Journal of Institutional Economics (2022), **18**, 901–918
doi:10.1017/S1744137422000029

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Liberal egalitarian justice in the distribution of a common output. Experimental evidence and implications for effective institution design

Giacomo Degli Antoni[1,2*], Marco Faillo[3], Pedro Francés-Gómez[4] and Lorenzo Sacconi[2,5]

[1]Department of Law, Politics and International Studies, University of Parma, via Università 12, 43121 Parma, Italy, [2]EconomEtica, c/o University of Milano-Bicocca, via Bicocca degli Arcimboldi 8, 20126 Milano, Italy, [3]Department of Economics and Management, University of Trento, via Inama 5, 38122 Trento, Italy, [4]Department of Philosophy I, Campus de la Cartuja, 18071 Granada, Spain and [5]Department of Italian and Supranational Public Law, University of Milan, via Festa del Perdono 7, 20122 Milano, Italy
*Corresponding author. Email: giacomo.degliantoni@unipr.it

(Received 22 July 2021; revised 20 January 2022; accepted 20 January 2022; first published online 22 February 2022)

Abstract

We present an experiment that sets up a context of production of a common output obtained by using production means that are randomly and unequally distributed. Before the production phase, subjects must choose a distributive principle for the output division, under ignorance of the allocation of the production means. Subsequently, they make a distributive choice fully aware of their luck and performance. The aim of the experiment is to test, first, whether ordinary subjects in an impartial situation are capable of converging on a fair principle of distribution – able of redressing the arbitrariness of the initial production means allocation; and second, whether these same ordinary subjects are capable of actually following that principle in a real distributive choice that excludes coercion, reputation effects and other forms of social pressures. The main finding is that a distributive rule that redresses initial inequalities is both accepted *ex-ante* and actually applied *ex-post* by most individuals. Our conclusion is relevant for the issue of realism of normative theories of justice and the possibility of institution design aimed at implementing distributive justice principles and policies.

Key words: game theory; inequality; institution design; justice theory; norms

JEL Classification: C72; C91; D02; D63; H80

1. Introduction: aim, overview and relevance for institution design

Productive institutions such as firms should be efficient but also, ideally, fair.

Much of new-institutional economics however purports to design institutions from the sole viewpoint of efficiency, neglecting distributive concerns – some examples in the theory of the firm, are: Hansmann (1988), Williamson (1975), Grossman and Hart (1986), Hart and Moore (1990). The general point is made by Kaplow and Shavell (2002) and, in the social contract tradition, by Buchanan (1975). This may be because working in the perspective of efficiency (Pareto optimality), institution design can be shown to be incentive compatible, given that Pareto-optimality appeals to mutual interest. Adopting the perspective of justice seems too demanding for a realistic effort of institution design because real-world motivations are too removed from what would be necessary for a spontaneous adherence to principles of justice. It is no coincidence that 'opportunism' is the assumed model of behavior (Williamson, 1975, 1985).

This is in contrast with much theoretical work on justice that is aimed at devising a hypothetical choice model to test the justice of fundamental social institutions. Several traditions of ethical and

political thought have proposed hypothetical original positions from which individuals allegedly choose or agree on principles of social justice. This *ex-ante* hypothetical choice is conceived as a context of ignorance or uncertainty so that the parties cannot be defending particular interests. While this method is hypothetical, it suggests realism insofar as the choice on principles is derived from prudentially rational motives. Once agreement is reached, and the veil of ignorance lifted up, it is claimed that just institutions, framed according to agreed principles, will be complied with and remain stable thanks to the development of a behavioral attitude called the 'sense of justice' (Rawls, 1971). This is elicited by the very agreement on principles, plus mutual beliefs of reciprocal compliance.

It must be noted that even in the theory of justice tradition it is not uncontroversial whether distributive justice may enter the domain of economic institutions. Some authors have affirmed that the extension to productive institutions is intrinsic to Rawlsian theories of justice and should be developed (Fia and Sacconi, 2019; Sacconi, 2006, 2011a, 2011b); others have taken the opposite view (Mansell, 2013; Singer, 2015). Our research question may settle these controversies, at least in so far as they concern the question whether the request of designing productive institutions (like firms) according to principles of distributive justice could be too demanding.

In this paper we use an experiment to answer three related questions:

1) Does the introduction of an *ex-ante* veil of ignorance process – where agreement over a non-binding rule for distribution has to be reached – identify a dominant rule?
2) Do people comply with the agreed-upon distributive rule when they actually decide how to allocate a shared output?
3) Is the actual distribution when there is such an *ex-ante* procedure different from the one obtained in its absence?

Our results suggest that, first, subjects predominantly select a rule of distribution that can arguably be identified with liberal-egalitarian justice; second, people do generally comply, but more so when (a) the *ex-ante* process is an informal deliberation carried out in natural language rather that formal bargaining, and (b) when they expect their co-player to comply; and third, the distribution in treatments with 'veil of ignorance' is far more consistent with the ideals of fairness than it is in the control treatment.

These results are relevant for institutional design in at least two ways.

First, they show that by introducing an *ex-ante* collective choice procedure offering agents the opportunity of reaching an impartial agreement on distributive principles the pattern of behavior changes substantially. We understand such a behavior as entirely consistent with the 'equilibrium view' of institutions: endogenous regularities of behavior (conform to normative principles) emerge from strategic interaction and are reflected in players' mental representation by means of reciprocal beliefs (see Aoki, 2001, 2007; Hindrinks and Guala, 2015 and the following debate: Aoki, 2015; Searle, 2015; Sugden, 2015; see also Hodgson, 2006).

A natural interpretation of our experimental results is that devising a collective choice procedure that permits subjects to agree under a veil of ignorance induces the actual convergence on a certain principle of distributive justice. And then, agents' behavior will *de facto* converge on playing individual strategies compliant with the agreed principle, even if this is not *prima facie* in their self-interest

The second way in which the results are relevant for institutional theory is in that it experimentally supports the bearing of the social contract view of institutions (see, for an analogous test, Ostrom *et al.*, 1992). This experiment takes the social contract, so to speak, to the lab, and shows that some social contract theoretical predictions are consistent with the observed behavior of subjects in experimental conditions (for previous results on the same line of research see Faillo *et al.*, 2015; Sacconi *et al.*, 2011; Sacconi and Faillo, 2010).

Now, one may ask whether the eventual agreement about distributive justice requires the veil of ignorance procedure. May it not be derived through a procedure under full information? Such a test could be done, but it would concern a different phenomenon, or alternatively its results would have an unclear interpretation.

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

Consider first bargaining under full knowledge. In this case results would be affected by the disagreement option (exit option) and the bargaining options (i.e. the shape of the bargaining frontier representing possible agreements). Assuming that the disagreement option was null for both and the possibility of the agreement were symmetrical the outcome should be egalitarian (Harsasyi, 1977; Nash, 1950). But this would reflect equality in the players' bargaining position, not their undertaking a normative impartial perspective. Assume on the contrary that the exit option was strongly asymmetrical. Players would be expected to consider their personal product as the concession limit; so they would accept nothing less. This leads to questions about bargaining game theory, which is not our focus here.

