



Does impartial reasoning matter in economic decisions? An experimental result about distributive (un)fairness in a production context

(¿Importa el razonamiento imparcial en las decisiones económicas?
Un resultado experimental sobre (in)justicia distributiva en un contexto de producción)

Laura MARCON*, Pedro FRANCÉS-GÓMEZ, Marco FAILLO

ABSTRACT: The Rawlsian veil of ignorance should induce agents to behave fairly in a distributive context. This work tried to re-propose, through a dictator game with giving and taking options, a sort of original position in which reasoning behind the veil should have constituted a moral cue for subjects involved in the distribution of a common output with unequal means of production. However, our experimental context would unwittingly recall more the Hobbesian state of nature than the Rawlsian original position, showing that the heuristic resource to the Rawlsian idea of a choice behind the veil is inefficacious in distributive contexts.

KEYWORDS: justice, social contract, veil of ignorance, experiments, moral cues.

RESUMEN: *El velo de ignorancia rawlsiano debería inducir una conducta justa en contextos distributivos. Este estudio intentó, mediante un juego del Dictador con opciones de dar y tomar, re-crear una especie de posición original en la que el razonamiento tras el velo debería haber sido una señal moral para sujetos que debían distribuir una ganancia común conseguida con medios de producción desiguales. Sin embargo, el diseño experimental resultó recordar más al estado de naturaleza hobbesiano que a la posición original de Rawls, y demostró que el recurso heurístico a la idea de una decisión tras el velo de ignorancia es ineficaz en un contexto de producción y distribución.*

PALABRAS CLAVE: *justicia, contrato social, velo de ignorancia, experimentos, señales morales.*

* **Correspondence to:** Laura Marcon. Department of Philosophy I, University of Granada. Campus de la Cartuja, Calle del Prof. Clavera, s/n (18011 Granada) – marcon4@ugr.es ; lamarcon@correo.ugr.es – https://orcid.org/0000-0003-1514-507X

How to cite: Marcon, Laura; Francés-Gómez, Pedro; Faillo, Marco (2020). «Does impartial reasoning matter in economic decisions? An experimental result about distributive (un)fairness in a production context»; *Theoria. An International Journal for Theory, History and Foundations of Science*, 35(2), 217-233. (https://doi.org/10.1387/theoria.21011).

Received: 11 July, 2019; Final version: 06 February, 2020.

ISSN 0495-4548 - eISSN 2171-679X / © 2020 UPV/EHU



This article is distributed under the terms of the
Creative Commons Attribution 4.0 International License

I. Introduction

This paper reports an unexpected experimental result and contributes to the debate about other-regarding vs self-interested behaviours in non-cooperative games.

The experimental approach was intended to check the impact of ‘reasoning behind a veil of ignorance’ on an individual’s distributive decision. By making subjects think on the distributive decision they would be asked to make *before* they knew the one relevant difference that the experimenter was going to introduce between the two persons that should split the earnings, this experiment was designed as a partial test of this feature of Rawls’s theory. The veil, intended as a moral cue, should have induced a reflection from an impartial perspective, leading subjects to put themselves in the shoes of the least advantaged person once the veil would have been removed. Drawing on the philosophical assumptions of Kantian constructivism, we supposed that the mere conception of a distributive choice behind the veil of ignorance should create a normative stance towards the actual distributive decision, even if experimental subjects were not acting within the rules of any known institution. And we further supposed that subjects’ behaviour, exposed to a clear moral cue in distributing a common output, would adjust in line with their normative conclusion.

In fact, although the Rawlsian original position and the veil mechanism are abstractions, not traceable in everyday life, they are useful tools for studying individual behaviour, in situations of potential conflict between personal interest and the common good (Faillo *et al.*, 2015). It is a question of understanding whether there is a correspondence between Rawls’s theoretical apparatus on distributive justice and how, *de facto*, people behave in distributive contexts. The first empirical studies on Rawls’s theory (Frohlich *et al.*, 1987; Frohlich & Oppenheimer, 1990, 1992; Lissowski *et al.*, 1991) showed how the maximin principle defended by Rawls has no counterpart in the laboratory: subjects prefer to maximize the income after having established a floor constrain. Other results followed from questions on how subjects perceive the principles of allocation: Scott *et al.* (2001) and Michelbach *et al.* (2003) distinguished four principles of allocation —equality, efficiency, need, and merit— demonstrating how these principles often intervene simultaneously and in an interdependent manner so subjects can formulate judgments on distributive justice.

In parallel with this literature on how different allocation principles intervene to form judgments on distributive justice, there exists an ever-growing experimental literature on non-cooperative games. These experiments mostly investigate underlying motivations in giving behaviour. The assumption in experimental economics is that to give away money is costly for the individual and therefore it should be expected only within a framework of social, institutional or moral obligations coercively imposed. However, in some economic games, such as the Dictator, people seem to share for no reason whatsoever.

Some factors have been detected to explain why a fair distribution got the drop on the rational choice theory predictions: normative and empirical expectations established by social norms amongst players (Hoffman *et al.*, 1996; Bicchieri, 2006; Krupka & Weber, 2013); the reputation effect —subjects feel observed, hence judged, and they do not want to contradict their self-image (Servátka, 2009); the frame itself, conveyed by the game. In relation to the frame effect: if the taking option is also provided in a Dictator’s game, subjects may perceive choices as conflicting messages, deciding to take from the other person because it is one of the available alternatives (List, 2007; Levitt & List,

2007; Barsdley, 2008; Franzen & Pointner, 2012). Rigdon and colleagues (2009) used even a minimal social cue of a visual type, called watching-eyes configuration, demonstrating how its introduction had an impact in terms of increasing generosity. Another very important element is the endowment, or better, how it accrues to the subjects: if it is ‘*manna from heaven*’, namely offered by experimenters at the beginning of the game, or whether it is the product of subjects’ effort, behaviour on the final distribution changes (Faillo *et al.*, 2019).

