# Moral Hypocrisy, Power and Social Preferences

Aldo Rustichini
Marie Claire Villeval

May 2012

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Moral Hypocrisy, Power and
# Social Preferences

**Aldo Rustichini**
*University of Minnesota*


**Marie Claire Villeval**
*University of Lyon, CNRS-GATE
and IZA*

# ABSTRACT

# Moral Hypocrisy, Power and Social Preferences[*]

We show with a laboratory experiment that individuals adjust their moral principles to the situation and to their actions, just as much as they adjust their actions to their principles. We first elicit the individuals' principles regarding the fairness and unfairness of allocations in three different scenarios (a Dictator game, an Ultimatum game, and a Trust game). One week later, the same individuals are invited to play those same games with monetary compensation. Finally in the same session we elicit again their principles regarding the fairness and unfairness of allocations in the same three scenarios. Our results show that individuals adjust abstract norms to fit the game, their role and the choices they made. First, norms that appear abstract and universal take into account the bargaining power of the two sides. The strong side bends the norm in its favor and the weak side agrees: Stated fairness is a compromise with power. Second, in most situations, individuals adjust the range of fair shares after playing the game for real money compared with their initial statement. Third, the discrepancy between hypothetical and real behavior is larger in games where real choices have no strategic consequence (Dictator game and second mover in Trust game) than in those where they do (Ultimatum game). Finally the adjustment of principles to actions is mainly the fact of individuals who behave more selfishly and who have a stronger bargaining power. The moral hypocrisy displayed (measured by the discrepancy between statements and actions chosen followed by an adjustment of principles to actions) appears produced by the attempt, not necessarily conscious, to strike a balance between self-image and immediate convenience.

Corresponding author:

Marie Claire Villeval
GATE (Groupe d'Analyse et de Théorie Economique)
CNRS - University of Lyon
93, Chemin des Mouilles
69130 Ecully
France
E-mail: villeval@gate.cnrs.fr

---

I. INTRODUCTION

Most individuals promote principles of behavior based on fairness and selflessness and derive utility from being perceived as generous and fair towards others (Benabou and Tirole, 2006; Andreoni and Bernheim, 2009). Parents teach their children that they should behave respectfully and kindly toward other persons; politicians emphasize dedication to serving others; businessmen promote corporate social responsibility. Real behavior may differ and a darker side of human nature can be revealed. Individuals destroy the resources of others because of envy (Mui, 1995; Maher, 2010; Charness *et al.*, 2010; Harbring and Irlensbusch, 2011) or for the joy of destruction (Zizzo and Oswald, 2001; Abbink and Sadrieh, 2009); the power of public office sometimes leads politicians to use it for their personal gain (Aidt, 2003); feelings of entitlement push leaders to take more than followers from a common resource (de Cremer and van Dijk, 2005). How do people form their ideas of fairness, and how do they reconcile their stated norms of fairness and the temptation of more selfish actions that may alter their perception as fair people? In this paper, we study how individuals try to maintain consistency between stated principles and real actions, and how they take into account the current situation to adjust principles that appear universal. We test how much they maintain their image of fairness by adjusting their actions to suit their fairness principles, and how much instead they adjust their principles to justify their actions. The study of this interaction is the study of moral hypocrisy (Batson et al., 1997, 1999), defined as the motivation to appear moral, to himself and to others, without acting according to the moral prescriptions. Since one of the people that hypocritical behavior wants to deceive may be the individual who produces that behavior, hypocrisy is not necessarily a conscious deception. But we also test whether a position of relative advantage (as the ability to propose an ultimatum) affects principles that should not, by their very nature of universal rules, depend on the position.

Our analysis requires a new approach to identify the individuals' fairness principles separately from the observation of actions. Experimental analysis of strategic choices claims to detect moral principles through the observation of the decisions made by the individuals. These norms and social preferences are posited and inferred indirectly from behavior when the latter differs from the equilibrium play determined under selfish preferences. The assumption is that people do not play selfishly because they are guided by some fairness principles (Ostrom, 2000), which may be driven by distributional concerns (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), a concern for increasing social welfare (Charness and Rabin, 2003), or conditional cooperation (Fischbacher and Gächter, 2012). These principles are identified indirectly and thus possibly in an incomplete or biased way. In addition, by

construction a single decision is usually observed.[1] Assuming that this decision delimitates a single value, below which all the other decisions would have been considered by the individual as being unfair and above which all decisions are judged fair, would be arbitrary. Thus, this approach delivers a partial vision of individuals' principles and does not permit to understand the dynamics between principles and actions.

When other-regarding preferences are explicitly elicited in the literature, they are then not compared with the corresponding actions to check whether the individuals' actions comply with their principles. In these experiments individuals are asked to assess the action of other players, not theirs. In Falk and Fischbacher (2006) individuals have to rate the kindness of various hypothetical divisions of an endowment made by another player in a bilateral exchange.[2] However, social psychologists studying moral hypocrisy have shown that individuals evaluate more negatively the moral transgression of fair principles when this transgression is enacted by others than when enacted by themselves (Valdesolo and deStefano, 2008). Therefore, it is unlikely that the principles stated on the fairness of others' actions also define the benchmark used by individuals to evaluate their own actions. Krupka and Weber (2008) elicit *a priori* social norms using both ratings of social appropriateness of actions and incentive compatible coordination games and use them to predict behavior in dictator games. Individuals have to rate the appropriateness of behavior according to their beliefs about what others consider as appropriate. Focusing on socially shared norms and not on personal principles, this technique enables studying norm compliance but not the strategies developed by individuals to cope with possible inconsistencies between their own principles and actions.

Differently from these approaches, we adopt the methodological view that both actions and statements of principles constitute social behavior. Actions determine immediate rewards; actions together with statements also modify the opinion of others. Individuals correctly anticipate that actions and statements matter. They strike a compromise between monetary consequences and reputation, understanding that statements are judged in view of actions, and actions interpreted in view of statements. Thus, to understand social behavior we need to measure both. From the point of view of the experimental design, the originality of our paper is threefold. First, we elicit directly the individuals' stated principles of fairness and unfairness, in two steps: initially behind the veil of ignorance and

---

[1] This remark does not apply to Fichbacher and Gächter (2012) who ask players to choose an action for each possible action of the other group members in a public good game. This procedure enables measuring social preferences through a set of actions. But it is not intended to identify whether individuals have a conscious perception of their fairness principles.
[2] Similarly, in Rabin (1993)'s theory an action is perceived as fair if it reveals the good intention of another player.

then after individuals know their role in two-player games. We define the individuals' stated principles as the set of shares they consider personally as being fair and the set of shares they consider as being unfair. This allows us to define intervals of fairness and unfairness that do not have to complement each other perfectly and that indicate the wiggle room people allow to themselves. By focusing on personal norms more than on social norms, we allow discrepancies between statements and action and we analyze how players cope with them. Second, we allow these statements to differ according to whether one adopts the point of view of an advantaged player or of a disadvantaged player. By not imposing that norms are perceptions of what fair behavior is behind the veil of ignorance, we can analyze whether the fairness principles are conditioned by the power of the individual. Third, our paper aims at analyzing the links between stated principles and actions, which may be complex because actions may follow principles but principles may also adjust to the selected actions. Indeed, self-justification can lead individuals to modify their principles regarding what is fair and what is unfair after having made selfish decisions. To this aim, we explicitly measure principles at various moments in time, i.e. both before and after actions are taken, and we relate the variation of these stated principles to the individuals' actions in similar situations.

In our experiment subjects participated in two consecutive sessions. In the first session we elicited the individuals' statements regarding the fairness and unfairness of all possible shares between two hypothetical players in three different scenarios. This elicitation is done behind the veil of ignorance and from the two players' perspectives. The scenarios correspond to a Dictator game, an Ultimatum game, and a Trust game. They also make a hypothetical – not incentivized- decision in each of these scenarios. At least one week later, the same individuals are invited to play a Dictator game, an Ultimatum game and a Trust game for real. In the final part of the second session, we elicit again their principles regarding the fairness and unfairness of all possible shares in the same three scenarios. This design allows us to measure both i) how the actions in the second session diverge from the statements enounced during the first session and ii) whether the principles enounced during the second session conform more to the initial statements or to the actions taken in the same session. In addition, we are able to determine whether actions respect more the principles initially stated when the other player is not passive (i.e. in the Ultimatum game) than when he cannot react (i.e. in the Dictator and the Trust games).