Now consider the case of a free discussion on principles of distribution in a state of complete information. It is certainly possible that some participants would claim some redress for the unlucky party on the basis of impartiality, fairness or sympathetic feelings. The reasons for the agreement would be all but impossible to disentangle. The interpretation of whatever results would be extremely uncertain. It would be utterly difficult to answer our research question since this would require to know whether the deliberating parties adopted a normative stand characterized by impartiality. Let us recall that the goal of the experiment is to test the motivational force and realism of an *impartial* agreement. The most parsimonious way to introduce an impartial bargaining procedure seemed to be the adoption of the 'veil of ignorance', as was used by Rawls, Harsanyi and other moral and political philosophers.

The article is organized as follows. In the next section (section 2) the theoretical framework for the experiment is laid out. A general characterization of the principle we hypothesized would be chosen *ex ante* is offered in this section. Section three is devoted to the experimental design and procedures. The fourth section justifies and describes operational hypotheses. Section five presents the results. The final section discusses the meaning of these results, describes its implications for institutional design, acknowledges the limitations of this study and suggests future research.

2. Impartial agreement and compliance: Theoretical background and experimental evidence

In this section, we set out to explain the theoretical basis for the operational hypotheses.

Since the experiment introduces a 'veil of ignorance' procedure *ex-ante*, we draw upon theories of distributive justice that use that method. Several principia and criteria for justice have been argued to derive from impartial choice or agreement –among them Harsanyi's maximization of average utility principle (Binmore, 1989; Harsanyi, 1982; Traub *et al*., 2005). However, liberal egalitarian principles (henceforth LE) occupy a prominent place (Binmore, 2005; Brock, 1979; Dworkin, 1981a, 1981b; Gauthier, 1986; Rawls, 1971; Roemer, 1986; Sen, 2009). Our theoretical conjecture is that the *ex-ante* agreement would make subjects converge on LE. In a hypothetical impartial agreement there is no reasonable basis for assigning people different levels of basic resources or rights, but it may be justified that people are rewarded according to their merit or contribution to common endeavors. This is an incomplete characterization of LE, but it suffices to explain our predictions regarding the distribution rule chosen *ex-ante*.

This characterization views LE as a twofold principle: a principle of equality (in resources) and a principle of reward related to contribution.[1] This twofold principle implies that, when the starting point is not equal, LE may require a redistribution of what immediately derives from personal work. This is the formulation of the principle that will be adopted:

Liberal Egalitarian Redress Rule: if voluntary contributions to the production of a common output are made by means of arbitrarily allocated endowments, the difference in the outputs that agents contribute should not translate into an identical difference in the final distribution; on the contrary, the differential outputs obtained from differential endowments should be *distributed equally*.

The rationale for this formulation will become clear after the description of the experiment. At this point, it is important to stress that even this simplified LE principle requires a complex cognitive operation. People may be expected to agree on equal shares; or on simpler entitlement rules. Under this

[1]For a model of constitutional and post-constitutional contracts that formalizes the intuition of LE as a two-tier view of justice see Brock (1979); Sacconi (2006; 2011a); Fia and Sacconi (2019).

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

light, the first question posed above turns into whether people –in a suitably fair position– are capable of agreeing on a simplified LE principle of distribution that redresses arbitrary inequalities in productive means.

Let's move now to the problem of compliance. In a context without external enforcement, in which the agreement must be hold up through the subjects' voluntary compliance, once the veil is lifted, agents may wish to change their *ex-ante* choice. Hence our second question: Would ordinary agents, in particular those who have worked from an advantaged position, stick to the rule of distribution they chose behind a veil of ignorance? We ask this question in the most radical way. The interaction we design virtually invites agents to be self-serving. But the subjects will have just agreed –or so we expect– that LE distribution rule *should* be applied. This is about whether it is realistic to assume a motivation to comply *ex-post* with a rule that is seen as justified *ex-ante*.

Both mainstream economics and new institutional economics would anticipate a negative answer to these two questions. The fictitious agents of armchair philosophers may well agree on LE, but ordinary people in a lab would be expected to focus on simpler distribution rules. Second, maximizing individual payoff is the dominant move in the game played *ex-post*. A self-interested-with astuteness agent will exploit the opportunity to circumvent the *ex-ante* agreement and will defect. However, if Rawls's conception of a sense of justice and the theory of conformity preferences (see section 4 below) hold, they would account for positive answers to these questions.

It is not the first time that the twofold element in LE principles has been explored in experimental economics (Konow, 2000, 2001, 2005). Cappelen *et al.* (2010), Cappelen *et al.* (2014) and Mollerstrom *et al.* (2015) present evidence that people can distinguish legitimate entitlements: people tend to justify the appropriation of what is in control of the agents, while they tend to be egalitarian about what is not.

An important strand of related experimental literature has taken up Rawls's second principle, which represents the liberal component in liberal egalitarianism (Brickman, 1977; Frohlich *et al.*, 1987; Frohlich and Oppenheimer, 1990, 1992; Jackson and Hill, 1995; Michelbach *et al.*, 2003; Yaari and Bar-Hillel, 1984). Also, for the first part of our design, experiments have been conducted about the effects of the choice behind the veil of ignorance on stated preferences for redistribution (e.g. Becchetti *et al.*, 2018; Schildberg-Hörisch, 2010).

However, no study (except Sacconi and Faillo, 2010, and Faillo *et al.*, 2015) has considered the motivational problem of *ex-post* compliance with a principle chosen behind the veil.

The novelty of our experiment is that subjects are asked not so much to make a judgment about possible distributions, but actually to opt for a division rule and then face the choice, affecting their own income, of whether to implement or ignore it.

3. Experimental design and procedures

The experiment had three treatments labeled 'Noveil', 'Bargaining' and 'Chat'. In all treatments, subjects were matched in pairs and asked to perform a task; one subject was randomly given 6 minutes and the other 10 minutes to perform the task; this variable instantiates the random allocation of production means in the experiment. Subjects generated an amount of money that depended on the output of the task. At the end of the task, each subject was asked how to divide the total produced by the pair. They could choose whatever division by opting for a percentage to be assigned to each member of the pair. Moreover, in order to provide explicit representation to distributions associated with principles of justice discussed in the literature, we identify five 'division rules' – described below – which may be selected by subjects. Once the subjects decided, one of the two members was randomly selected and her choice implemented.

Henceforth, we will name the decision taken at the end of the task as '*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, before performing the task and before knowing who would have 6 or 10 minutes to perform it, the members of each pair had to agree on one of the five-division rules. We refer to this phase as the '*ex-ante* agreement'. For the sake of comparability, we will also refer to the division choice

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

as the '*ex-post* choice' in the Noveil treatment, even if in the Noveil there was no '*ex-ante* agreement.' Subjects were aware of all of the phases of the experiment from the beginning, before they made their first choice.[2]

3.1. Noveil treatment

The treatment consisted of a practice phase, a task phase, and a division phase. We describe the 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.

3.1.1 The task

The task consisted of encoding words. In each pair, one subject was randomly given a time of 10 minutes to perform the task; the other was given six. Information about the time limits was given just before the task. A sequence of words appeared on the subjects' screens. 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 incorrect) was written for the current word. A countdown on the screen displayed the time remaining. The total production (i.e. the number of tokens – one token = 0.15 euro) generated in the task corresponded to the number of words correctly encoded by the two subjects.

At the end of the task, subjects were informed about the total production of the pair (total number of words encoded correctly), individual production of each member of the pair, individual productivity (words/minute), production and productivity of the subject with 10 minutes both in the first 6 minutes and in the subsequent four.