Given this experimental literature, the actual game, explained in detail below, was a variation of the Dictator Game¹ (DG) that was initially used to question the assumptions of economic rationality. On average, people give on average 30%, which seems to show that ordinary people are not as selfish as economic theories assume. The introduction of Dictator with Taking changed this view, since in this version of the game, money is given to *both* players, and then only one of them (the dictator) has to decide whether to keep her portion, give part of it to the other player, or take some from the other. On average, subjects in this game *take* from their pairs, as pointed above, what seems to be a very selfish behaviour.

In our case, the experimental currency was earned by both players through a real-effort task rather than simply given to them by the experimenter, and each member’s contribution to the pair’s total output was common knowledge between them. The experiment introduces an asymmetry among players so that one member of the pair suffers an unjustified ‘disadvantage’ relative to the other. We expected our subjects to be aware of the need to redress the unjustified inequality. Our prediction was that whatever the average distribution was in the base treatment, introducing ‘reasoning behind the veil’ would move the average distribution towards a more ‘liberal egalitarian’ pattern.

In this framework, the liberal egalitarian perspective consists in a type of distribution that should compensate for the initial disadvantage – in our case, a shorter time limit to perform the task, which is a mere random element. The assumption is that those with more time available could produce more and therefore claim a ‘right’ to a greater income. So, to repeat, the main objective is to verify the effectiveness of the veil as a moral cue: without a veil we expect that subjects would have chosen a distribution based on merit (everyone gets what she/he has produced); while introducing the veil, we expect that subjects, realizing that the advantageous position will not be the result of merit but of mere fortune, should seek to level this inequality during the distribution stage, by introducing some re-distribution from the advantaged party to the disadvantaged one. In fact, reasoning behind the veil should lead each person to think of the possibility that she is the disadvantaged party and, wishing to protect herself from undeserved bad luck, establish a redress mechanism.

By isolating this element, we expected to gather empirical support for the Rawlsian contractual argument: if distribution changes after introducing deliberation behind the veil, this would imply that the original position story does track the constitution of our moral intuitions and norms regarding distributive justice. We had two arguments to sup-

¹ An experimental situation in which a subject is asked to split an amount of money, generally given by the experimenter, between himself and a second subject, who plays no role in the game, but simply receives what the ‘dictator’ decides.

port our working hypotheses: first, reasoning behind the veil of ignorance should direct our subjects to a generally egalitarian split of the cake; second, the fact that the thought experiment implied a normative reflection should make the rules of property and merit that are part of common morality even more salient.

We thought that the reasoning behind the veil meant introducing a moral cue within the decision-making process, given the previous experiments in moral psychology and behavioural ethics. A strong line of research in this area has worked particularly on the underlying dynamics about people's honesty and dishonesty. Some results show that trying to behave honestly is perceived as an effort, as a practice that is not automatic but requires strength of will and commitment (Aquino & Reed, 2002). Others point out that, on the one hand, people tend to justify their dishonest behaviour (Shalvi *et al.*, 2015), but, in many situations, the self-concept of being and perceiving themselves as moral persons is a motivation that leads people to be more honest (Mazar *et al.*, 2008). Ayal and colleagues (2015) have shown how the use of certain moral cues have an effect on human actions even in conditions of anonymity. One of the moral cues they used is called 'reminding', which, as confirmed by data, has been salient in influencing behaviour "utilizing principles of right and wrong, specific examples of morals and 'do's and don'ts', and even slogan" (Ayal *et al.*, 2015, p. 739).

So, we expected the thought experiment to measurably move whatever result we obtained in a baseline design (with no veil of ignorance, and no moral cue) towards a more equal split. However, the predicted move towards egalitarianism was not observed. This left us puzzled. We revised our procedures, method and assumptions —see discussion sections below— and found no significant flaw. We conjecture a philosophical explanation for our result, namely, that the *mere* use of a moral theory as a thought experiment is not enough for eliciting the behaviour prescribed by the theory. This is a major amendment to Social Contract theory —and in general to moral theories that rely on counter-factual reasoning. But we need to be careful with our conclusion; since this result is so opposed to literature on moral cues, further research is surely required.

The paper is structured as follows: the next section describes the experimental design and hypotheses in detail; results are presented in section three; section four summarizes and discusses our findings and explanatory conjecture.

II. *Experimental design and hypotheses*

In this study we compare two treatments, one called NOVEIL, which constitutes our baseline, and a second one named VEIL. In both 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, therefore making almost sure that their 'contribution' would be different, due in great part to a chance event (whether they have six minutes or ten minutes) that happens *before* the task begins and the earnings are collected; (iv) participants played a DG in which they can just keep their earned endowment, give part of their endowment to the other or take a part of the other's endowment; in other words, they can distribute the pair's total earnings as they wish. The underlying assumption behind condition (iii) is that the person with more available minutes has an advantage, *and* a corresponding responsibility; the foresee-

able 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.

Let us briefly note that the use of time as a basic resource to represent initial inequality has no precedent in experimental literature. The reason may be that more time may imply an extra effort, and therefore it may be taken to represent a disadvantage rather than an advantage. In this experiment this is not the case, since the task is easy enough, and the working time short enough in any case. We are sure—in part from the debriefing questions registered after the experiment—that having more time was invariably interpreted by the subjects as having an advantage. Other forms of representing initial inequality could have been used; but we found that working time required less intervention in the design of the rules of the game and was easily identified as a difference that happened *before* the task itself began.