We identify two main channels through which environment affect norms: moral hypocrisy and power. The violation by later actions[3] of the fairness principles initially stated behind the veil of ignorance provides the first evidence of moral hypocrisy. Since we observe whether statements are revised after the choice of an action to fit this action, we have the additional proof provided by the difference between initial and revised principles. In most situations, individuals increase the range of shares acceptable as fair after playing the game for real money compared with their initial statement. The discrepancy between hypothetical and real behaviors is larger in games where real behavior has no strategic consequence (Dictator and Trust games) than in games where the other player can react to the decisions (Ultimatum game).

By using different games that vary the power of the players (i.e. their ability to impose their preferred outcome), we can also study whether the adjustment of principles to the situation is influenced by relative power. We find that both sides, the one in the advantageous and disadvantageous position, do, by accepting the fact that allocations are biased in favor of the powerful. [4]

Moral hypocrisy is not systematic: those who adjust the most their principles to their action are those who behave more selfishly and those who are more powerful. These findings confirm that although it is rational to take selfish decisions from an economic point of view, individuals care about their self-image. By being hypocritical (i.e., by acting in violation of initial statements and adjusting statements of fairness downwards), they both pursue their self-interest and try to keep up appearances of pro-social motivations. This is an indirect confirmation of the importance of moral norms and aspirations in our societies.

Our design and data do not allow us to test two alternative possible views: conscious or unconscious hypocrisy. The first suggests that hypocrisy is a conscious attempt to claim a moral norm as universal and at the same time to violate the norm. The second is that hypocrisy is based first of all on self-deception, and individuals are not fully aware of the contradiction between principles and actions. In particular, we chose not to remind subjects, at the beginning of the second session, what their statements had been a week earlier. This design strategy leaves the door open for unconscious hypocrisy to manifest itself.

---

[3] This corresponds to the notion of "decoupling" in the theory of hypocrisy in organisations (Meyer and Rowan, 1977, 1978).
[4] In political theory, the link between power and hypocrisy is discussed in Runciman (2008).

The remainder of this paper is organized as follows. Section II gives a brief overview of the related literature. Section III describes the experimental design and procedures. Section IV presents the hypotheses we test, and Section V develops the results of the test. Section VI discusses our results and concludes.

II.  RELATED LITERATURE

Moral hypocrisy has been so far underexplored in economics although the analysis of social preferences has been blooming in the last decade.[5] A somewhat related approach is the analysis of strategic choice of ignorance as an excuse to selfish behavior in moral dilemmas: individuals avoid information susceptible to reveal harmful consequences of their action on others because it would affect their social image negatively.[6] Such an illusory preference for fairness has been identified by Dana, Weber and Kuang (2007) (see also Larson and Capra, 2009; Grossman, 2010; van der Weele, 2012). Indeed, fairness decreases substantially when the link between fairness and outcome is obfuscated. The choice to play fair is frequently motivated by the willingness to appear fair more than by the willingness to produce a fair outcome and this is why greater anonymity leads to more selfish transfers in the dictator game (Andreoni and Bernheim, 2009; Ariely *et al.*, 2009). Our approach is different in that individuals in our experiment cannot strategically ignore the consequences of selfish actions. We study instead whether they strategically adjust their principles of fairness to fit their action to preserve their self-image.[7]

Social psychologists have studied moral hypocrisy for a while. The notion of moral hypocrisy has been conceptualized by Batson et al. (1997, 1999, 2002) to explain the discrepancies between the norms held by individuals and their actions (see also Stone et al., 1997; Watson, Teague and Papamarcos, 2006). In this view, individuals want to appear moral without bearing the cost of morality. An implication of moral hypocrisy is the fact that the individuals' evaluation of their own transgressions of norms differs from that of transgressions by others (Valdesolo and DeSteno, 2007, 2008). The underlying mechanism is the individuals' general tendency to relieve

---

[5]  As we focus here on social preferences, we ignore the analysis of hypocrisy within organizations motivated by the search of a favor (Robin, Rusinowska and Villeval, 2012). Hypocrisy in organizations has been investigated more frequently in political sciences than in economics, notably regarding international organizations (Krasner, 1999a, 1999b; Lipson, 2011), and in the sociology of organizations where it results from the diversity of norm systems leading to inconsistencies between managers' talks and decisions (Brunsson, 1989, 1995, 2003; Meyer and Rowan, 1977, 1978; Huzzard and Östergren, 2002).

[6] On the importance of social image in economic models, see Bernheim (1994), Benabou and Tirole (2006), Ellingsen and Johanesson (2008).

[7] We focus more here on self-image than on social image as what is observed by others is the action and not the initial or revised fairness principles. On self-image and self-signaling, see notably Bodner and Prelec (2003) and Benabou and Tirole (2006). On an attempt to isolate self- from social signaling, see Grossman (2010).

themselves of their responsibility (Bandura, 1990, 1996). Self-deception (an "active misrepresentation of reality to the conscious mind" according to von Hippel and Trivers, 2011) is also involved in this disengagement of self-regulatory mechanisms: individuals do not adjust their actions to their principles, but they try to convince themselves that serving their own interests does not violate their principles (Trivers, 1985, 2000, 2009, 2011). Identifying hypocrisy in our setup is easier than in these experiments in social psychology because, rather than by a general statement of principles imposed exogenously, a moral criterion here is measured by an observable stated admissible interval of transfers to another player. Moreover, our design leaves the participants with no possibility to relieve themselves from the responsibility of their actions. Finally, we focus on the adjustment of principles after actions have been undertaken.

### III. THE EXPERIMENT

*III.1. Experimental design*

As we mentioned, the experiment consists of two consecutive sessions, both using three different scenarios. We use a within-subject design to observe the stated principles and actions of the same individuals placed both in strategic and non strategic situations.

*Scenarios and elicitation of principles of fairness and unfairness*

The first scenario corresponds to the Dictator game.[8] The second scenario corresponds to the Ultimatum game.[9] The third scenario corresponds to the Trust game.[10] We use these scenarios to elicit the participants' principles regarding the fairness and unfairness of various transfers in session 1 and at the end of session 2. Precisely, in the first session,

---

[8] The participant A receives an endowment of 10 points and he has to decide how many points, between 0 and 10, he is willing to transfer to B. B is passive. A earns the difference between his initial endowment and his transfer to B, while B earns the points transferred by A.

[9] A receives an endowment of 10 points and he makes an offer indicating how many points, between 0 and 10, he is willing to transfer to B. B has to decide on whether he accepts or he rejects A's offer. If his offer is accepted, A earns the difference between his initial endowment and his transfer to B, while B earns the points transferred by A. If A's offer is rejected, both participants earn nothing.

[10] Both A and B receive an initial endowment of 5 points. A decides how many points, between 0 and 5, he is willing to transfer to B. The amount transferred is tripled by the experimenter. B decides how many points he is willing to send back to A, between 0 and three times the amount sent by A plus 5 points. A earns the difference between his initial endowment and his transfer to B plus the amount sent back by B, while B earns the difference between his initial endowment augmented by three times the amount transferred by A minus the points sent back to A.

we ask for a statement on what they consider fair and unfair in both roles[11] and a hypothetical choice (the transfer they would make if they had to decide in a real game). In the Ultimatum and Dictator games participants evaluate the fairness and unfairness of the allocations made by the sender, while in the Trust game, they evaluate the fairness and unfairness of transfers by the *second mover*, conditional on two hypothetical initial transfers of the first mover. We then use the corresponding games to observe the participants' actions in session 2.