3.1.2 The division phase and the rules

In the division phase ('*ex-post* choice'), each subject was asked to choose how to divide the income generated by the pair in the task. She could do this either by choosing a percentage from 0 to 100% of the total income for herself or by choosing one of the five-division rules. Subjects saw on their screens the final payoffs corresponding to the application of each rule. The option of free percentages also characterized the *ex-post* division choice in the two treatments with the agreement (see below). The possibility of choosing a free percentage put compliance with the rule agreed behind the veil of ignorance in the worst condition to be realized. Percentages ensured that no subjects complied with the agreement due to the lack of alternatives.

The five-division rules were:

Rule 1 – Equal split: each subject obtains exactly half of the total output produced by the pair.
Rule 2 – One gets all: one subject obtains the total output produced. Choosing this rule means asking for 100% of the total output produced by the pair.
Rule3 – One subject gets what she has produced: each subject obtains exactly what she has produced through her word-coding activity.
Rule 4 – Time independent division: each subject obtains what she has produced through her activity during the first 6 minutes; what is produced in the last 4 minutes by the subject who has 10 minutes is split evenly between the two subjects.
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.

Rule 1 is an application of a pure egalitarian principle. Rule 2 reflects pure opportunism in which each party, if selected to play the dictator role, is entitled to appropriate the whole output. Rule 3 is based on the idea of compensation directly proportional to contribution. Rule 4 is the liberal

[2]Instructions and zTree screenshots are included in the SOM - section IV and V.

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

egalitarian redress rule described in section 2. Rule 5 rewards individuals' productivity without considering their actual contributions.

Note that, according to Rule 4, the subject endowed with more time will spend an additional effort for which she is not fully compensated (her production in the extra time is to be split equally). Those who accepted Rule 4 have discounted this cost. Our admittedly simplified LE redress rule accommodates situations like our experiment, where the difference –the unpaid effort of four extra minutes of quite simple work– was practically negligible; and so, it should not matter very much to the subjects. Moreover, concern for this issue is overridden by consideration of the focal unfairness in the situation –i.e. the sharp inequality of the initial allocation of practical opportunity to work. In any case, the imperfection in covering extra effort costs introduces some attrition that worked against –and made bolder– the conjecture concerning the willingness to opt for the LE Rule 4 that we aimed to corroborate with this experimental design.

Subjects could read the text of the rules, and they were also shown the payoff which they would obtain if that rule was applied for the division of the output. Once both members of the pair made their decisions, one of them was randomly selected and her decision was implemented.

3.1.3 The practice phase

Before starting the task, subjects could practice, individually and for 5 minutes, with the rules by using a simulation platform. They could read the rules on their screen and choose one of them. They could also insert the number of words encoded by the person with 6 minutes and by the person with 10 minutes both in the first 6 minutes and in the remaining 4 minutes, and they could decide the person whose final choice would be selected.

In a typical session of the Noveil treatment, a participant starts with the practice phase, then the experiment begins with the task, in which she is assigned either 6 or 10 minutes. Once the task is completed, after having received the information concerning the performance in the task, each subject makes her choice, selecting one of the rules or a free percentage, knowing that a random draw will determine whether her choice or the choice of her counterpart will be implemented. Finally, she is informed of the outcome of the draw and the payoffs of the pair.

3.2. Bargaining treatment

In the Bargaining treatment, the practice, the task and the division (or '*ex-post* choice') phases were the same as in the Noveil. However, the task and the division phases were preceded by a stage in which the members of the pair, before knowing the time allocation, could reach an *ex-ante* agreement on one of the same five rules by means of a bargaining procedure. The procedure consisted of a maximum of 13 rounds. In the first six rounds, subjects simultaneously chose one of the rules, proposing it for the final division of the total product. 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, it was considered an agreement. Pairs unable to reach an agreement in the first six rounds accessed a second bargaining stage of four sequential choices. Each sequential choice consisted of an offer and, if rejected by the receiver, a counter-offer. At the beginning of each sequential choice, one of the members of the pair was randomly selected to make the first offer. If the recipient of the offer rejected it, then she had to make a counter-offer that might be accepted or refused by the counterpart. Pairs that failed to reach an agreement in this second stage moved to a final sequence of three further simultaneous choices. The subjects knew that the rule was not going to be enforced, but they also knew that, if they failed in reaching the agreement, they would be excluded from the experiment and would be asked to fill in a questionnaire unrelated the experiment. In this case, their earning would be equal to the show-up fee of 3 euros. All pairs reached the agreement within the 13 rounds.[3]

[3]See SOM, Section IV – Instructions Bargaining treatment for the rationale behind the bargaining procedure adopted.

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

The agreement phase was preceded by the practice phase, as in the Noveil treatment.

In the *ex-post* choice, subjects were reminded of the rule chosen by their pair (the rule also appeared with a different background color) in the *ex-ante* agreement. They could choose either a percentage of the total product, a division of the total product corresponding to an application of the agreed rule, or a division corresponding to the application of a different rule. The final payoffs corresponding to the application of each of the five rules appeared on the subjects' computer screens.

3.3. Chat treatment

In the Chat treatment, subjects also had to reach an agreement on one of the five-division rules in order to access the task and the *ex-post* division phase. The *ex-ante* agreement procedure was based on an anonymous chat. Subjects were given 5 minutes for discussion. Communication of personal information, PC number, threats, promises of side payments, and the use of offensive language were prohibited. Once members of the pair reached an agreement, within the limit of 5 minutes available to discuss through the chat function, they had to click on the same rule on a choice screen similar to the final division choice screen. Selecting the same rule on the screen after having agreed to it in the chat was a way to make clear that the agreement had been reached and that there was no misunderstanding about it.[4] All the pairs succeeded in choosing the same rule (it took an average of 3.75 minutes). As in the Bargaining treatment, they knew that the agreement would not binding, but if they failed to reach the agreement they would be excluded from the experiment.

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, a division corresponding to the application of the agreed rule, or a division corresponding to the application of a different rule.

3.4. Beliefs and questionnaire

In all treatments, at the end of the *ex-post* choice, before a subject knew if her choice had been selected for payment, first- and second-order beliefs were elicited by asking what she believed the other member of the pair had chosen (any of the five rules or a percentage of the total product) and what she believed the other member believed about her own 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.[5]

3.5. Sessions and procedures

The experiment was programmed using zTree (Fischbacher, 2007) and conducted at the EGEO laboratory of the University of Granada. Subjects were paid a 3-euro show-up fee. 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.

At the beginning of each session, participants were welcomed, asked to draw lots and 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. 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.

[4]If two subjects chose a different rule, a warning message appeared and they could make another choice. Only one mistake was allowed.

[5]The questionnaire is included in the SOM, section IV.

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

4. Hypotheses

We assume that making an agreement under the veil of ignorance would elicit a preference for the LE redress principle.[6] It is also assumed that the empirical suppositions behind theories for LE – as exemplified by the Rawlsian postulate of a sense of justice and the theory of conformity preference (Checchini Manara and Sacconi, 2019; Grimalda and Sacconi, 2005; Sacconi, 2007, 2011b, 2011a; Sacconi and Faillo, 2010; Sacconi and Grimalda, 2007) – are correct. According to the conformity preference theory, what counts in engendering preferences for conformity is (i) participation in the impartial agreement, and (ii) that the agreement elicits the mental model of an agent who –having agreed– simply intends to carry out the agreed action and, by default, believes that the other agreeing party will also comply.