In our baseline treatment (NOVEIL), subjects were randomly assigned computer-cubicles and they were anonymously paired with someone else in the room. They were informed that they were going to perform a task—which was coding words, translating them in numbers by using a conversion table—for which one member of the pair would have six minutes, and the other member ten minutes. They were informed that right before the task began, they would know whether they have six or ten minutes: this would show on their monitors. Once they were done with the task, they were informed about each member's performance (total number of words coded and productivity measured as words *per minute*). The same list of words was presented to every participant. 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 = €0.20. They received one token for each word correctly coded. At this point, each member of the pair had to decide how to divide the total output of the pair by claiming any percentage for herself. Once both members of the pair made their choices, one member was randomly selected and her choice was implemented. So, each participant decided under the expectation that her choice had a 50% probability of being her real final payoff. There were no further rounds, and each subject participated just once. Once the member with six minutes consumed her task-time, she would be playing a game unrelated with the experiment and with no effect on her payoff.

The VEIL treatment was the same as the NOVEIL except that, before the time limit was assigned and the subjects proceeded to the task and decision phases, they were asked how they think the total output should be divided, by stating how much (what percentage) they should claim for themselves. At this moment they knew the details of the game, but they ignored whether they will be given six or ten minutes, so they decided behind a veil covering the information about their labour time. Finally, the subjects knew that their choice at this time had no effect on their final choice after the task. This *ex ante* phase lasted for two minutes, which we calculated is plenty of time. We deliberately gave them time to spare. The goal was to make them think; we made clear that they could change their option any time until the questionnaire closed. Their choice was recorded, and then they proceeded as in the previous treatment: they were informed about the time assigned to each of them and they worked on the task. After the task, they could confirm their choice behind the veil, or change it.

In both treatments, instructions were read aloud by one the experimenters and a set of control questions were proposed to make sure that participants understood the instructions. Participants were students recruited at the University of Granada in May 2013. 50.6% of the participants were females, the average age was of 22.27 years, 96% were Spanish, with an average number of previous experiments of 2.13 (max = 9); 29% were enrolled in the Economics program and 61% in the Management program; 10 were enrolled in different programs (see table 1 for the data on the two treatments' samples). We conducted 4 sessions of 20 participants for each treatment, for a total of 80 participants. The average of payments was of €11.30 (included a show-up fee of €3). The experiments were run at the Egeo Lab (University of Granada). The experiment was programmed by using the z-Tree platform (Fischbacher, 2007). We used a between-subjects design; no subject participated in more than one treatment.

Table 1. The characteristics of the samples (standard deviations in parentheses)

	NOVEIL	VEIL
Age	22.3 (1.83)	22.3 (2.02)
Gender (% of female)	46.2	55.0
Nationality (% of spanish)	96.2	97.5
Major (% of economics and management students)	87.5	92.5
Number of experiments in which the subject took part.	2.15 (1.46)	2.12 (1.42)

We were interested in testing the two following hypotheses, respectively related to treatments:

Hypothesis 1: Agents in treatment NOVEIL should choose a distribution that tracks individual earnings.

It should be a cost, in moral terms, to steal from another who has earned a certain amount of money by working – namely the taking option. In addition, if social norms (about work, effort, and desert) carry over to the laboratory, they should weigh towards this distribution. In other words, even if agents find themselves in complete anonymity and may fear no punishment, the design of the experiment seems to call for them to keep what they have earned and leave to their partner what (s)he has earned.

Hypothesis 2: Agents in the VEIL treatment should choose a distribution that approximately tracks a liberal-egalitarian principle.

Agents in this treatment are subject to the thought experiment of the veil of ignorance. This provides an impersonal and impartial point of view that elicits a fairer and less selfish

behaviour: each individual is drawn to think as if (s)he could have more or less endowment, with no reason (randomly). From this each participant is aware that the design involves an unjustified inequality; and they have the power to ‘correct’ it by choosing to distribute the common output in a more egalitarian way –even if, since they have data about individual productivity, we never expected convergence on pure egalitarian distributions. From a moral reasoning perspective, our hypothesis 2 means that the Rawlsian assumption about the moral capacities of people prevail over the Hobbesian assumption; in other words, the hypothesis implies that moral reasoning and moral conclusions would have an effective power to shape behaviour even in absence of external public authority.

III. Results

Starting from the evidence on the task, we run a two-way ANOVA to assess the effect of treatment and time on the total number of words encoded (table 2). We do not observe any significant difference between treatments in terms of amount of words encoded ($F(1,156) = 0.00$, $p = 0.97$) while, as expected, the production of the participants with ten minutes is higher than that of participants with six minutes in both the treatments ($F = 333.3$, $p < 0.01$; interaction of treatment and time: $F = 0.00$, $p = 0.96$). Participants received one token for each word correctly encoded, so the numbers in table 2 correspond also to participants’ average earnings. The level of productivity (words per minute) is 5.15 for the participants with six minutes and 5.26 for those with ten minutes in both the treatments (time: $F = 0.60$, $p = 0.44$; treatment: $F = 0.00$, $p = 0.97$; interaction: $F = 0.00$, $p = 0.99$).

Table 2. Production (correctly encoded words)

		Time	
		6 minutes	10 minutes
Treatment	NOVEIL	30.90 (5.82)	52.62 (8.45)
	VEIL	30.92 (5.37)	52.67 (9.62)

Means, standard deviations in parentheses.

We can then put forward our first result.

Result 1. Production and productivity.

The production of participants with the same endowment in terms of time is the same across treatments. In both treatments, the level of production of participants with ten minutes is higher than that of participants with six minutes. There are no differences in the level of productivity, neither between treatments nor between subjects with different time groups.