In part 1 (Dictator game) of the first session, the participants are requested to put themselves in the position of participant A. They have to evaluate the fairness of all possible shares transferred by A to B, then the unfairness of all possible shares. Then we ask what choice they would make if involved in a real game. After that, we ask them to consider the situation from the point of view of B, the passive player, and to evaluate the fair and unfair shares. In part 2 (Ultimatum game) of the first session, the structure of the decisions is similar. Participants have to report the sets of shares they consider as fair or unfair, in the shoes of the participant A, then in the position of B; they also make a hypothetical choice. In part 3 (Trust game) of the first session participants are presented two Trust game scenarios. In the first scenario, A transfers 1 point out of 5 to B and keeps 4 points for himself; thus, B can send back between 0 and 8 points to A. In the second scenario, A transfers 4 points out of 5 to B and keeps 1 point for himself; thus, B can send back between 0 and 17 points to A. Participants have to judge the fairness and unfairness of all possible amounts sent back by B to A, first in the position of player B and next in the position of player A. They also make a hypothetical choice in each scenario as a player B.

In the second session, participants play the Dictator game (part 1), the Ultimatum game (part 2) and the Trust game (part 3) for real, before providing evaluation of principles in the three corresponding scenarios (part 4).[12] Part 4 replicates exactly the first three parts of the first session, this time without asking the hypothetical choice. Roles in the game are assigned randomly. The Ultimatum game is played with the strategy method. The B participants are not told the actual choice of their co-participant A before the end of the session. So, they have to

---

[11] After stating their own principles regarding the fair and unfair shares, the participants were also requested to indicate which shares they believe most people consider as being fair and unfair and which share they believe most people would transfer in the same position. We do not report these beliefs in this paper because our analysis focuses here on the correlation between statements participants make for themselves and their hypothetical and real behavior.

[12] The order between the three scenarios has been kept constant across sessions. In session 1, this order ensures the same progressive increase in the degree of complexity for all the participants. At the beginning of part 4 in session 2, the three scenarios are reminded all together, so the order of each scenario should not matter. It should also be noted that in session 2, principles are elicited after the three games have been played for real and not after each game has been played. This is designed to avoid that after the first game, participants adjust their decisions in the next games knowing that they will also have to report their principles.

decide on whether accepting or rejecting each possible offer made by A. Similarly, the Trust game is played under the strategy method. The participants B are not told the amount actually sent by their co-participant A. Thus they have to decide how many points they are willing to send back to A for each possible amount sent by him. These games are played one-shot with a random re-matching of participants and roles (A or B) after each game.

The two sessions were separated by at least one week since we did not want the participants to remember precisely their stated principles before making their decisions in session 2 and to feel committed to follow these principles due to a consistency bias induced by the experimental design.[13] This is also the reason why we did not remind the participants their initial principles at the beginning of session 2.

*Elicitation of principles technique*

As we mentioned, we ask subjects to provide explicit statements of what they consider fair allocations, but let them communicate intervals rather than values. These reported fairness and unfairness intervals constitute what we call "stated principles". There are several reasons for this design. First, individuals may not view a fair share as a single precise value: more likely they think there is a range of acceptable values. Even if they do have a single number in mind, social interactions never force us to precise numerical statements, but leave us with the discretion of vagueness. We let our participants use this strategic uncertainty. We ask them to determine, in the position of participant A, the set of what they consider fair shares transferred to B. The specific question was *"What do you consider as being fair shares between A and B?"[14]*. On the computer screen a bar with two cursors, graduated from 0 to 100%, and a box detailing the choices are displayed (see two examples in figure 1). Moving the left cursor indicates the minimum fraction going to player B that is considered fair; moving the right cursor indicates the maximum value. This method enables to determine *i*) whether the midpoint of the fairness band is 50%, and *ii*) how large is the fairness band, its width indicating the wriggle room that individuals allow to themselves.

---

[13] From that perspective, we could have chosen a longer time interval between the two sessions, but this could have increased the risk of losing participants and it could have created some dissatisfaction since total payment was made only at the end of the second session.

[14] The French original text of the Instructions is "*Qu'est-ce que vous considérez comme des partages justes entre A et B ?*". We did not retain the alternative word "*équitable*" that evokes equality and is probably less neutral.

I consider as being fair all the shares in which A transfers to B at least 28% and at most 89% of the amount.

I consider as being fair all the shares in which A transfers to B at least 43% and at most 53% of the amount.
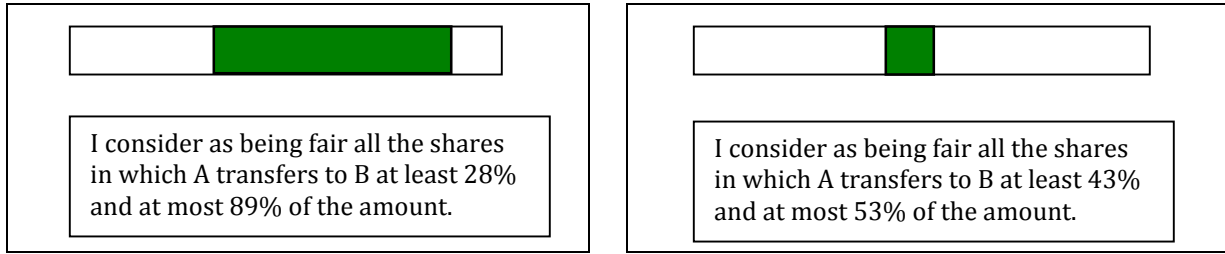
Figure 1. Determination of the sets of fair shares - Two examples of players' screens

Next, the participants (still in the position of player A) are requested to evaluate the unfairness of the shares transferred by A to B in the same scenario.[15] Another bar, graduated from 0 to 100%, and a box detailing the choices are displayed on the screen (see two examples in figure 2). By moving the left cursor the participants indicate the share below which shares are considered as unfair, and by moving the right cursor they indicate the share above which shares are considered as unfair. This method has the same virtues as for the determination of fairness principles. Note that there may be some overlapping between principles of fairness and unfairness and some shares may be considered as neither fair, nor unfair. In other words, we allow our participants to be inconsistent or indifferent about some shares.



I consider as being unfair all the shares in which A transfers to B less than 8% or more than 83% of the amount.

I consider as being unfair all the shares in which A transfers to B less than 34% or more than 64% of the amount.
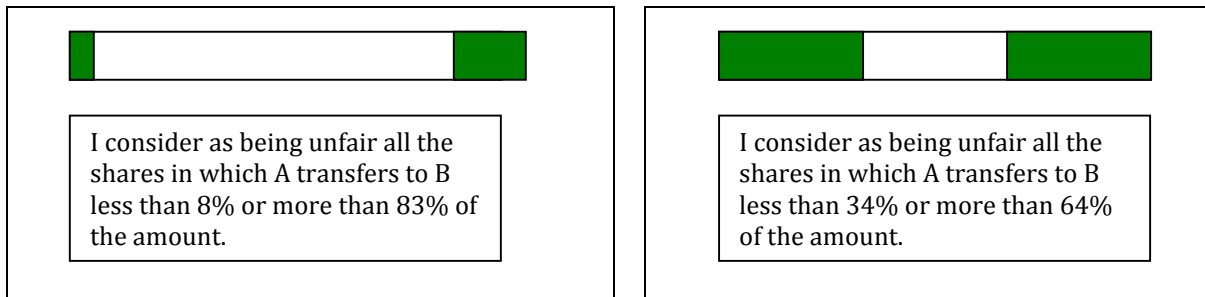
Figure 2. Determination of the sets of unfair shares – Two examples of players' screens

*III.2. Procedures*

The experiment consists of 5 sets of 2 sessions each conducted at the laboratory of the *Groupe d'Analyse et de Théorie Economique* in Lyon, France. Undergraduate students from the local engineering and business schools were invited via the ORSEE software (Greiner, 2004). Individuals committed to participate in two sessions separated by one week and they were informed that the earnings made in the two sessions would be paid only at the end of the

---

[15] The French original text of the Instructions is "*Qu'est-ce que vous considérez comme des partages injustes entre A et B ?*".

second session.  In total, 83 individuals participated in the two sessions, with 52% of females.[16] None of them already participated in a bargaining game or in a Trust game before. The experiment was computerized, using the REGATE platform (Zeiliger, 2000).

Upon arrival at the first session, the participants extracted a tag from a bag indicating their computer name. They also received another tag with a password. They were instructed to bring back this tag for the second session; otherwise they would not be allowed to participate.[17] The participants were informed that they would be paid €8 at the end of the second session for participating in the first session. They received sets of instructions for each scenario (see Appendix B) after completion of the reports regarding statements and hypothetical decision in the previous scenario, explaining the scenario and the rules for the expression of principles.[18] They had to answer a comprehension questionnaire and questions were answered in private. Participants were informed at the beginning of the first session that their responses in this session would in no case influence their participation in the second session or the content of the second session.