Hypothesis 1: In the *ex-ante* choice, subjects should focus on Rule 4 (time-independent division).

LE is not the most symmetrical or simple rule. Nevertheless, it is the focal rule in a moral sense. Self-interest is canceled under the veil of ignorance and therefore salience must be found in what results from impartial reasoning. Rule 4 (the LE redress rule) is the rule that requires less justification, given an impartial standpoint: it neutralizes unjustified inequality. But it does not require that production from an eventually harder-working party be shared. The extra time does not ground any claim; performance, within the limits of equal time, may. Ideally, time (endowment) should be equal if the game is to be fair. But since the subjects are aware that the distribution will be unequal, the best they can do to rectify the random inequality is to apply some redress mechanism.

Hypothesis 2: Subjects generally comply *ex-post* with the rules they have selected by agreement *ex-ante*. Principles that may not have been followed in the non-agreement treatment (Noveil) prove highly followed in other treatments, as an effect of the agreement itself.

The sort of normative reasoning and agreement induced by the veil of ignorance has the power to turn the subsequent interaction into a moralized one. We suggest that people will tend to see the distributive problem as a decision situation in which the rational thing to do is to comply with the justified principle.

This conjecture, in the context of our design, rests on the assumption that agents possess what Rawls called a 'sense of justice' (Rawls, 1971, ch. VIII). They are ready to choose according to agreed principles of justice, even if this is personally costly, provided there is a common expectation that others will do the same. *Ex-ante* agreement, in conjunction with the formation of mutual expectations of compliance, entails an attitude of reciprocal compliance. If this assumption holds, our subjects should comply with their agreed choice.

As a combination of Hypothesis 1 and Hypothesis 2 we put forward our third hypothesis:

Hypothesis 3: The frequency of *ex-post* choices of Rule 4 in the two treatments with an agreement is significantly higher than in the Noveil treatment.

Hypothesis 4: Chat treatment induces more compliance than the Bargaining treatment

Even if anonymity is preserved, the mere fact that there is a chat open for several minutes, in which the subjects may use natural language to write down their thoughts about the situation and to exchange considerations about the meaning of the different rules, makes a difference. This form of reaching an agreement makes the normative nature of the situation even clearer. In addition, the

[6]One may wonder if the agreed rule in the *ex-ante* choice affects productivity If Rule 4, that we suppose being the focal rule in a moral sense behind the veil of ignorance, undermined subjects' productivity, a conflict between efficiency and distributive justice would arise. In general, as we will see in the next section, this is not the case.

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

chat exchange reduces the social distance[7] and may increase the disposition not to take advantage of one's partner.

Hypothesis 5: The high level of compliance in the *ex-post* choice is correlated to a high level of beliefs aligned with compliance, i.e. the subjects' first and second-order beliefs predict reciprocal compliance when we observe that they comply in the *ex-post* decision.

We draw hypothesis 2 from the idea of a 'sense of justice', and from the formalization of this idea through the theory of conformity preferences. According to this theory, a preference for conformity is activated by means of an impartial agreement under the condition that the parties hold first (what agents believe others will do) and second-order (what agents believe others believe they will do) positive beliefs about conformity (Faillo *et al.*, 2015; Grimalda and Sacconi, 2005; Sacconi, 2007; Sacconi and Faillo, 2010). Therefore, if we predict a high degree of compliance, we must also predict the corresponding first and second-order beliefs aligned with conformity.

5. Results

Data have been analyzed by performing nonparametric tests and multivariate analyses which allowed us to control both for socio-demographic characteristics of participants in the experiment and for aspects connected to the experimental conditions (see below for the description of control variables).

RESULT 1: LE is preferred in the *ex-ante* agreement.
In the *ex-ante* agreement of the Chat and Bargaining treatments the majority of pairs agreed on the liberal egalitarian rule (Rule 4).

Figure 1 shows the percentages of subjects who chose the five rules across treatments.[8] Rule 4 was chosen by the majority of subjects in the *ex-ante* agreement – 57.5% of subjects in the Bargaining treatment and 57.89% in the Chat treatment.[9]

This evidence supports Hypothesis 1.

The choice of Rule 4 does not seem to negatively affect productivity. When we focus on the subsample of subjects involved in the Chat and Bargaining treatment, we find that subjects who agreed on Rule 1 have lower productivity than subjects who agreed on Rule 2, 4 and 5 (statistically significant result at 10% level)[10], while no specific trend characterizes subjects who agreed on Rule 4 (average

[7]Social distance decreases when "the "other" is no longer some unknown individual from some anonymous crowd but becomes an "identifiable victim" (Schelling, 1968). The experimental literature provided wide evidence of how *ex-ante* communication promotes altruism (e.g. Bohnet and Frey, 1999), cooperation (see Balliet, 2010 for a comprehensive review) and coordination (e.g. Cooper *et al.*, 2018).

[8]As a whole, in the Chat and Bargaining treatments, only 22 subjects (14.10%) opted for a percentage in the *ex-post* choice. Among them, 11 chose to equally split the total production and one subject opted for the 100%. We consider the *ex-post* choice of opting for 50% or 100% as equivalent to the *ex-post* choice of Rule 1 or 2, respectively. In fact, in terms of the *ex-post* division, opting for Rule 1 (Rule 2) in the *ex-post* choice is equivalent to 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 footnotes. With respect to the other subjects, the 55%, 70% and 75% has been opted for by one subject each; 80% and 90% by two subjects each and 60% by three subjects. In general, we do not find a clear correspondence between the percentage of the total production that would have been obtained by these subjects if they had complied with the rule agreed behind the veil and the percentage chosen in the *ex-post* division, with 7 out of 10 subjects who, *ex-post*, asked for a higher part of the total payoff.

[9]A test of proportions revealed that Rule 4 was chosen by a proportion of subjects significantly greater than 40% ($p < 0.01$) in both the Chat and in the Bargaining treatment.

[10]Two-sample Wilcoxon rank-sum (Mann–Whitney) test, p. = 0.0680 – Rule 1 vs. Rule 4; p = 0.0555-Rule 1 vs. Rule 5; p = 0.0943 Rule 1 vs. Rule 2 (consider however that the *ex-ante* choice of Rule 2 concerns only two subjects).