Table 3. Percentage and number of tokens (percentage \times total production) claimed after the task

		Percentage		Tokens	
		6 minutes	10 minutes	6 minutes	10 minutes
Treatment	NOVEIL	74.50% (19.20)	77.50% (17.05)	62.29 (19.10)	64.35 (15.17)
	VEIL	74.00% (19.18)	85.75% (13.93)	61.59 (18.12)	71.87 (16.10)

Means, standard deviations in parentheses.

Table 3 (columns 3 and 4) reports the number of tokens claimed by the participants after the task, obtained by multiplying the percentage claimed by the total production of the pair. In the NOVEIL, the amount of tokens claimed is significantly higher than the amount of tokens earned with the task, for both participants with six minutes and participants with ten minutes (Wilcoxon signed-rank test, participants with six minutes: $z = 5.51$, $p < 0.01$; participants with ten minutes: $z = 3.99$, $p < 0.01$).

Result 2. Individual production and claims in the NOVEIL treatment.

In NOVEIL treatment participants' claims are significantly higher than their individual production, independently on the time available for the task.

This result is related to hypothesis 1. We certainly expected a better fit between working-time and claim (claims from subjects with six minutes should be approximately 60% of the claim of subjects with ten minutes), and less distance between production and claim –or less ‘stealing’ from partners. Although there is a difference between claims by ten-minute type subjects and six-minute type subjects, the data do not support our hypothesis. However, it must be said that the main purpose of treatment NOVEIL was to establish a baseline against which to check the effect of the cue introduced in treatment VEIL.

As for hypothesis 2, the implementation of a liberal egalitarian principle, aimed at reducing the inequality due to pure luck, would have resulted in participants with six minutes in the VEIL treatment asking more than those in the NOVEIL, and participants with ten minutes in the VEIL treatment asking less than those in the NOVEIL.

Looking at the data, we observe that claims by participants with six minutes in the two conditions are not significantly different (Wilcoxon rank-sum - Mann-Whitney test, on percentages: $z = 0.19$, $p = 0.85$; on number of tokens: $z = 0.10$, $p = 0.91$). As for participants with ten minutes, those in the VEIL ask *even more* than those in the NOVEIL (on percentages: $z = 2.19$, $p = 0.03$; on number of tokens: $z = 2.11$, $p = 0.03$).

Result 3. *Ex post* claims across treatments.

Participants with six minutes in the VEIL treatment make the same claim of those in the NOVEIL. Participants with ten minutes in the VEIL treatment ask for more than those in the NOVEIL.

Based on result 3 we reject hypothesis 2.

Table 4. Percentage claimed before the task in the VEIL treatment

	Tokens
Six minutes	73.50 (17.76)
Ten minutes	79.50 (16.78)

Means, standard deviations in parentheses.

In the VEIL treatment, before knowing the distribution of time within the pairs, subjects were asked to choose a percentage of the total production they may claim after the task. Table 4 reports the average percentage chosen by the subjects, distinguishing between those who later would have been assigned six minutes and those who would have been assigned ten minutes. We observe that while *ex post* claimed percentages by participants with six minutes (table 3, row 2, column 1) are not statistically different from their claims *ex ante* (Wilcoxon signed-rank test, participants with six minutes: $z = 0.26$, $p = 0.79$), *ex post* percentage claims of participants with ten minutes (table 3, row 2, column 2) are even higher than their claims *ex ante* ($z = 2.97$, $p = 0.03$).

Result 4. *Ex ante* and *ex post* claims in the VEIL treatment.

In the *ex post* phase of the VEIL treatment, participants with six minutes confirm their choice *ex ante*, while participants with ten minutes claim more than what they have judged as the right claim *ex ante*.

Table 5 reports the results of an OLS estimation in which we control for socio-demographic characteristics and participants' experience with experiments. The results reported in column 2 confirm that claims of the subjects with ten minutes in the VEIL are slightly higher than those of subjects with ten minutes in the NOVEIL. Interestingly, females claim less than male but only in the VEIL treatment, and subjects with less experience with experiments claim less than more experienced ones only in the NOVEIL treatment. The results in column 2 confirm the correlation between claim *ex ante* and claim *ex post* in the VEIL treatment, which is weaker for the subjects with ten minutes who tend to ask more than the *ex ante* claim with respect to the subjects with six minutes.

Table 5. Determinants of individual claims after the task

	(1) OVERALL	(2) VEIL TREATMENT
AGE	-0.161 (0.720)	0.138 (0.611)
GENDER	-0.902 (3.890)	-4.185* (2.485)
NATIONALITY	6.256 (8.206)	-2.880 (8.343)
MAJOR	-2.779 (2.313)	0.891 (2.269)
N. OF EXPERIMENTS	3.647*** (1.328)	0.318 (0.735)
TEN	1.400 (3.882)	30.08*** (11.35)
VEIL	10.98* (5.806)	
TEN*VEIL	10.31* (5.531)	
GENDER*VEIL	-7.932 (5.554)	
N. OF EXPERIMENTS*VEIL	-3.605** (1.755)	
CLAIM <i>EX ANTE</i>		0.876*** (0.101)
CLAIM <i>EX ANTE</i> * TEN		-0.289** (0.143)
CONSTANT	70.41*** (19.88)	9.028 (17.72)
Observations	160	80
R-squared	0.157	0.681

OLS regression.

The dependent variable is the percentage claimed after the task (in the case of the VEIL treatment it corresponds to the *ex post* choice)

Results in column 2 refer only to the subsample of subjects who took part in the VEIL treatment.

GENDER takes value 1 if the subject is a female and zero otherwise.

NATIONALITY takes value 1 if the subject is Spanish and zero otherwise.