At the beginning of the second session, about one week later, the participants entered their individual password in the computer. They were informed that they would be paid €8 for participating in the second session in addition to the payment of one of the first three parts of this session that would be randomly drawn at the end of the session. They received the instructions for each part after completion of the previous part (see Appendix C). The order of games was the same as in the first session: first the Dictator game, then the Ultimatum game and finally the Trust game.  In each part, after a check for their understanding, we paired the participants and assigned randomly the two roles. It was made common information that in each game the decisions of the players would not be communicated to the other participant until the end of the session. Since the Ultimatum and the Trust games were played under the strategy method, the receivers in the Ultimatum game were not informed on their sender's decision and they accepted or rejected each possible amount sent by the sender.  Similarly, in the Trust game the receivers had

---

[16] Five individuals participated in the first session but did not show up at the second session.  One individual showed up in the second session and not in the first one; he was nevertheless accepted since we needed an even number of participants.  Of course, the data from these six participants are not included in the data analysis.
[17] The composition of the sessions differ between the first and the second sessions, as participants registered at the same time for two sessions but chose among the proposed schedules of each session separately. We did this on purpose to limit communication or coordination between participants. This is also why we were careful dismissing people at different moments at the end of the first session to avoid that they communicate with other participants.
[18] In each session, the scenarios were always administered in the same order: Dictator game, Ultimatum game, and Trust game. This does not allow us to control for a possible order effect across games but the sequence permits a progression in the degree of difficulty of the games, which has probably facilitated participants' understanding.

to decide on the amount of points sent back to the sender for each possible amount chosen by the sender. Last, we distributed a final set of instructions reminding the three scenarios and the rules for expressing the principles of fairness and unfairness. After completion of this final part, the participants received a feedback on the decisions of their co-participants in the first three parts of the second session. Finally, one of the first three parts was randomly selected for payment.

Payment was made in a separate room by a secretary who was not aware of the content of the experiment. This fact was made common information to all subjects from the very beginning of the first session. The first session of the experiment lasted on average 75 minutes and the second 90 minutes. Each participant earned an average of €28.99 (standard deviation: €8.18).

## IV. A THEORY OF SOCIAL NORMS AND HYPOCRISY: PREDICTIONS

What is the natural benchmark for abstract norms? Consider for example the Dictator game: a decision made behind the veil of ignorance might suggest fairness to be defined as equal splitting, as long as there are no reasons to think that the position of Dictator deserves special treatment because it is earned or deserved.[19] When this ideal translates into actual proposals, the advantageous bargaining position might creep into the moral reasoning and bias the actual proposal in favor of the Dictator. The same might happen in the Ultimatum and Trust games. Is this the case? Or instead even fairness ideals take into account the relative bargaining power of the two sides, and deviate from equal split rules? We take as the null hypothesis that individuals use the equal splitting rule independently of the role in the game.

**Hypothesis 1:** The fairness and unfairness bands are symmetric around the equal split rule. We test this hypothesis in section V. 1.

Abstract moral principles are naturally compared by individuals with real behavior, even their own. When real behavior is not readily available for comparison, principles are not constrained to some measure of consistency with actual choices, and are free to follow ideal positions that may reflect well on those who utter them. This holds both for statements as well as for hypothetical behavior. Instead, when the check given by real behavior is possible,

---

[19] Equity is the reference point in distributional models (Fehr and Schmit, 1999; Bolton and Ockenfels, 2000). In Falk and Fischbacher (2006) where individuals have to evaluate the kindness of another's action, the equitable share of payoffs is the reference standard to determine whether a share is kind or not. In several experiments on social norms, equal sharing is seen as more socially appropriate (Krupka and Weber, 2008; Andreoni and Bernheim, 2009).

the cost attached to the failure to implement a self-flattering statement is the bad signal of being inconsistent. Thus, we predict that when real behavior can be used to judge the consistency of facts and statements, the latter ones are more prudently in tune with real behavior.

**Hypothesis 2:** Stated fairness criteria will be farther from equal allocation after participants have made actual transfer choices which violate those criteria, moving the stated norm closer to actual behavior. We test this hypothesis in section V. 2.

Our next focus is on the relation between the expressed social preferences, strategic reasoning and behavior. We hypothesize that a strategic thinking mode may or may not be activated as players contemplate the game and they are cued to view the situation alternatively as a strategic setting or as an abstract moral setting. The three games differ in important respects in the way ethical and strategic considerations interact at the moment of formulating moral criteria. In the Ultimatum game the moral evaluation of the proposed allocation offered by the sender may have practical consequences because he knows that the receiver has to choose between accept or reject. Thus, the Ultimatum game naturally cues a strategic view even when the norm is stated as abstract. This was especially clear in our experimental setting since the scenario of the Ultimatum game was presented to the participants just after the scenario of the Dictator game, making the second mover – and therefore the strategic dimension of the game - more salient. In the other two games evaluations are void of practical consequences. In the Dictator game moral norms and the strategic reasoning never interact: the evaluation given by the second player has no consequence because he cannot act to affect the outcome in any way. Similarly, in the Trust game the evaluation given of the second mover's action has no strategic implications because no one can change payoffs anymore. Thus, Dictator game and Trust game do not cue strategic thinking in the abstract setting.

The strategic cue is instead active when participants, as in our experiment, evaluate the fairness of moves ex-post. When in our second session the individual is asked to choose and then to provide a new fairness statement, he now considers the question of the fair allocation together with the practical implication that he has just witnessed of an allocation among people. The practical consequence has now more weight (or some weight, as opposed to none) and the fairness statements reflect the power and opportunity of the first mover. In the Ultimatum Game, this is not new, because the real implication of the statement had been present from the start, so the change in the

perception of the statement is smaller and the adjustment smaller. The opposite analysis holds for the Dictator game and the Trust game.

**Hypothesis 3:** Real choices are more selfish than hypothetical ones. The discrepancy between the two will be larger in games where the real behavior has no strategic consequences (Dictator and Trust games) than in games where consequences are possible (Ultimatum game). We test this hypothesis in section V. 3.

Finally, and as a consequence of the previous hypotheses, we test the hypothesis that the adjustment of principles to actions differs across individuals, depending on the generosity of their actual choice and on their bargaining power in the games. We expect that moral hypocrisy is associated with more selfish actions and more powerful positions.

**Hypothesis 4:** At the individual level, power and selfishness increase the need for adjustment of stated principles to actions. We test this last hypothesis in section V. 4.

## V. EXPERIMENTAL RESULTS

*V. 1. The fairness criteria*

Our data show that even abstract fairness or unfairness principles take into account the relative bargaining power of the two sides. This is true independently of the point of view participants are asked to take. Table 1 displays the lower and upper bounds of the fairness and unfairness principles in each game and from the perspective of each player, as reported in the first session.

(*Insert Table 1 here*)

Table 1 indicates that the upper and lower bounds of stated fair shares are clearly in an asymmetric position with respect to an equal share. In the Dictator game, the midpoint of the upper and lower values for the dictator's point of view on fairness is 40.92, which is significantly different from 50 (t-test, two-tailed, $p<0.001$).[20] In the Ultimatum game, the midpoint of the upper and lower values for the sender's point of view is 41.57, which is also significantly different from 50 ($p<0.001$). In the Trust game, an equal share between the two players would require a transfer back of 2 in case of a low transfer by player A and a transfer back of 8 in case of a high transfer by A. The

---

[20] In this section, all tests are two-tailed t-tests with the individual as the unit of observation, unless specified otherwise.

14

midpoint of the upper and lower values for the receiver's point of view is 7.34 in case of the high transfer, which is significantly different from 8 ($p = 0.002$). In contrast, the midpoint is 2.39 in case of a low transfer, which is significantly *higher* than the transfer required by an equal share ($p<0.001$).