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

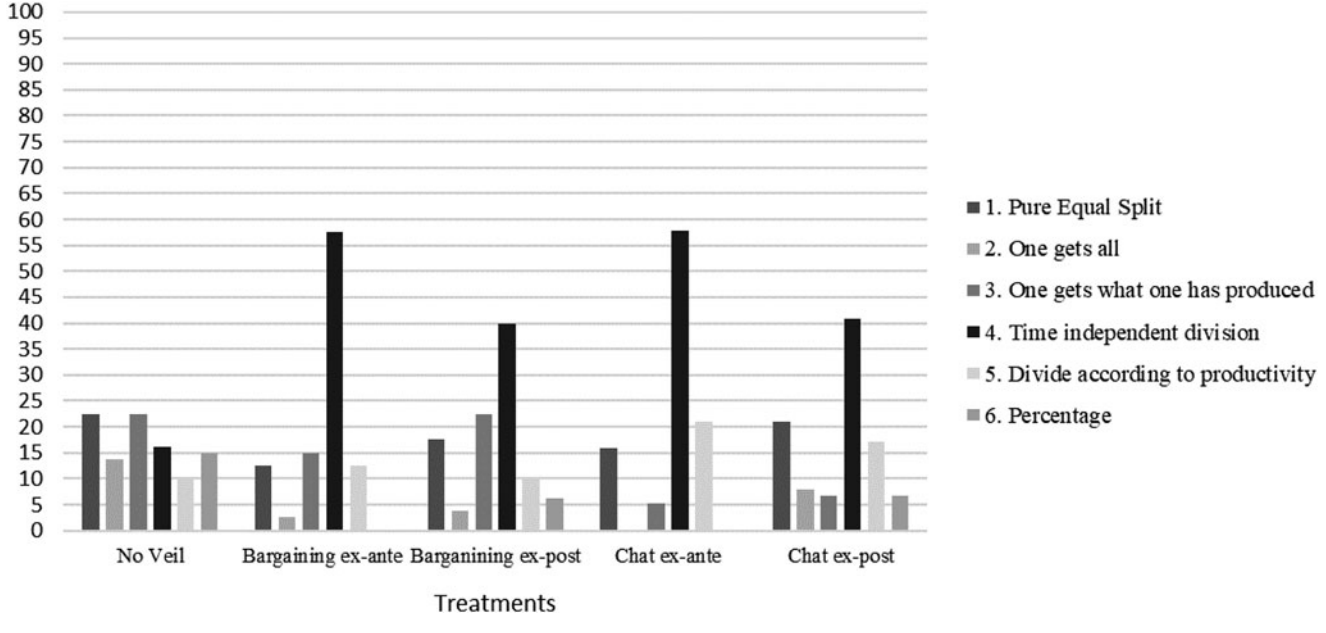

Figure 1. Rules chosen across treatments (percentage values).

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

productivity: Rule 1 = 4.311; Rule 2 = 5.485; Rule 3 = 4.669; Rule 4 = 4.686; Rule 5 = 4.858). Moreover, no statistically significant difference emerges when we compare the productivity of subjects who agreed on Rule 4 in the Chat and/or in the Bargaining treatment and subjects involved in the Noveil (Two-sample Wilcoxon rank-sum (Mann–Whitney) test, $p. = 0.2116$ – Chat; $p = 0.1477$-Bargaining; $p = 0.1013$ Chat and Bargaining).[11]

RESULT 2: *Ex-post* compliance with *ex-ante* agreement is high.
Most of the members of pairs who reached an agreement behind the veil of ignorance complied with it.

Overall, 60% of subjects complied with the rule agreed in the *ex-ante* phase. Table 1 shows the level of compliance across treatments and rules chosen in the agreement. 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 higher percentages of compliance were observed with Rule 1 and 5 in the Chat treatment (the determinants of compliance are discussed below).

Table 2 shows the dynamics concerning the rule chosen in the *ex-ante* agreement and those selected in the *ex-post* division choice.

When considering subjects who did not comply with Rule 4, we find that they mainly opted for Rule 3 (37%) and Rule 1 (26%).

The high level of compliance in the *ex-post* choice is in line with Hypothesis 2.

RESULT 3 The high level of compliance that characterizes subjects who agreed on the liberal egalitarian rule (Rule 4) ex-ante generates a significantly higher frequency of ex-post choices of Rule 4 in the two treatments with agreement than in the Noveil treatment.

The combination of *ex-ante* and *ex-post* choices in the two treatments with agreement results in a significantly higher frequency of choices of Rule 4 in the *ex-post* phase with respect to the Noveil treatment (Pearson $\chi^2(1)$: Bargaining vs. Noveil, pr = 0.001; Chat vs. Noveil, pr = 0.001); no difference emerges between the Bargaining and Chat treatments (Pearson $\chi^2(1)$, Pr = 0.920).

To check for the treatment effect on the *ex-post* choice of Rule 4, we ran a Logit regression (see Appendix – Table A1). The dependent variable was a binary indicator (*Rule_4_ex-post*) which took value 1 if subjects opted for a division consistent with Rule 4 in their *ex-post* choice. The independent variables of primary interest were the two dummies identifying the treatment in which subjects were involved, i.e. *Chat* and *Bargaining*. Estimates included socio-demographic characteristics – age, sex, income, propensity for financial risk, religious orientation, a propensity to trust strangers–, variables connected to the experimental conditions – the number of words encrypted in the task, the number of words encrypted per minute as a measure of subjects' productivity in the task– and the fact of having already taken part in experiments.[12] All the estimation results commented in this section are robust to

[11]One may also wonder if subjects who agreed on Rule 4 tend to reduce their productivity in the last 4 minutes. Through Two-sample Wilcoxon rank-sum (Mann–Whitney) tests we find that this is not the case, neither when we consider all subjects who agreed on Rule 4 ($p = 0.1937$) nor when we consider subjects who agreed on Rule 4 and comply with it in the *ex-post* choice ($p = 0.6064$).

[12]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.

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

Table 1. 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(5)	50(1)	50(6)	54.35(25)	20(2)
Chat	91.67(11)	*No obs.*	50(2)	68.18(30)	68.75(11)

Table 2. Subjects' choice transition considering the *ex-ante* agreements and *ex-post* decisions

Rule	1 – ex-post	2– ex-post	3– ex-post	4 – ex-post	5– ex-post	Percentage
1 –ex-ante	16 (72.73%)	0	0	0	3	3
2 – ex-ante	1	1 (50%)	0	0	0	0
3 – ex-ante	1	3	8 (50%)	3	0	1
4 – ex-ante	9	4	13	55 (61.11%)	5	4
5 – ex-ante	3	1	2	5	13 (50%)	2

Compliance rate in parentheses. The last column reports the number of subjects who opted for a percentage value in the *ex-post* choice.

the consideration of the previous control variables (see section I of the Supplementary Online Material – SOM – for variable legend and descriptive statistics).

The estimates reported in Table A1 show that 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 (Column 1 – first two lines).[13] Meanwhile, no difference emerges between Chat and Bargaining (Column 1 – last line).[14] These results show that Rule 4 is chosen significantly more in the treatments with agreement than in the Noveil treatment: 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.

Included in column 2 of Table A1 is the dummy variable *Rule_agr_4*, equal to 1 if subjects opted for Rule 4 in the *ex-ante* agreement. This variable, which captures the role of the agreement in subjects' *ex-post* choice, significantly affected the decision to select a division consistent with that rule in the *ex-post* choice. 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 Noveil; i.e. this confirms the role of agreement in explaining why a greater percentage of subjects choose Rule 4 *ex-post* in the Chat and Bargaining treatment.

Result 3 supports hypothesis 3 concerning the frequency of ex-post selection of Rule 4 in the two treatments with an agreement.

RESULT 4: Chat is more effective than Bargaining in inducing compliance.

Table A2 in the Appendix shows the econometric analysis of the determinants of compliance. Column 1 shows the effect of the two treatments on compliance. Estimates refer only to subjects in Chat and Bargaining. With respect to Table A1, we added to the control variables the payoff associated with the rule agreed in the *ex-ante* agreement (*Payment_agreement*) and 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

[13]When we consider interaction terms (see footnote 12), we find that the level of significance disappears with respect to the Bargaining treatment.