MAJOR takes value 1 if the subject is enrolled in an economics or management program and zero otherwise.

N. OF EXPERIMENTS is the number of previous experiments in which the subject took part.

TEN takes value 1 if the subject is endowed a time of ten minutes and zero otherwise.

VEIL takes value 1 if the treatments VEIL and zero otherwise.

CLAIM *EX ANTE* is the percentage claimed in the *ex ante* phase of the VEIL treatment.

Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.10

In conclusion, our data support neither of our hypotheses.

IV. Summary and discussion

In the experiment reported here a majority of young students stole from an anonymous peer. They did this while knowing that the earnings that each had contributed to the total to be split was obtained through a work they performed – a work that while not particularly hard, was not entirely effortless. Even though the game was anonymous (they did not know from whom they were taking the money), it was a very homogeneous population of students. This fact would have made us predict a higher degree of empathy. According to Binmore (2005), this kind of preferences are used to find the fair solution to coordination problems. In Seabright's reading of *Natural Justice* (Binmore, 2005), we find condensed into a very effective phrase the function that empathetic preferences should play: "individuals must have empathetic preferences – when imagining themselves in the situation of others, not their own" (Seabright, 2005, p. 35). Furthermore, "morality serves to coordinate between the possible equilibria of social life" (Seabright, 2005, p. 36), so the moral rules internalized by evolution should help us to resolve potential situations of social inequality. The introduction of the veil, applied to an individual rather than a collective choice, should direct behaviour to a fair split, levelling to some extent the initial fortuitous disparities. Considering the role that internalize moral principles usually have in the process of deliberation, we hypothesized the prevalence of the social norms about effort and merit, the unwritten rules about earnings and possession; we hypothesized further that a reflection towards fairness, reinforced by the individual position of ignorance at the time of the thought-experiment, would help bending individual decisions towards equality. Despite the formulation of our working hypotheses, we were not particularly optimistic about the effect of these rules on the average behaviour of our experimental population, but we definitely expected at least some departure from standard results in DG with taking. In the NOVEIL treatment, the fact that income was 'earned' through a task may activate the social norms of merit and desert, and induce claims tracking individual's contributions. In the VEIL treatment, reflection on the unjustified inequality of resources may induce sympathy and the social norm of fair distribution for these circumstances – liberal-egalitarian principles. To our dismay, the only significant difference between baseline treatment and treatment VEIL was *contrary* to our hypothesis: subjects with ten minutes-time asked, on average, more than their counterparts in the baseline treatment. The rejection of our hypotheses left us baffled and wanting an explanation.

Let us note first that this result may be related to other experiments in which the most advantaged subjects were reluctant to give. These experiments usually involve effort and luck (Erkal *et al.*, 2011; Rey-Biel *et al.*, 2018; Konow, 2000). However, in our design, having more time to perform the task does not imply merit, effort or recognizing some specific abilities the player has. The veil of ignorance should have functioned to highlight that the initial disparity was not due to a difference in ability amongst subjects but to a merely external and random element. As stated by Erkal *et al.* (2011, p. 3332) "in real life, earnings are determined not only by effort, but also by luck. Evidence suggests that giving behavior might be different when luck affects earnings and that people are more likely to receive support when they have been negatively affected by luck (Christina M. Fong 2007)". Nev-

ertheless, when subjects believed that they have earned a certain reward resulting from an effort, considerations on fairness ideals and on how material goods should be redistributed would not have a significant impact.

We have reached the conclusion that the setting in which the game was made—marked by anonymity and non-iteration—created a context in which there was a lack of motivation to act in other-regarding ways. As Eckel and Grossman emphasize:

The decision makers cannot identify each other, nor do they have enough information to know if their partner is poor or otherwise deserving of their generosity; thus there is little or no basis for altruism to play a part in the decision. Furthermore, as subjects cannot be identified by either the experimenter or other subjects, there is no role for social esteem to affect the decision. Only self esteem (or warm glow) remains. With little motivation for other-regarding behavior, it is not surprising that the subjects' behavior closely approximates the game-theoretic predictions for noncooperative, non repeated games with selfish, payoff-maximizing subjects. (Eckel & Grossman, 1996, p. 182)

However, some features of our design could make us think of fairer behaviours: the effort needed for production—as opposed to the 'manna from heaven' situation common in other experiments—and the veil of ignorance at the individual level, proved completely irrelevant. The veil of ignorance should have had an effect in the final distribution. If everyone had an effective sensitivity to moral cues, the fact of allowing subjects to think in advance what percentage of their product to bring home and which one to offer should have fostered a fairer behaviour.

A first explanation could be related to a misunderstanding of the experiment by the subjects. However, this possibility was ruled out from the start: not only we did run check questions after reading the instructions, but the subjects were well versed in the workings of experimental sessions, and most of them had some training in economics or management. If anything, our subjects' 'mistake' was to grasp the experiment all too well. They saw the true nature of the choice at their disposal and were not influenced by a setting that involved working in pairs, or pooling together earnings, or making a distributive choice. They understood that at the end of the day, what they were going to get was determined simply by their choice—multiplied by the 0.5 chance of being actually selected for final payoff—and they responded wisely—in economic terms. They were, on average, very good monetary payoff maximizers.

The fact that our subjects were students, belonging to the same group of reference, could have caused gamesmen behaviour, rather than increased sympathy, because the context would not be perceived as a situation in which real interests are at stake. This interpretation is in line with the results of Hoffman *et al.* (1996): they saw that “when a double blind procedure, intended to guarantee the complete social isolation of the individual's decision (no one including the experimenter or any subsequent observer of the data could possibly know any subject's decision), was used, 64 percent of the offers were \$0 with only 8 percent offering \$4 or more” (Hoffman *et al.*, 1996, pp. 653-54).