The same asymmetry holds for the unfairness bands. The area defined by the lower bound of unfair shares is always significantly smaller than the area defined by the upper bound. In the Dictator game, the largest share to the receiver that is still considered unfair is 18.93 (S.E. = 1.68) while the upper value for unfairness bound is 68.67 (S.E. = 1.76) when participants are asked to take the point of view of the dictator ($p<0.001$).[21] In the Ultimatum game, the lower bound is 20.24 (S.E. = 1.58) and the upper bound is 66.14 (S.E. = 1.86) from the point of view of the sender ($p<0.001$). In the Trust game, the lower bound is 1.07 (S.E. = 0.09) and the higher bound is 4.18 (S.E. = 0.19) when participants are asked to take the point of view of the receiver in case of a low transfer by player A ($p<0.001$); the corresponding values are 4.39 (S.E. = 0.29) and 10.86 (S.E. = 0.33) in case of a high transfer ($p<0.001$).[22]

Interestingly, the asymmetries are also observed when one takes the perspective of the less favored player, and in the same direction, favoring the stronger player: see Table 1. The midpoint of the upper and lower values of fairness is also significantly different from 50 in both the Dictator game ($p<0.001$) and the Ultimatum game ($p = 0.018$) in the perspective of the receiver. Similarly in the Trust game, the midpoint differs from 2 in case of the low transfer ($p<0.001$) and from 8 in case of the high transfer ($p = 0.015$) from the point of view of the sender. The same asymmetry is observed regarding the unfairness bands. When participants are asked to take the point of view of the receiver, the width of the lower unfair band is significantly smaller than that of the upper band in the Dictator game ($p = 0.031$) and the Ultimatum game ($p = 0.026$). When participants are asked to take the point of view of the sender in the Trust game, the difference is also significant in the case of a low transfer but not in the case of a high transfer ($p<0.001$ and $p = 0.031$, respectively). In summary, the player in the weaker position makes statements that anticipate that the stronger side will use the power, rather than (for example) claiming that allocations should be independent of the bargaining position.

---

[21] These paired *t*-tests compare the width of the area delimited by the lower bound and the minimum possible transfer and the width of the area delimited by the upper bound and the maximum possible transfer.
[22] Due to the fact that participants determine the unfair shares without being reminded which shares they defined as being fair, there may some overlapping between fair and unfair shares. This does not necessarily denote confusion, but rather a margin of error. This concerns a small minority of players. For example, in session 1 (session 2, respectively) we observe overlapping between the lower bounds of fairness and unfairness in 14.46% of the players in the Dictator game (15.66%, respectively), 7.23% in the Ultimatum game (2.41%), 8.43% in the Trust game with a low transfer (12.05%) and 10.84% in the Trust game with a high transfer (9.64%).

In fact, there is no difference between the lower bounds of shares that are considered fair and unfair in the two points of view. Consider for example the Dictator game, where strategic considerations are least likely to influence the individual's criteria. In the first session, the mean lower bound on the share given to the receiver that is considered fair is 24.19 (S.E. = 1.79) in the position of the dictator; the value is 24.52 (S.E. = 1.53) in the position of the receiver. These values do not differ significantly ($p = 0.831$). Similarly, there is no difference between the two points of view about the lower bound of the unfair shares in this game ($p = 0.151$) as well as in the Ultimatum game ($p = 0.134$).

These results lead us to reject hypothesis 1:

**Result 1:** Abstract principles of fairness and unfairness take into account the bargaining power of the players, conceding more to the more powerful player. This is true for statements made from the perspective of both the most and the least powerful players, in both cases favoring the most powerful player.

*V. 2. Behavior and principles: From real acts to words*

Table 2 displays the evolution of fairness and unfairness principles between the first and the second sessions for each game and in each player's perspective. In the second session, principles are stated after players made real decisions; we can therefore relate the evolution of principles between the two sessions to the actual decisions.

(*Insert Table 2 here*)

Table 2 shows how players in the position of a higher bargaining power adjust their principles in the second session. In the Dictator game, dictators adjust downward both lower and upper bounds of the fairness statement. In the first session they consider that 22 is the minimum to consider a transfer as being fair and that a share is fair up to 59.44; in the second session, 15.8 is the minimum and 56.34 is the maximum. These differences are significant ($p = 0.031$ for the lower bound in a two-tailed t-test; $p = 0.087$ for the upper bound in a one-tailed t-test). If one considers midpoints of fairness bands, the asymmetry of fair shares relative to the equal split is increased in the second session compared to the first one. The difference between the midpoints in the first and the second sessions is significant in the Dictator game ($p = 0.031$) and in the Trust game when the transfer is high ($p = 0.012$). The same conclusion applies to the evolution of the unfairness principles that are also adjusted downwards (see the lower part of Table 2). In contrast, the asymmetry of fair shares is not significantly modified in the second session in the Ultimatum game where the player's offer can still be rejected by the receiver ($p = 0.761$). The difference is not significant either in the

Trust game when the initial transfer is low ($p = 0.465$). The fact that individuals adjust their principles in the Dictator game and the Trust game with a high transfer but not when strategic motives are present or in the Trust game with a low transfer confirms that the adjustment of principles in the direction of more selfish actions in the second session cannot be explained by a learning effect. Indeed, if people were adjusting in the second session because they have learned to play the equilibrium of the games, they would have adjusted their principles in all games; this is not what we observe. Similarly, the adjustment cannot be explained by an anchoring effect of the actions, otherwise we should have observed the same process of adjustment in all games. The adjustment of principles affects in no game the size of the fairness and unfairness bands ($p$-values are always higher than 0.100): the moral wriggle room is not increased, nor decreased.

A striking result is that players who have been assigned a weaker role in the bargaining also adjust downwards their principles of fairness and unfairness when placed in the perspective of the other player. For example, Table 2 shows that in the first session, they consider that 26.33 is the minimum to consider a transfer as being fair in the Dictator game; in the second session, instead they consider that 20.74 is the minimum ($p = 0.029$). The upper bound of fair shares is, however, not modified (55.90 in the first session and 55.31 in the second session, $p = 0.805$). This shows that for players who have a lower bargaining power, not only the definition of fairness becomes more asymmetric relative to an equal sharing (the midpoint is lower in session 2 than in session 1, $p = 0.075$ in a one-tailed test), but also the range of shares that are considered as fair is increased after playing the game for real ($p = 0.047$). In addition, for these players the lower bound of unfair shares is also adjusted downwards in the second session ($p = 0.088$). The adjustment of the participants selected as receivers is, however, slightly smaller than the dictators. Indeed, while the lower bounds of fairness are not significantly different in the first session between actual dictators and receivers, they become significantly different in the second session (Mann-Whitney U-test, $p = 0.077$).

In the Trust game with a high transfer, the senders also relax their norm of fairness in the perspective of the receiver in the second session: the lower bound of the fair back transfers is adjusted downwards between the first and the second sessions ($p = 0.066$, one-tailed), moving the midpoint to the left ($p = 0.109$, one-tailed). In contrast, in the Ultimatum game the receivers like the senders do not adjust the lower bound of their fairness principle when placed in the perspective of the sender ($p = 0.368$). They do not adjust either the lower bound ($p = 0.193$) or the upper bound of unfairness principles ($p = 0.640$). However, they increase marginally the upper bound of the fairness

principle ($p$ = 0.086), so that the midpoint of lower and upper bounds of fairness principles is moved slightly to the right compared to the first session ($p$ = 0.060).

This analysis leads to the following results that support Hypothesis 2.

**Result 2a:** In most games, the players with a higher bargaining power adjust their principles to their actions by relaxing their norm of fairness and by extending the range of fair allocations to their advantage.

**Result 2b:** The players with a lower bargaining power also adjust downwards their principles of fairness and unfairness when placed in the perspective of the more powerful player.

*V. 3 Hypothetical and real choices*

Our findings are in agreement with Hypothesis 3: In the Dictator and Trust games, where the evaluation of the action of the other has no consequence, hypothetical and real choices differ. In the Ultimatum game, where the evaluation has consequences, the two choices do not differ.