[14]When the interaction terms are considered (see footnote 12), the difference between Chat and Bargaining emerges for Men, who opt for Rule 4 more in the Chat than in the Bargaining treatment.

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

subjects who agreed on Rule 4. The first variable, which is weakly statistically significant, controls for the possible role of material incentives in affecting the decision to comply.[15] The dummies concerning the rule chosen *ex-ante* reveal that the rule characterized by the higher level of compliance is Rule 1. No differences concern the level of compliance with respect to the other rules.[16]

Table A2 – column 1 shows that the level of compliance in the Chat treatment is higher than in the Bargaining treatment.[17]

RESULT5: Compliance is explained by the alignment of beliefs.
In the *ex-post* choice of Chat and Bargaining treatment the majority of participants who complied with the agreed rule believed that: (i) their counterpart complied by choosing the ex-ante agreed rule; (ii) their counterpart believed that they had done the same (alignment of beliefs and choice).

This result confirms our fifth hypothesis about the role of beliefs in favoring compliance in the *ex-post* decisions.

In Table A2 – columns 2, 3 and 4– in the Appendix, we analyze the relationship between the decision to comply with the agreement and subjects' beliefs. Estimates consider only subjects involved in the Chat and Bargaining treatment. With respect to the estimates presented in Table A2 – column 1, we added the variable *Belief_aligned_compliance* which takes the value of 1 for subjects who believed that their counterpart would comply (first-order belief) and, at the same time, believed that their counterpart believed that they would comply (second-order belief). The significance of this variable in the regression presented in Table A2 – column 2, 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 beliefs, despite the differences characterizing the Chat and the Bargaining treatments, is also correlated with compliance when we consider separately the sub-sample of subjects involved in each of these two treatments, even though the level of statistical significance is lower when the considered sub-sample is the Bargaining (Table A2, columns 3 and 4; see SOM – Section III for additional analysis of beliefs across treatments).[18] Finally, when we include the *Belief_aligned_compliance* variable in the estimate, the significance of the *Chat* dummy disappears. This reveals that the positive effect of the Chat treatment on compliance is entirely explained by the role of beliefs that characterize this treatment.

5. Meaning, limitations and implications for effective institution design

This paper presents an economic experiment based on introducing a procedure for non-binding agreement in ignorance before the playing of a dictator with taking where the amount to be distributed is produced by the players through an activity they perform with unequal resources. Since players are free to claim as much as they want, and the agreement is non-binding, opportunism seems to be the conventional-knowledge prediction (Williamson, 1975, 1985). In fact, the control treatment without

[15]In our experiment subjects can always choose a percentage in the *ex-post* decision, so they always have the possibility to increase the payoff associated with the *ex-ante* agreed rule by choosing, *ex-post*, a percentage up to 100%. However, it could be interesting to explore whether low levels of payoff associated with the *ex-ante* agreed rule have any effect on compliance. What we observe is that the difference between the payoff obtained by subjects who complied with the agreement after having agreed on a rule, and the payoff that would have been obtained by complying by subjects who did not, is not statistically significant (Two-sample Wilcoxon–Mann–Whitney rank-sum test, $p = 0.2375$).

[16]Wald-tests available upon request.

[17]When we consider possible differences between men and women (see footnote 12), we find that this result holds only for men.

[18]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, (see footnote 8), and when the analysis focuses on the sub-sample of subjects involved in the Bargaining treatment (column 4), the *Belief_aligned_compliance* variable becomes statistically insignificant, albeit very close to the 10% level (11.1%).

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

the agreement procedure showed results in line with the literature on dictator with taking (Bardsley, 2008; Cappelen *et al.*, 2013; Faillo *et al.*, 2019; Korenok *et al.*, 2013; List, 2007). However, distributive decisions in treatments with the *ex-ante* agreement procedure widely depart from standard dictator games. A significant number of players adhere to the rule they voluntarily chose; and the chosen rule is for the most part a redress rule that we identify with liberal-egalitarian justice. This is an addition to the extant literature on distributive principles inspired by Rawls's second principle of justice, generally focused on the choice of a distributive scheme, rather than the choice of a distributive principle itself, and neglecting the motivational component (Frohlich *et al.*, 1987; Frohlich and Oppenheimer, 1990; 1992; Jackson and Hill, 1995; Lissowski *et al.*, 1991; Michelbach *et al.*, 2003; Yaari and Bar-Hillel, 1984). In our case, for the first time subjects distribute the income they generate through a real-effort task. The experiment shows a general tendency to agree on a re-dress principle and a strong motivational effect of that agreement.

Furthermore, our results show that the *ex-ante* impartial procedure for agreement does change voluntary individual behavior in an experimental context that captures key elements of productive organizations. Our claim is that this result is relevant for institutional economics insofar as it effectively challenges the assumption of opportunism in firms (Grossman and Hart, 1986; Hansmann, 1988; Hart and Moore, 1990; Williamson, 1975).

Most experimental subjects correctly analyzed arbitrary endowment inequality as injustice, accepted the rationality of redress, and then proceeded to act upon this principle. Their action in compliance with the redress rule is explained both by the fact that they needed to reach an *ex-ante* agreement, and by the fact that they did believe that the agreement was authoritative for the two parties, as is showed by their predominant mutual beliefs.

The experiment focuses on three simple questions inspired by the idea that institutional design based on principles of justice may not be unrealistically demanding. The conclusion is that it is not. Ordinary people are persuaded by normative criteria when they reason from an impartial situation. Furthermore, they are ready to act upon agreed fair rules, even when they focus on the individual decision of complying or not with the agreed criteria, a perspective that could activate the frame of a non-cooperative game. The *ex-ante* agreement seems to elicit a framing that activates the disposition to conform when other participants are taken to be participating in common action. Our study is a first step in buttressing an important set of normative theories which are usually criticized for being too removed from reality. Our conclusion is that, right or wrong, they are assuming nothing that cannot be assumed of most ordinary people.

Moreover, as shown by the data from the Chat treatment, the compliance effect is stronger if the agreement involved natural-language deliberation. This finding invites further exploration about the cognitive and motivational role of explicit agreement; in particular, agreement reached through participatory natural-language deliberation, as opposed to a consent-based agreement (Hielscher *et al.*, 2015).

At this point, one may ask how we can derive normative suggestions for institutional design from an experiment which should have mere descriptive or explanatory value. Are we falling in the famous naturalistic fallacy? Note that any suggestions are derived from the fact that the relevant normative principles are accepted by the parties themselves through their agreement. We suggest that if you want to obtain – in a controlled environment like a lab – a behavior consistent with a policy of redress of an unjust gain, you have just to devise rules of collective choice that allow agents to impartially decide by themselves the principles of distribution. No strong assumptions about agent's motivations or specific incentives are required. The principle will be *de facto* complied with by means of individual choices based on reciprocal expectations, in a way that may be understood as a stable institution in Aoki's equilibrium-sense (Aoki, 2001, 2007).

What we learn from our result is that in case the conditions created in our experimental settings can be approached in institution design, *de facto* adhesion to the LE principle of justice is not a 'mirage'.[19] It may seem a long shot from actually implementing justice in productive institutions. This contribution should be carefully developed in order to explore how to implement the idea of an *ex ante*

[19]Hayek (1976) dubbed the idea of social justice, a 'mirage'.