On a related line, Fehr *et al.* (2006) found that students attending majors in Economics and Management are less sensitive to egalitarian concerns. However, this would not explain the surprising result that our ten-minute subjects claimed on average *more* in the treatment with a veil. We were aware of this possible effect, so we were not surprised by the rate of 'stealing' in the baseline treatment. But the null effect of the individual reflection behind the veil still wants an explanation.

Another possible reason for the observed proficiency in maximizing behaviour combined with their utter disregard for social norms or lack of pro-social preferences might be that the subjects were recruited over a platform Orsee: people included in the platform were ‘used’ to do economic experiments, so it may be thought that they are sophisticated economic choice-makers. However, we have excluded also this explanation because the recruitment of experimental subjects through registration on platforms is a widespread and established method to conduct experiments in the laboratory.

One potential explanation might have to do with the problem of stability of intentions, as it applied to the stability of the principles chosen behind the veil. The subjects might have been sensitive to the moral cue, and correctly concluded that the final distribution had to be re-distributive, and still fail to follow this conclusion in practice (McClennen, 1989; Klosko, 1994; Barry, 1995). There is no logical implication that choosing a distribution criterion behind the veil will bring about a corresponding behaviour. The problem of stability can be read as the problem of compliance: why should subjects act according to a certain ideal of fairness—even if it is an ideal they chose as their preferred option before—, when they could obtain a greater gain by deviating?

In our case, there is little difference between *ex ante* claim and *ex post* decision; so that the *ex ante* choice can hardly be identified as an ideal to follow. Leaving this aside, there are two reasons why this line of reflection would not be wholly satisfactory as an explanation. First, our design excludes the interpretation of the *ex ante* choice as a plan. Our subjects know from the beginning that their final choice is up to them and that they will decide only after they know all the facts (whether they had six or ten minutes, how much the pair produced, etc.). So, there is nothing in the design implying that their first choice is a kind of plan they have to follow or rule they have to abide by. The initial choice should work as a mere moral cue, that is, as a reminder of the fact that there was an unjustified inequality, and that they had the opportunity to re-dress for it.

Second, the issue of stability of principles is quite complex, involving psychological conditions for the stability of intentions and conceptions of rationality and morality. The contribution of this paper is to clarify one aspect of that dynamic, and the experiment abstracts as much as possible of the deeper philosophical issues related to the problem of stability. Our hypothesis tried to relate the veil of ignorance as moral cue with an adjustment of individual behaviour in a prosocial and egalitarian direction.

The difficulties with this explanation do not imply that our result is irrelevant to the debate of stability. After all, the experiment shows that whatever thoughts are derived from reasoning behind the veil, they do not carry over to the real choice situation. But from our data we cannot conclude whether this is a problem of stability or a problem of irrelevance: we cannot know whether our subjects correctly concluded that initial inequality called for re-dress but then failed to make a re-distributive choice (either *ex ante* or *ex post*), or they, also correctly, found that reasoning behind the veil was irrelevant in the situation.

Upon reflection, the puzzling fact of stealing from peers should not be surprising. Note that the final distribution of money (how much one gets) depends on three factors: effort (own, and one’s pair), decision, and chance (being selected for final payoff). After all, this is what gambling is—think of State lottery: working people voluntarily bet part of their legitimate earnings and subject themselves to a chance event. They know that the most likely outcome is that someone else will get that money. And in the event that they win, they experience no remorse in taking money from other working people, maybe poorer than

them. And this is because the whole practice is voluntary in the first place and based on free choices. Our subjects might have approached the experiment in this ‘gambling mood’ and this would explain the result.

The problem is that this explanation would be applicable to virtually all economic experiments involving populations of students with some previous experiments in their records.

It might be the case that, by making the experiment one-shot and anonymous dictator’s choice —no chance of reciprocity, no need to regard the other member of the pair’s attitudes, beliefs or dispositions— with the conspicuous absence of punishment by any kind of authority, we gave our subjects a true Gige’s ring to act as they pleased (Plato, 2000, 359d). And our finding is that they behaved as Trasymachus and Hobbes would have predicted, rather than the way Socrates, Kant or Rawls would.

In conclusion, two interesting elements emerge from this study: the effectiveness of the veil of ignorance and the actual behavioural adherence to moral principles given the social dimension in which we are immersed. We wish to comment on each in turn.

Regarding the first point, this study shows that the veil of ignorance, as moral device, has a low effect on individual choice when used as counter-factual *ex ante* thought experiment. This is in contrast with experiments where the choice behind the veil was the result of an actual agreement among parties (Sacconi & Faillo, 2010; Schildberg-Hörisch, 2010; Faillo *et al.*, 2015). In these cases, impartiality and impersonality —conveyed from the veil— had an impact on participants’ behaviour. So, it would be interesting to study the conditions under which the veil might work, inducing pro-social behaviour, on individual choice – there are many occasions in our everyday life when we have to decide how to behave, morally or not, without prior agreements. In general, moral philosophy tends to make use of counter-factual reasoning. It may be argued that counter-factual arm-chair philosophy is not intended to *motivate* action, but to justify, or explain. However, criticism about the lack of motivating power was taken very seriously by leading contractarian philosophers, like John Rawls or David Gauthier. While acknowledging that their theories, being rationalistic and laid out in analytic language, were not particularly enticing, they did defend that, properly understood and once the distance between ideal theory and non-ideal social conditions are taken into account, they should give people reasons to act as they prescribe. We assumed that reasoning behind the veil is a kind of toy moral theory for a particular case. Given the setting of the experiment, rational people should get to the conclusion that equal —or approximately equal— split would be more justified. They generally did reach this conclusion. Now, the fact that they did not take this conclusion as a sufficient reason, while actual agreements reached behind the veil do seem to create new normative reasons for action, is enlightening. It is a result that has potential interest in institutional settings. It speaks of the relevance of the social in inducing real cognitive and motivational changes in individual agents.