Let's consider first the difference between hypothetical transfers in the first session and real transfers in the second session in the Dictator game. Not surprisingly, senders are much less generous when making real choices than hypothetical ones. 34.15% of the dictators play the equilibrium (0) for real in the second session while only 14.63% of them did so when making a hypothetical choice. The average hypothetical transfer is 2.66 (S.E. = 0.24), the average real one is 1.78 (S.E. = 0.29).[23] The difference is significant according to a paired t-test ($p$ = 0.006) and a Kolmogorov-Smirnov test (p<0.001, corrected). The left panel of Table 3 illustrates the differences between the two distributions of choices. Regarding the Trust game, Table 4 reports the hypothetical and actual amounts sent back by the receivers in the cases of both a low transfer (one out of five points) and a high transfer (four out of five points) from the sender.

(*Insert Tables 3 and 4 here*)

In the Trust game like in the Dictator game, the real amounts are significantly lower than the hypothetical amounts and by a sizeable amount. In the case of a low initial transfer, the zero transfer represents only 7.14% of the hypothetical choices of the receivers but 38.10% of their actual decisions. In case of a high transfer, the

---

[23] The mean actual transfers in the three games are consistent with the usual values observed in these games (Plott and Smith, 2008); we do not think that having reported principles in session 1 has modified actual decisions in session 2.

corresponding values are 0% and 23.81%. The picture is the same when one considers the percentages of back transfers that would equalize payoffs between senders and receivers. In case of a low transfer, 52.38% of the hypothetical choices but only 19.05% of the real choices are equal to 2; in case of a high transfer, 26.19% of the hypothetical choices but none of the real choices are equal to 8. The average amount sent back in the second session if the sender made a low transfer is 0.88 (S.E. = 0.137) whereas the average hypothetical amount in the first session was 1.74 (S.E. = 0.150). The hypothetical and actual values differ significantly (t-test: $p < 0.001$; Kolmogorov-Smirnov test: $p < 0.001$, corrected). The average amount sent back in the second session if the sender has sent 4 points is 4 (S.E. = 0.447) whereas the average hypothetical amount was 6.38 (S.E. = 0.334). The difference is large and significant (t-test: $p < 0.001$; Kolmogorov-Smirnov test: $p < 0.001$, corrected).

Instead, there is no large difference between hypothetical and real choices for the sender in the Ultimatum game (see right panel of Table 3). Indeed, the equilibrium play (0 or 1) represents 7.14% of the hypothetical choices of the senders and only 4.76% of their actual decisions. The mean actual transfer to the receiver is 3.50 (S.E. = 0.17) whereas the mean hypothetical transfer was 3.43 (S.E. = 0.18). The difference is not significant (t-test, $p = 0.691$) although the distributions of hypothetical and actual transfers differ significantly (Kolmogorov-Smirnov test: $p < 0.001$, corrected). The amount of the transfers is consistent with the hypothesis that strategic considerations drive both hypothetical and real choices: the hypothetical transfer is 28.95% higher than in the Dictator game and the actual transfer is even 96.63% higher (both percentages in the Ultimatum game are significantly higher than in the Dictator game, $p < 0.001$).

This analysis supports our next result.

**Result 3:** In the Dictator and Trust games, hypothetical and real choices differ, and the change favors the first mover. The change is smaller or absent in the Ultimatum game.

*V. 4 Actual fairness and hypocrisy*

In this last sub-section, we complement the previous analysis by relating the relative fairness of the players to the adjustment of their stated principles depending on their bargaining power. We study whether the most hypocritical players are or not the most selfish and powerful ones.

In each game the sample has been divided in three categories based on the transfers actually made in session 2 by the senders in the Dictator and the Ultimatum games, and by the receivers in the Trust game. The categories correspond to transfers below the median, equal to the median, and above the median. To assess hypocrisy, we consider i) the differences between fairness principles stated in session 1 and actual transfers, ii) the difference between hypothetical and actual transfers, and iii) the adjustment of principles between the two sessions. For each game and each category of players, Figures 3 to 6 display four bars indicating respectively the mean lower bound of the fairness principle stated in session 1, the mean lower bound of the fairness principle stated in session 2, the mean hypothetical choice in session 1 and the mean actual choice in session 2.[24] In the Dictator game (Figure 3) and in the Ultimatum game (Figure 4), the principles (expressed in a percentage scale) have been rescaled between 0 and 10 to enable the comparison with hypothetical and actual transfers that are expressed between 0 and 10.

In addition, Table 5 indicates, for each game and each category of actual transfer relative to the median, the percentage of variation between the mean actual transfer and the lower bound of fairness stated in session 1 (column 1), the percentage of variation between the mean fairness principles stated in session 2 and in session 1 (column 2), and the percentage of variation between the mean actual and hypothetical transfers (column 3). Each column includes the *p*-value of two-tailed *t*-tests comparing respectively actual transfers and principle stated in session 1, principles stated in sessions 1 and 2, hypothetical and actual transfers.

(*Insert Figures 3 to 6 and Table 5 here*)

Let's consider first the Dictator game and the Trust games. Figures 3, 5 and 6 and Table 5 indicate that in these games the participants who actually transfer less than the median exhibit the highest discrepancy between their hypothetical transfer and their actual transfer. The difference is -86.49% in the Dictator game, -100% in the Trust game with a low transfer, and -89.25% in the Trust game with a high transfer. The amount transferred by these individuals is significantly lower than the lower bound of their fairness principle stated in session 1 and the difference is large (-88.97% in the Dictator game, -100% in the Trust game with a low transfer, and -89.40% in the Trust game with a high transfer). For these players, except in the Trust game with a low transfer, the fairness principle is revised downward and significantly so (-34.65% in the Dictator game and -25.40% in the Trust game with a high transfer). The revision is not sufficient, however, to include the actual decisions in the fairness bands, but

---

[24] We display only the lower bounds of the fairness principles for simplification. The same figures could of course be displayed with the midpoints of fairness intervals, the lower and upper bounds of unfairness principles.

this allows the players to reduce the distance with what they consider as fair. In the Dictator game, it should also be noted that the players who transferred less than the median reported on average a much more ambitious fairness principle than the players whose transfer is above the median (2.38 vs. 1.59). In the Trust game with a low transfer, the receiver's hypocrisy is less necessary because a low back transfer can be easily justified by negative reciprocity in reaction to the first mover's low level of trust. It should be also acknowledged that in such a context there are less possible returns and therefore lesser possibilities to adjust the fairness bands.

Individuals whose actual transfer is at the median level exhibit also a significant discrepancy between hypothetical and actual transfers in the Dictator and Trust games but its importance is twice smaller than for those who transferred less than the median (-40.55% in the Dictator game, -41.52% in the Trust game with a low transfer, and -39.39% in the Trust game with a high transfer). Average transfers do not differ significantly from the lower bound of the fairness principles stated in session 1. Still, these participants adjust the lower bound of fairness downward, except in the Trust game with a low transfer (-46.73% in the Dictator game and -26.09% in the Trust game with a high transfer). They want to appear more generous than what they state as the minimum fair allocation.

In contrast, the participants who transferred more than the median do not adjust their principle of fairness in session 2. In the Dictator game, this may be due to the fact that they transfer more than their hypothetical choice (+31.43%) and more than the lower bound of fairness stated in session 1 (+162.85%). In the Trust game with a low transfer, actual choices do not differ significantly from the hypothetical ones and they exceed significantly the lower bound of fairness (+42.31%). In the Trust game with a high transfer, mean transfers are slightly lower than the hypothetical ones (-13.11%) but they remain significantly higher than the lower bound of fairness stated in session 1. Overall, these participants are relatively consistent with their initial principles and they do not need adjusting their principles to convince themselves that they are fair.

In the Ultimatum game behavior is relatively different. The participants whose transfer is below the median send significantly less than their hypothetical choice but the difference is much smaller than in the other games (-14.03%) and these transfers still exceed significantly the lower bound of fairness. If these players adjust their principles downward in session 2, it is probably because they want to keep the image of sending much more than the minimum fair allocation. The participants at or above the median do not adjust their principles: they actually send the same or even a significantly higher amount (+25%) than their hypothetical choices in session 1 and this exceeds

significantly the lower bound of fairness stated in session 1 (+46.79% and +50.01%, respectively). This analysis leads to our last results that confirm Hypothesis 4:

**Result 4a:** Hypocrisy is stronger in individuals who transfer less.

**Result 4b:** More power is associated with more hypocrisy.


VI. DISCUSSION AND CONCLUSION

The results of our experiment suggest the general rule that the practical implications of stated normative principles about actions affect the stated norms, and *vice versa*. This intrusion of strategic evaluation into normative setting occurs in many ways and the influence goes in many directions.