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

impartial agreement. This is beyond the scope of this article. However, let's note that the message is partly similar to Ostrom *et al*., (1992). And our results go even beyond Ostrom, Walker and Gardner: first we focus on distributive justice of a common output, not efficiency in the consumption of a CPR. Second, this is not just pre-play communication but impartial and impersonal procedures. This extends the scope of *ex-ante* agreements beyond the limits of small communities. Third, the explanation is based on the explicit hypothesis that utility functions are shaped according to a 'sense of justice'. The result is an institution that does not give up the equilibrium property, but consists in a self-sustaining equilibrium based on conformity preferences.

Acknowledgement. We wish to thank Cristina Bicchieri, Tom Donaldson, two anonymous reviewers and editors of the *Journal of Institutional Economics*, and the participants in the international workshop "The Social Contract in Corporate and Economic Ethics", University of Granada, May 2015, for their helpful and insightful comments and suggestions. We also thank the staff of the EGEO lab of the University of Granada for their support in the organization of the sessions. Remaining errors are solely the responsibility of the authors. This research was funded by a grant from the Spanish Ministerio de Economía y Competitividad – MINECO Projects BENEB (FFI2011-29005) and BENEBII (FFI2014-56391-P). This article was published in Open Access thanks to the support of the Erasmus + Programme of the European Union, within the Jean Monnet Module 'Explaining, Exploring and Expanding the European Peace' (Project no. 611350-EPP-1-2019-1-IT-EPPJMO-MODULE).

Financial support. This research was funded by a grant from the Spanish Ministerio de Economía y Competitividad - MINECO Projects BENEB (FFI2011-29005) and BENEBII (FFI2014-56391-P). This article was published in Open Access thanks to the support of the Erasmus+ Programme of the European Union, within the Jean Monnet Module "Explaining, Exploring and Expanding the European Peace" (Project no. 611350-EPP-1-2019-1-IT-EPPJMO-MODULE).

Conflict of interest. No conflict of interest.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017//S1744137422000029

References

Aoki, M. (2001), *Toward A Comparative Institutional Analysis*, Cambridge: the MIT Press.

Aoki, M. (2007), 'Endogenizing Institutions and Institutional changes', *Journal of Institutional Economics*, **3**(1): 1–31.

Aoki, M. (2015), 'Why is the Equilibrium Notion Essential for A Unified Institutional Theory? A Friendly Remark on the Article by Hindriks and Guala', *Journal of Institutional Economics*, **11**(3): 485–488.

Balliet, D. (2010), 'Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review', *The Journal of Conflict Resolution*, **54**(1): 39–57.

Bardsley, N. (2008), 'Dictator Game Giving: Altruism or Artefact?', *Experimental Economics*, **11**(2): 122–133.

Becchetti, L., G. Degli Antoni, S. Ottone and N. Solferino (2018), 'Performance, Luck and Equality: An Experimental Analysis of Subjects' Preferences for Different Allocation Criteria', *The BE Journal of Economic Analysis & Policy*, **18**(1): 20160259.

Binmore, K. (1989), 'Social Contract I: Harsanyi and Rawls', *The Economic Journal*, **99**(395): 84–102.

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

Bohnet, I. and B. S. Frey (1999), 'Social Distance and Other-Regarding Behavior in Dictator Games: Comment', *American Economic Review*, **89**(1): 335–339.

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., E. Ø. Sørensen and B. Tungodden (2010), 'Responsibility for What? Fairness and Individual responsibility', *European Economic Review*, **54**(3): 429–441.

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

Cappelen, A. W., U. Nielsen, E. Ø. Sørensen, B. Tungodden and J. R. Tyran (2013), 'Give and Take in Dictator Games', *Economics Letters*, **118**(2): 280–283.

Cecchini Manara, V. and L. Sacconi (2019), *Compliance with Socially Responsible Norms of Behavior: Reputation vs. Conformity*, Econometica working papers series, No. wp73, http://www.econometica.it/wp/wp73.pdf

Cooper, R., D. V. DeJong, R. Forsythe and T. W. Ross (2018), 'Communication in Coordination Games', in J. F. Shogren (ed), *Experiments in Environmental Economics* (Vol. **1** vols, Abingdon: Taylor and Francis Inc, pp. 345–378.

Dworkin, R. (1981a), 'What is Equality? Part 1: Equality of Welfare', *Philosophy and Public Affairs*, **10**(3): 185–246.
Dworkin, R. (1981b), 'What is Equality? Part 2: Equality of Resources', *Philosophy and Public Affairs*, **10**(4): 283–345.
Faillo, M., S. Ottone and L. Sacconi (2015), 'The Social Contract in the Laboratory. An Experimental Analysis of Self-Enforcing Impartial agreements', *Public Choice*, **163**(3–4): 225–246.
Faillo, M., M. Rizzolli and S. Tontrup (2019), 'Thou Shalt not Steal: Taking Aversion with Legal Property claims', *Journal of Economic Psychology*, **71**: 88–101.
Fia, M. and L. Sacconi (2019), 'Justice and Corporate Governance: New Insights From Rawlsian Social Contract and Sen's Capabilities Approach', *Journal of Business Ethics*, **160**(4): 937–960.
Fischbacher, U. (2007), 'z-Tree: Zurich Toolbox for Ready-Made Economic Experiments', *Experimental Economics*, **10**(2): 171–178.
Frohlich, N. and J. A. Oppenheimer (1990), 'Choosing Justice in Experimental Democracies with Production', *American Political Science Review*, **84**(2): 461–477.
Frohlich, N. and J. A. Oppenheimer (1992), *Choosing Justice: An Experimental Approach to Ethical Theory*, Berkeley: University of California Press.
Frohlich, N., J. A. Oppenheimer and C. L. Eavey (1987), 'Choices of Principles of Distributive Justice in Experimental Groups', *American Journal of Political Science*, **31**(3): 606–636.
Gauthier, D. (1986), *Morals by Agreement*, Oxford: Clarendon Press.
Grimalda, G. L. and L. Sacconi (2005), 'The Constitution of the Not-For-Profit Organisation: Reciprocal Conformity to Morality', *Constitutional Political Economy*, **16**(3): 249–276.
Grossman, S. J. and O. D. Hart (1986), 'The Costs and Benefits of Ownership: A Theory of Vertical and Lateral integration', *Journal of Political Economy*, **94**(4): 691–719.
Hansmann, H. (1988), 'Ownership of the Firm', *Journal of Law, Economics, & Organization*, **4**(2): 267–304.
Harsanyi, J. C. (1982), 'Morality and the Theory of Rational Behaviour', in A. Sen and B. Williams (eds), *Utilitarianism and Beyond*, Cambridge, MA: Cambridge University Press, pp. 39–62.
Harsasyi, J. C. (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge, Mass: Cambridge University Press.
Hart, O. and J. Moore (1990), 'Property Rights and the Nature of the firm', *Journal of Political Economy*, **98**(6): 1119–1158.
Hayek, F. A. (1976), *The Road to Serfdom*, London: Routledge.
Hielscher, S., M. Beckmann and I. Pies (2015), 'Participation Versus Consent: Should Corporations Be Run According to Democratic Principles?', *Business Ethics Quarterly*, **24**(4): 533–563.
Hindriks, F. and F. Guala (2015), 'Institutions, Rules, and Equilibria: A Unified theory', *Journal of Institutional Economics*, **11**(3): 459–480.
Hodgson, G. M. (2006), 'What are Institutions?', *Journal of economic issues*, **40**(1): 1–25.
Jackson, M. and P. Hill (1995), 'A Fair Share', *Journal of Theoretical Politics*, **7**(2): 169–180.
Kaplow, L. and S. Shavell (2002), *Fairness Versus Welfare*, Cambridge, Mass: Harvard UP.
Konow, J. (2000), 'Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions', *American Economic Review*, **90**(4): 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. (2005), 'Blind Spots: The Effects of Information and Stakes on Fairness Bias and Dispersion', *Social Justice Research*, **18**(4): 349–390.
Korenok, O, E. L. Millner and L. Razzolini (2013), 'Impure Altruism in Dictators', *Journal of Public Economics*, **97**(1): 1–8.
Lissowski, G. T. Tyszka and W. Okrasa (1991), 'Principles of Distributive Justice', *Journal of Conflict Resolution*, **351**: 98–119.
List, J. A. (2007), 'On the Interpretation of Giving in Dictator games', *Journal of Political Economy*, **115**(3): 482–493.
Mansell, S. (2013), *Capitalism, Corporations and the Social Contract: A Critique of Stakeholder Theory (Business, Value Creation, and Society)*, Cambridge: Cambridge University Press.
Michelbach, P. A., J. T. Scott, R. E. Matland and B. H. Bornstein (2003), 'Doing Rawls Justice: An Experimental Study of Income Distribution Norms', *American Journal of Political Science*, **47**(3): 523–539.
Mollerstrom, J., B. A. Reme and EØ Sørensen (2015), 'Luck, Choice and Responsibility – An Experimental Study of Fairness views', *Journal of Public Economics*, **131**: 33–40.
Nash, J. (1950), 'The Bargianing Problem', *Econometrica*, **18**(2): 155–162.
Ostrom, E., J. Walker and R. Gardner (1992), 'Covenants With and Without A Sword: Self-Governance is Possible', *American Political Science Review*, **86**(2): 404–417.
Rawls, J. (1971), *A Theory of Justice*, Cambridge, Mass: Harvard University Press.
Roemer, J. (1986), 'The Mismarriage of Bargaining Theory and Distributive justice', *Ethics*, **97**(1): 88–110.
Sacconi, L. (2006), 'A Social Contract Account For CSR as Extended Model of Corporate Governance I: Rational Bargaining and Justification', *Journal of business ethics*, **68**(3): 258–281.
Sacconi, L. (2007), 'A Social Contract Account for CSR as Extended Model of Corporate Governance (II: Compliance, Reputation and Reciprocity)', *Journal of Business Ethics*, **75**(1): 77–96.

