, in: *Public Choice* 163, 225–246

—/M. Rizzolli/S. Tontrup (2019), Thou Shalt Not Steal: Taking Aversion with Legal Property Claims, in: *Journal of Economic Psychology* 71, 88–101

Fehr, E./I. Schurtenberger (2018), Normative Foundations of Human Cooperation, in: *Nat Hum Behav.* 2, 458–468

Fischbacher, U. (2007), z-Tree: Zurich Toolbox for Ready-Made Economic Experiments, in: *Experimental Economics* 10, 171–178

Freeman, S. (2019), Original Position, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL: https://plato.stanford.edu/entries/original-position/

Frohlich, N./J. A. Oppenheimer (1992), *Choosing Justice: An Experimental Approach to Ethical Theory*, Berkeley

Gilbert, M. (2014), *Joint Commitment: How We Make the Social World*, Oxford

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

Haidt, J. (2007), The New Synthesis in Moral Psychology, *Science* 316, 998–1002

Hobbes, T. (1991), *Leviathan*, Cambridge

Hoffman, E./K. McCabe/V. Smith (1996), Social Distance and Other-Regarding Behaviour in Dictator Games, in: *The American Economic Review* 86, 653–660

Huanga, K./J. D. Greene/M. Bazerman (2019), Veil-of-Ignorance Reasoning Favours the Greater Good, in: *Proceeding of the National Academy of Sciences of the United States of America* 116, 23989–23995

Kok-Chor, T. (2008), A Defense of Luck Egalitarianism, in: *The Journal of Philosophy* 105, 665–690

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

— (2001), Fair and Square: The Four Sides of Distributive Justice, in: *Journal of Economic Behavior and Organization* 46, 137–164

— (2003), Which Is the Fairest One of All? A Positive Analysis of Justice Theories, *Journal of Economic Literature, American Economic Association* 41, 1188–1239

— (2005), Blind Spots: The Effects of Information and Stakes on Fairness Bias and Dispersion, in: *Social Justice Research* 18, 349–390

Marcon, L./P. Francés-Gómez/M. Faillo (2020), Does Impartial Reasoning Matter in Economic Decisions? An Experimental Result about Distributive (Un)fairness in a Production Context, in: Theoria 35, 217–233

Oleson, P. E. (2001), *An Experimental Examination of Alternative Theories of Distributive Justice and Economic fairness*, UMI Microform 3016508, Bell & Howell Information and Learning Company

Rawls, J. (1999[1971]), *A Theory of Justice (Revised edition)*, Cambridge/MA

Rodriguez-Lara, I./L. Moreno-Garrido (2012), Self-interest and Fairness: Self-serving Choices of Justice Principles, in: *Experimental Economics* 15, 158–175

Sacconi, L. (2011), A Rawlsian View of CSR and the Game Theory of Its Implementation (Part III: Conformism and Equilibrium Selection), in: Sacconi, L./G. Degli Antoni (eds.), *Social Capital, Corporate Social Responsibility, Economic Behavior and Performance*, Basingstoke, **??–??**

—/G. Grimalda (2007), Ideals, Conformism and Reciprocity: A Model of Individual Choice with Conformist Motivations, and an Application to the Not-for-Profit Case, in: Bruni, L./P. L. Porta (eds.), *Handbook on the Economics of Happiness*, 532–569

—/L. Faillo (2008), *Conformity, Reciprocity and the Sense of Justice How Social Contract-based Preferences and Beliefs Explain Norm Compliance: The Experimental Evidence*, Discussion paper 14, University of Trento

—/— (2010), Conformity, Reciprocity and the Sense of Justice. How Social Contract-based Preferences and Beliefs Explain Norm Compliance: The Experimental Evidence, in: *Constitutional Political Economy* 21, 171–201

—/—/S. Ottone (2011), Contractarian Compliance and the 'Sense of Justice': A Behavioral Conformity Model and Its Experimental Support, in: *Analyse & Kritik* 33, 273–310