Norms that appear abstract take into account the bargaining power of the two sides. Individuals in advantageous position anticipate how future possible behavior might deviate from tight moral standards set in those statements, and make them less stringent beforehand; their stated principles deviate from the norm of equal sharing. Individuals in weaker positions also anticipate that the stronger side will take advantage of the position, and are willing to make this fact into a norm, declaring fair the unequal allocation favoring the opposite side. Thus, power intrudes in ethical norms, in the mind of the weak and in that of the strong, and bends the norm in favor of the powerful.

Individuals adjust the range of fair shares and unfair shares after playing the game for real money compared with the initial statement they gave when the criteria of fairness and unfairness were elicited as universal but inconsequential norms. Individuals in an advantageous as well as in a disadvantageous position make the adjustment in the same direction. Moral hypocrisy is used as a tool to manage the tradeoff between the immediate convenience of the actions and the conflict these actions create with principles. It balances the need to maintain a social reputation in the future and the convenience of a present choice. Moral hypocrisy may be based on self-deception, but this could be in turn used to deceive others (von Hippel and Trivers, 2011). If people are conscious of the importance of the evaluation of intentions (Rabin, 1993; Charness and Rabin, 2002), it is all the more important for them to adjust their stated principles to their actions. The importance of the adjustment is also reduced when another player has behaved selfishly, i.e. in the Trust game when the sender expressed a low level of trust.

The adjustment of principles to actions is considerably smaller or absent when there are strategic reasons that dictate prudence and fairness in deciding transfers. The discrepancy between hypothetical and real behavior is larger when the action being judged has no further consequence (as the first move in the Dictator and the second move in the Trust game) than when it does (as the first move in the Ultimatum game). In our experiment, if an allocation has the strategic value of affecting future actions, then it is probably perceived differently already in the first session, in spite of the fact that this session is all about abstract judgments. This difference in perception is reflected in the lack of adjustment of stated norms and hypothetical choices in the second session: since the action was evaluated as strategic in the first session already, facing the fact that it has real consequences of affecting payoffs of individuals is not a novelty, as it is in the first session. The fact that individuals adjust their statements in games where they are the last movers but not when strategic motives are present confirms that the adjustment of principles in the direction of more selfish actions in the second session cannot be explained by a learning effect. Indeed, if people were adjusting in the second session because they have learned to play the equilibrium of the games, they would have also adjusted their principles after playing the Ultimatum game. This is not what we observe. Similarly, learning cannot explain the diverging changes between hypothetical and actual transfers according to the level of actual transfers relative to the median. Our findings illustrate that a stronger power facilitates the enunciation of generous ethical principles that may be distant from real actions because individuals do not examine the situation as a strategic one when they express them. This could also contribute to explain that, for example, politicians express more generous views in their discourses before being in responsibility.

Individuals who adjust more their statements also behave more selfishly. These results should suggest a reconsideration of the significance of norms and social preferences. They also suggest the more modest task of re-evaluating how these norms are elicited. In economics, social preferences are usually identified as a given share elicited through actions in a game. This definition clearly fails to measure the strength of the support the individual gives to the preference, and does not disentangle the interplay between normative and strategic considerations. In our experiment, we let norms to be defined by intervals, which gives a better measure of the wiggle room allowed and reveals that social preferences may be contextual and flexible.

The discrepancy between statements and acts and the subsequent adjustment of fairness and unfairness statements pose the question of a systematic study of moral hypocrisy in strategic behavior. Occurrence of hypocrisy is due to the fact that people build an identity when stating their initial principles and hypothetical choices in session

23

1. Since there is no cost under the veil of ignorance for looking pro-social, most of them build an excessively generous image of themselves. However, our design forces people to make decisions in the second session when acting in accord with pro-social principles is costly and at that moment they cannot remain ignorant of their true identity (there is no opt-out option in our environment). After making choices that are usually less pro-social than the initially stated principles, especially if they get the highest bargaining power, they observe that they lied to themselves. They adjust their principles in the direction of a better alignment with their actions to keep up the appearance of being pro-social and to maintain a positive self-image. By reducing the distance between the fairness principles and their actual behavior, people may convince themselves that their actions do not hurt their morale since they are closer or belong to a more permissive fairness interval. Hypocritical individuals are those who reported initially the most generous statements and who then behaved the most selfishly. In contrast, the individuals who behave more generously (by transferring more than the median) do not feel the need to reevaluate their stated principles.

REFERENCES

Abbink, K., Sadrieh, A. (2009). The pleasure of being nasty. *Economics Letters*, 105(3), 306-308.

Aidt, T.S. (2003). Economic Analysis of Corruption: A Survey. *The Economic Journal*, 113, F632-F652.

Andreoni, J., Bernheim, B.D. (2009). Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, 77(5), 1607-1636.

Ariely, D., Bracha, A., Meier, S. (2009). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, 99(1), 544-555.

Bandura, A. (1990). Selective activation and disengagement of moral control. *Journal of Social Issues*, 46, 27–46.

Bandura, A. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, 71, 364–367.

Batson, C.D., Kobrynowicz, D., Dinnerstein, J.L., Kampf, H.C., Wilson, A.D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, 72, 1335–1348.

Batson, C.D., Thompson, E.R., Seuferling, G., Whitney, H., Strongman, J. (1999). Moral Hypocrisy: Appearing Moral to Oneself Without being so. *Journal of Personality and Social Psychology*, 77, 525–537.

Batson, C.D., Thompson, E.R., Chen, H. (2002). Moral Hypocrisy: Addressing Some Alternatives. *Journal of Personality and Social Psychology,* 83, 330–339.

Benabou, R., Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review* 96(5), 1652-1678.

Bernheim, B. D. (1994). A Theory of Conformity. *Journal of Political Economy*, 102, 841–877.

Bodner, R., Prelec, D. (2003). Self-signaling and Diagnostic Utility in Everyday Decision Making. In Brocas, I., and Carillo, J. (Eds.), *The Psychology of Economic Decisions*. Oxford: Oxford University Press.

Bolton, G.E., Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *American Economic Review*, 90 (1), 166–193.

Brunsson, N. (1989). *The Organization of Hypocrisy: Talk, Decisions and Action in Organizations*. New York: John Wiley & Sons.

Brunsson, N. (1995). Ideas and Actions: Justification and Hypocrisy as Alternatives to Control, in Bacharach, S., Gagliardi, P., Mundell, B. (Eds.), *Research in the Sociology of Organizations*, Vol. 13: Studies of Organizations in the European Tradition, Greenwich, CT:, JAI Press, 211-235.

Brunsson, N. (2003). Organized Hypocrisy, in Czarniawska, B. Sevón, G. (Eds.), *The Northern Lights: Organization Theory in Scandinavia*. Oslo: Copenhagen Business School Press, 201-222.

Charness, G., Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117, 817-869.

G. Charness, D. Masclet, M.C. Villeval. (2010). Competitive Preferences and Status as an Incentive: Experimental Evidence. IZA Discussion Paper 5034, Bonn.

Costa, P.T., McCrae, R.R., (1992). *Revised Neo Personality Inventory (Neo-Pi-R) and Neo Five-Factor Inventory (Neo-Ffi) Professional Manual*. Odessa: FL: Psychological Assessment Resources.

Costa, P.T., McCrae, R.R. (2004). A Contemplated Revision of the Neo Five-Factor Inventory. *Personality and Individual Differences 36*, 587-96.

de Cremer, D., van Dijk, E. (2005). When and why leaders put themselves first: Leader behaviour in resource allocations as a function of feeling entitled. *European Journal of Social Psychology,* 35, 553-563.

Dana, J., Weber R.A., Xi Kuang, J. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1), 67-80.

Ellingsen, T., Johannesson, M. (2008). Pride and Prejudice : The Human Side of Incentive Theory, *American Economic Review*, 98, 990-1008.

Falk, A., Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54, 293-315.

Fehr E., Schmidt, K.M. (1999). A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics*, 114, 817-868.