Secondly, these results might contribute in defying how to distinguish social and moral norms, when such norms are collectively recognized, what they include and how they are perceived by the decision makers. Although far from our original intent, what emerged seems to be that moral reflections are not so binding on behaviour. Observability, and the consequent awareness resulting from a choice made in public, is the fundamental feature present in a social dynamic, where the risk of social punishment or exclusion is high. Not only that, observability allows each subject to show himself in a certain way —appearing right rather than being right— because in social contexts we often want to give a certain image of ourselves and then do not betray it, neither in the eyes of others nor in our own.

This would be a confirmation of how social norms are behavioural rules that arise from conditional preferences in certain contexts. Empirical and normative expectations are essential ingredients for deciding to follow a social norm and to comply with it, even when a decision is the result of an individual deliberation, not influenced by any agreements. Without a social environment in which these conditions are created, moral motivations, potentially existing in *foro interno*, do not apply externally (Bicchieri, 2006).

Following this line, we may say that a moral cue of this type results to be much less effective than a minimal social cue as the one proposed by Rigdon and colleagues (2009). In that case, presenting a very simple watching-eyes configuration had an impact on giving behaviour, while our findings show how inducing a reflection from an impartial point of view is not enough under conditions of total anonymity and of individual choice. It seems that moral cues have an effect on human behaviour only if they are already proposed as ethical principles. For example, reading the Ten Commandments before an experiment in total anonymity had a real weight in leading subjects' choices towards honest behaviour (Gino *et al.*, 2012). Unlike this type of reminder, our findings show that providing time to make subjects reflect from an impartial point of view does not have the same cognitive and motivational force: it seems much easier to take care of moral principles when someone else makes us remember them, instead of finding the ethical answer by our own.

Despite our hypotheses were rejected, this study might contribute to better understand how moral cues can intervene in the decision-making process and to remember how important the distinction between social and moral norms is, both in the study of prosocial behaviour and in the relationship between motivation and action.

Acknowledgement

Funding for the research reported in this paper was provided by the Spanish Ministry of Economy through Research Grants FFI2011-29005 and FFI2014-56391P.

REFERENCES

- Aquino, K., & Reed A. (2002). The Self-Importance of Moral Identity. *Journal of Personality and Social Psychology*, 83(6), 1423-1440.
- Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three Principles to REVISE People's Unethical Behavior. *Perspectives on Psychological Science*, 10(6), 738-741.
- Barkan, R., Ayala, S., & Ariely, D. (2015). Ethical dissonance, justifications, and moral behaviour. *Current Opinion in Psychology*, 6, 157-161.
- Bardsley, N. (2008). Dictator game giving: Altruism or artefact?. *Experimental Economics*, 11(2), 122-133.
- Barry, B. (1995). John Rawls and the Search for Stability. *Ethics*, 105(4), 874-915.
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.
- Binmore, K. (2005). *Natural Justice*. Oxford: Oxford University Press.
- Cappelen, A.W., Nielsen, U.H., Sørensen, E.Ø., Tungodden, B., & Tyran, J.R. (2013). Give and take in dictator games. *Economics Letters*, 118(2), 280-283.
- Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior and Organization*, 149, 74-87.

- Degli Antoni, G., Faillo, M., Francés-Gómez, P., & Sacconi, L. (2016). Distributive Justice with Production and the Social Contract. An Experimental study. *Economia Etica*, 60, 1-51.
- Eckel, C., & Grossman, P. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior*, 16, 181.
- Erkal, N., Gangadharan, L., & Nikiforakis, N. (2011). Relative Earnings and Giving in a Real-Effort Experiment. *The American Economic Review*, 101(7), 3330-3348.
- Faillo, M., Rizzolli, M., & Tontrup, S. (2019). Thou shalt not steal (from hard-working people): An experiment on respect for property claims. *Journal of Economic Psychology*, 71, 88-101.
- Faillo, M., Ottone, S., & Sacconi, L. (2015). The social contract in the laboratory. An experimental analysis of self-enforcing impartial agreements. *Public Choice*, 163(3-4), 225-246.
- Fehr, E., Naef, M., & Schmidt, K.M. (2006). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment. *The American Economic Review*, 96(5), 1912-1917.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10, 171-178.
- Franzen, A., & Pointner, S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior & Organization*, 81, 74-81.
- Frohlich, N., Oppenheimer, J.A., & Eavey, C.L. (1987). Laboratory Results on Rawls's Distributive Justice. *British Journal of Political Science*, 17(1), 1-21.
- Frohlich, N., & Oppenheimer, J.A. (1990). Choosing Justice in Experimental Democracies with Production. *American Political Science Review*, 84(2), 461-477.
- Frohlich, N., & Oppenheimer, J.A. (1992). *Choosing Justice: An Experimental Approach to Ethical Theory*. University of California Press.
- Gino, F., Ayal, S., & Ariely, D. (2012). Self-Serving Altruism? When Unethical Actions That Benefit Others Do Not Trigger Guilt. *Harvard Business School Working Paper*, No. 13-028.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-125.
- Herne, K., & Suojanen, M. (2004). The Role of Information in Choices Over Income Distributions. *Journal of Conflict Resolution*, 48(2), 173-193.
- Hobbes, T. (2011). *Leviatano*. (G. Micheli Trans.). Milano: BUR Biblioteca Univ. Rizzoli.
- Hoffman, E., McCabe, K., & Smith, V. (1996). Social Distance and Other-Regarding Behavior in Dictator Games. *The American Economic Review*, 86(3), 653-660.
- Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7(3), 346-380.
- Khadjavi, M. (2015). On the interaction of deterrence and emotions. *Journal of Law Economics and Organization*, 31(2), 287-319.
- Kessler, J.B., & Leider, S. (2012). Norms and contracting. *Management Science*, 58(1), 62-77.
- Klosko, G. (1994). Rawls's Argument from Political Stability. *Columbia Law Review*, 94(6), 1882-1897.
- Konow, J. (2000). Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *The American Economic Review*, 90(4), 1072-1091.
- Korenok, O., Millner, E.L., & Razzolini, L. (2014). Taking, giving, and impure altruism in dictator games. *Experimental Economics*, 17(3), 488-500.
- Krupka, E.L., & Weber, R.A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary?. *Journal of the European Economic Association*, 11(3), 495-524.
- Levitt, S.D., & List, J.A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?. *Journal of Economic Perspectives*, 21(2), 153-174.
- Lissowski, G., Tyszka, T., & Okrasa, W. (1991). Principles of Distributive Justice: Experiments in Poland and America. *Journal of Conflict Resolution*, 35(1), 98-119.
- List, J.A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3), 482-493.
- Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45, 633-644.