Sacconi, L. (2011a), 'A Rawlsian View of CSR and the Game Theory of Its Implementation (Part II: Fairness and Equilibrium)', in L. Sacconi, M. Blair, E. Freeman and A. Vercelli (eds), *Corporate Social Responsibility and Corporate Governance: The Contribution of Economic Theory and Related Disciplines*, Basingstoke: Palgrave Macmillan, pp. 157–193.
Sacconi, L. (2011b), 'A Rawlsian View of CSR and the Game Theory of its Implementation (Part III: Conformism and Equilibrium Selection)', in L. Sacconi and G. Degli Antoni (eds), *Social Capital, Corporate Social Responsibility, Economic Behavior and Performance*, Basingstoke: Palgrave Macmillan, pp. 194–252.
Sacconi, L. and M. Faillo (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. and 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 P. L. Porta and L. Bruni (eds), *Handbook of Happiness in Economics*, Cheltenham Northampton, Mass: Elgar, pp. 532–570.
Sacconi, L., M. Faillo and S. Ottone (2011), 'Contractarian Compliance and the 'Sense of Justice': A Behavioral Conformity Model and its Experimental support', *Analyse&Kritik*, **33**(1): 273–310.
Schelling, T. C. (1968), 'The Life You Save May Be Your Own', in S. Chase (ed), *Problems in Public Expenditure Analysis*, Washington, DC: Brookings Institution, pp. 127–162.
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.
Searle, J. R. (2015), 'Status Functions and Institutional Facts: Reply to Hindriks and Guala', *Journal of Institutional Economics*, **11**(3): 507–514.
Sen, A. (2009), *The Idea of Justice*, Cambridge, Mass: Harvard University Press.
Singer, A. (2015), 'There Is No Rawlsian Theory of Corporate Governance', *Business Ethics Quarterly*, **25**(1): 65–92.
Sugden, R. (2015), 'On 'Common-Sense Ontology': A Comment on the Paper by Frank Hindriks and Francesco Guala', *Journal of Institutional Economics*, **11**(3): 489–492.
Traub, S., C. Seidl, U. Schmidt and M. V. Levati (2005), 'Friedman, Harsanyi, Rawls, Boulding – or Somebody Else? An Experimental Investigation of Distributive justice', *Social Choice and Welfare*, **24**(2): 283–309.
Williamson, O. E. (1975), *Markets and Hierarchies*, New York: The Free Press.
Williamson, O. E. (1985), *The Economic Institutions of Capitalism*, New York: The Free Press.
Yaari, M. E. and M. Bar-Hillel (1984), 'On Dividing Justly', *Social Choice and Welfare*, **1**(1): 1–24.

Appendix: Econometric results

Table A1. The determinants of the choice of the rule

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

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The last line of Table reports Wald-tests useful for comparing subjects' behavior in the Chat and the Bargaining treatments.
Full estimates results are reported in section II of the SOM.

https://doi.org/10.1017/S1744137422000029 Published online by Cambridge University Press

Table A2. The determinants of compliance

	(1) Logit Whole sample	(2) Logit Whole sample	(3) Logit Sub-sample of subjects involved in the Chat	(4) Logit Sub-sample of subjects involved in the Bargaining
Dependent variable: *Compliance*				
Chat	0.973** (0.407)	0.635 (0.454)		
Payment_agreement	0.0648* (0.0354)	0.0729* (0.0374)	0.208 (0.195)	0.0334 (0.0500)
Belief_aligned_compliance		1.993*** (0.456)	6.255*** (2.001)	1.162* (0.605)
Rule_agr_1	1.720** (0.731)	1.894** (0.771)	9.913 (15.76)	1.209 (1.003)
Rule_agr_2	−2.393 (2.045)	−1.565 (2.143)		−0.892 (2.383)
Rule_agr_3	0.280 (0.670)	0.0735 (0.705)	5.037* (2.769)	−0.221 (0.842)
Rule_agr_5	0.0659 (0.638)	0.0243 (0.690)	3.028 (4.038)	−1.094 (1.198)
Control variables	YES	YES	YES	YES
Constant	1.145 (14.66)	6.104 (16.43)	−11.22 (44.47)	25.52 (29.58)
Observations	156	156	76	80
Pseudo R^2	0.159	0.264	0.603	0.236

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
Full estimates results are reported in section II of the SOM.

Cite this article: Degli Antoni G, Faillo M, Francés-Gómez P, Sacconi L (2022). Liberal egalitarian justice in the distribution of a common output. Experimental evidence and implications for effective institution design. *Journal of Institutional Economics* **18**, 901–918. https://doi.org/10.1017/S1744137422000029