Sinnott-Armstrong, W./L. Young/F. Cushman (2010), Moral Intuitions, in: Doris, J. M. (ed.), *The Moral Psychology Handbook*, Oxford, 246–272

Voigt, S. (2015), Veilonomics: On the Use and Utility of Veils in Constitutional Political Economy, in: Imbeau, L. M./S. Jacob (eds.), *Behind a Veil of Ignorance? Power and Uncertainty in Constitutional Design*, 9–33

Young, H. P. (2007), Social Norms, *Discussion Paper Series* 307, Department of Economics, University of Oxford

Appendix: Econometric Analysis

Tab. 1A: Determinants of ex-ante choice of Rule 4

Dep. Variable: Rule 4 ex ante	(1) Probit
Agreement	0.742*** (0.242)
Rule Other	0.311 (0.236)
Age	-0.0277 (0.0394)
Gender	0.177 (0.196)
Experiments	-0.00751 (0.0129)
Constant	0.234 (0.856)
Agreement – Rule Other	0.43 (0.240)
Observations	174
Pseudo R^2	0.04
Log Likelihood	-115.011

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
The dependent variable is equal to 1 if the subject chooses Rule 4 before the task and 0 otherwise.
Agreement is equal to 1 if treatment is the Agreement treatment and 0 otherwise.
Rule Other is equal to 1 if treatment is the Rule Other treatment and 0 otherwise.
Age is the age of the subjects, Gender is equal to one if the subject is female and 0 otherwise, Experiment is the number of experiment the subject as taken part in the past.

Tab. 2A: Determinants of the ex-post choice of Rule 4

	(1)	(2)	(3)
Dependent variable:	Probit	Probit	Probit
Rule 4 ex-post	Full sample	Full sample	Rule Other only
Agreement	0.671***	0.713***	
	(0.252)	(0.258)	
Rule Other	0.324	0.352	
	(0.255)	(0.261)	
Ten		-0.577***	-0.536***
		(0.207)	(0.205)
Age	0.0141	0.0157	0.0195
	(0.041)	(0.042)	(0.0410)
Gender	0.602***	0.586***	(0.204)
	(0.206)	(0.210)	-0.0178
Experiment	-0.0190	-0.0179	(0.0145)
	(0.014)	(0.014)	
Rule Other 4			-0.0792
			(0.291)
Constant	-1.175	-0.969	-0.655
	(0.910)	(0.934)	(0.896)
Agreement – Rule Other	0.360	0.339	
	(0.250)	(0.470)	
Observations	174	174	58
Pseudo R	0.07	0.11	0.08
Log Likelihood	-103.89	-99.960	-33.817

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

The dependent variable is equal to 1 the subject chooses Rule 4 after the task and 0 otherwise.

Agreement is equal to 1 if treatment is the Agreement treatment and 0 otherwise.

Rule Other is equal to 1 if treatment is the Rule Other treatment and 0 otherwise.

Ten is equal to 1 the subject has 10 minutes and zero otherwise.

Rule Other 4 is equal to 1 if the other subject in the pair chose Rule 4 ex-ante.

Age is the age of the subjects, Gender is equal to one if the subject is female and 0 otherwise, Experiment is the number of experiment the subject as taken part in the past.

Tab. 4A: Determinants of compliance with Rule 4

Dep. Variable: Rule 4 ex ante	(1) Probit
Agreement	0.742*** (0.242)
Rule Other	0.311 (0.236)
Age	-0.0277 (0.0394)
Gender	0.177 (0.196)
Experiments	-0.00751 (0.0129)
Constant	0.234 (0.856)
Agreement – Rule Other	0.43 (0.240)
Observations	174
Pseudo R^2	0.04
Log Likelihood	-115.011

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
The dependent variable is equal to 1 if the subject chooses Rule 4 before the task and 0 otherwise.
Agreement is equal to 1 if treatment is the Agreement treatment and 0 otherwise.
Rule Other is equal to 1 if treatment is the Rule Other treatment and 0 otherwise.
Age is the age of the subjects, Gender is equal to one if the subject is female and 0 otherwise, Experiment is the number of experiment the subject as taken part in the past.