- McClennen, E. (1989). Justice and the Problem of Stability, *Philosophy & Public Affairs*, 18(1), 3-30.
- Michelbach, P.A., Scott, J.T., Matland, R.E., & Bornstein, B.H. (2003). Doing Rawls Justice: An Experimental Study of Income Distribution Norms. *American Journal of Political Science*, 47 (3), 523-539.
- Plato (2000). *The Republic*. (G.R.F. Ferrari, Ed.; T. Griffin, Trans.). Cambridge: Cambridge University Press.
- Rawls, J. (1980). Kantian Constructivism in Moral Theory. *The Journal of Philosophy*, 77(9), 515-572.
- Rey-Biel, P., Sheremeta, R., & Uler, N. (2018). When Income Depends on Performance and Luck: The Effects of Culture and Information on Giving. *MPRA Paper*, No. 83940.
- Rigdon, M., Ishii, K., Watabe, M., & Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology*, 30, 358-367.
- Sacconi, L., & Faillo, M. (2010). Conformity, reciprocity and the sense of justice. How social contract-based preferences and beliefs explain norm compliance: the experimental evidence. *Constitutional Political Economy*, 21, 171-201.
- Sacconi, L., Faillo, M., & Ottone, S. (2011). Contractarian Compliance and the 'Sense of Justice': A Behavioral Conformity Model and Its Experimental Support. *Analyse & Kritik*, 33(1), 273-310.
- Scott, J.T., Matland, R.E., Michelbach, P.A., & Bornstein, B.H. (2001). Just Deserts: An Experimental Approach to Distributive Justice. *American Journal of Political Science*, 45(3), 749-67.
- Seabright, P. (2005). The evolution of fairness norms: an essay on Ken Binmore's *Natural Justice*. *Politics, Philosophy & Economics*, 5(1), 33-50.
- Servátka, M. (2009). Separating reputation, social influence, and identification effects in a dictator game. *European Economic Review*, 53(2), 197-209.
- Schram, A., & Charness, G. (2012). Social and Moral Norms in the Laboratory. *eScholarship, UC Santa Barbara, Departmental Working Papers*, 1-33.
- Schildberg-Hörisch, H. (2010). Is the veil of ignorance only a concept about risk? An experiment. *Journal of Public Economics*, 94, 1062-1066.
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-Serving Justifications: Doing Wrong and Feeling Moral. *Current Directions in Psychological Science*, 24(2), 125-130.

Laura Marcon is a PhD student at the University of Granada (Spain), in a co-tutelle thesis with the Doctoral School of Social Sciences, Doctoral program in Economics and Management at the University of Trento (Italy). Her doctoral thesis fits in the doctoral line of research "Normativity and rationality: organizations and public policy". Her main research interests are moral philosophy, experimental philosophy, moral psychology, behavioural economics. She was a short-term scholar from August to December 2018 in the Philosophy, Politics and Economics program at the University of Pennsylvania, under the supervision of Professor Cristina Bicchieri.

ADDRESS: Department of Philosophy I, University of Granada. Campus de la Cartuja, calle del Prof. Clavera, s/n (18011 Granada). E-mail: marcon4@ugr.es ; lamarcon@correo.ugr.es

Pedro Francés-Gómez is Professor of Moral Philosophy at the University of Granada. His main research interest is Social Contract Theory applied to business and organizational ethics. He has published in specialized international journals in the fields of Business Ethics and Political Philosophy: *Business Ethics: A European Review*; *Organizational Change Management*; *Journal of Business Ethics*; *Hobbes Studies*; *Telos*; *Revista Internacional de Filosofía Política*, *Isegoría*, and others.

ADDRESS: Department of Philosophy I, University of Granada. Campus de la Cartuja, Calle del Prof. Clavera, s/n (18011 Granada). E-mail: pfg@ugr.es

Marco Faillo is Associate professor, Economic Policy, Department of Economics and Management of the University of Trento, Italy. Since August 20th, 2018: National scientific qualification as Full professor of Economic Policy. He is external member of Institute of Economics of the Sant'Anna School of Advanced studies, Pisa, Italy. His main research interests are economics of institutions and organizations; behavioural and experimental economics; bounded rationality; ethics and economics.

ADDRESS: Department of Economics and Management, University of Trento. Via Inama 5. 38122 Trento (Italy). E-mail: marco.faillo@unitn.it