Fischbacher, U., Gächter, S. (2010), Social preferences, beliefs, and the dynamics of free-riding in public good experiments, *American Economic Review*, 100 (1), 541-556.

Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K., Macho, V., *Forschung und wissenschaftliches Rechnen* GWDG Bericht 63, Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Grossman, Z. (2010). Strategic ignorance and the robustness of social preferences. Working paper, University of California at Santa Barbara.

Harbring, C., Irlensbusch, B. (2011). Sabotage in Tournaments: Evidence from the Laboratory, *Management Science*, 57(4), 611-627.

Von Hippel, W., Trivers, R. (2011). The evolution and psychology of self-deception, *Behavioral and Brain Sciences*, 34, 1–56.

Huzzard, T., Östergren, K. (2002). When Norms Collide: Larning under Organizational Hypocrisy. *British Journal of Management*, 13, S47-S59.

Krasner, S.D. (1999a). Logics of Consequences and Appropriateness in the International System," in Egeberg, M., Laegrid, P. (Eds.), *Organizing Political Institutions: Essays in Honor of Johan Olsen*. Oslo: Universitetsforslaget, 181-213.

Krasner, S.D. (1999b). *Sovereignty: Organized Hypocrisy*. Princeton University Press, Princeton, NJ.

Krupka, E.L., Weber, R.A. (2008). Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? IZA Discussion Paper 3860, Bonn.

Larson, T., Capra, M. (2009). Exploiting moral wiggle room: Illusory preference for fairness? A comment. *Judgment and Decision Making*, 4(6), 467-474.

Lipson, M. (2006). Dilemmas of Global Governance: Organized Hypocrisy and International Organization. Paper presented at the annual convention of the Canadian Political Science Association, Toronto, Ontario, June 1-3.

Maher, B. (2010). Research Integrity: Sabotage! *Nature*, 467, 30 September, 516-518.

Meyer, J.W., Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83, 340-63.

Meyer, J.W., Rowan, B.. (1978). The Structure of Educational Organizations. In Meyer, M.W. (Ed.), *Environments and Organizations*. San Francisco: Jossey-Bass.

Mui, V.L. (1995). The economics of envy. *Journal of Economic Behavior & Organizations*, 26(3), 311-336.

Ostrom, E. (2000). Collective Action and the Evolution of Social Norms. *Journal of Economic Perspectives*, 14(3), 137-158.

Plott, C.R., Smith, V.L. Eds.) (2008). *Handbook of Experimental Economics Results*. Vol.1. Amsterdam: North Holland.

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

Robin, S., Rusinowska, A., Villeval, M.C. (2012). Ingratiation and influence within organizations. GATE Working Paper, Lyon.

Rolland, J.P., Parker, W., Stumpf, H. (1998). A Psychometric Examination of the French Adaptation of the Neo-Pir and Neo-Ffi. *Journal of Personality Assessment, 71*(2), 269-291.

Runciman, D. (2008). *Political Hypocrisy: The Mask of Power, from Hobbes to Orwell and Beyond*. Princeton: Princeton University Press.

Stone, J., Wiegand, A.W., Cooper, J., Aronson, E. (1997). When exemplification fails: Hypocrisy and the motive for self-integrity. *Journal of Personality & Social Psychology*, 72, 54-65.

Trivers, R. (1985). *Social evolution*. Menlo Park: Benjamin/Cummings.

Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences,* 907, 114–131.

Trivers, R. (2009). Deceit and self-deception. In Kappeler, P., Silk, J. (Eds.), *Mind the Gap. Tracing the Origins of Human Universals*. Heidelberg: Springer-Verlag, 373–393.

Trivers, R. (2011). *Deceit and self-deception. Fooling Yourself the Better to Fool Others.* New-York: Allen Lane, Penguin Books.

Valdesolo, P., DeSteno, D.A. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, 18, 689–690.

Valdesolo, P., DeSteno, D.A. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, 44 (5), 1334-1338.

Van der Weele, J. (2012). When ignorance is innocence: on information avoidance in moral dilemmas. SSRN working paper.

Watson, G.W., Teague,B.T., Papamarcos, S.D. (2006). Moral Hypocrisy: A Matter of Measure? In Wurtzel, A.E. (Ed.), *Trends in Contemporary Ethical Issues*, New-York: Nova Science Publishers, 1-14.

Zeiliger, R., (2000). A Presentation of Regate, Internet Based Software for Experimental Economics. http://www.gate.cnrs.fr/~zeiliger/regate/RegateIntro.ppt, GATE.

Zizzo, D., Oswald, A.J. (2001). Are People Willing to Pay to Reduce Others' Incomes? *Annales d'Economie et de Statistique*, 63-64, 39-62.

## Appendix A. Tables

Table 1: Fairness and unfairness principles, for the three games

| Principles | In the shoes of Player A | | In the shoes of Player B | |
|---|---|---|---|---|
| | Fairness | Unfairness | Fairness | Unfairness |
| *Dictator Game* | | | | |
| Lower bound | 24.19 (1.79) | 18.93 (1.68) | 24.52 (1.53) | 22.29 (1.98) |
| Upper bound | 57.65 (1.53) | 68.67 (1.76) | 62.53 (1.70) | 71.04 (1.81) |
| *Ultimatum Bargaining Game* | | | | |
| Lower bound | 24.71 (1.56) | 20.24 (1.58) | 27.29 (1.58) | 22.58 (1.57) |
| Upper bound | 58.45 (1.36) | 66.14 (1.86) | 66.42 (1.84) | 71.45 (2.0) |
| *Trust Game – low transfer* | | | | |
| Lower bound | 1.63 (0.80) | 1.40 (0.09) | 1.30 (0.90) | 1.07 (0.92) |
| Upper bound | 4.11 (0.19) | 4.54 (0.22) | 3.49 (0.16) | 4.18 (0.19) |
| *Trust Game – High transfer* | | | | |
| Lower bound | 5.83 (0.22) | 5.02 (0.24) | 4.99 (0.25) | 4.39 (0.29) |
| Upper bound | 11.11 (0.30) | 11.55 (0.35) | 9.69 (0.28) | 10.86 (0.34) |

Note : Standard errors are in parentheses.  N=83.

Table 2: Evolution of the principles between the first and second sessions, for the three games.

2.1. Fairness principles

| Principles | Actual Player A in session 2 | | | Actual Player B in session 2 | | |
|---|---|---|---|---|---|---|
| | Session 1 | Session 2 | % | Session 1 | Session 2 | % |
| *Dictator Game* | | | | | | |
| Lower bound | 22 (2.38) | 15.80 (2.21) | -28.18 | 26.33 (2.64) | 20.74 (2.14) | -21.23 |
| Upper bound | 59.44 (2.57) | 56.34 (2.80) | -5.22 | 55.90 (1.67) | 55.31 (2.33) | -1.06 |
| *Ultimatum Bargaining Game* | | | | | | |
| Lower bound | 25.24 (2.16) | 24 .31 (2.11) | -3.68 | 24.17 (2.27) | 25.90 (2.15) | +0.07 |
| Upper bound | 58.83 (2.06) | 58.83 (2.29) | 0 | 58.05 (1.80) | 61.37 (2.28) | +5.72 |
| *Trust Game – low transfer* | | | | | | |
| Lower bound | 1.41 (0.13) | 1.24 (0.14) | -12.06 | 1.17 (0.12) | 1.14 (0.12) | -2.56 |
| Upper bound | 3.71 (0.24) | 3.83 (0.29) | +3.23 | 3.29 (0.21) | 3.55 (0.23) | +7.90 |
| *Trust Game – High transfer* | | | | | | |
| Lower bound | 4.83 (0.42) | 4.24 (0.34) | -12.22 | 5.14 (0.28) | 4.55 (0.30) | -11.48 |
| Upper bound | 9.41 (0.44) | 9.12 (0.42) | -3 .08 | 9.95 (0.36) | 9.09 (0.34) | -8.64 |

2.2. Unfairness principles

| Principles | Actual Player A in session 2 | | | Actual Player B in session 2 | | |
|---|---|---|---|---|---|---|
| | Session 1 | Session 2 | % | Session 1 | Session 2 | % |
| *Dictator Game* | | | | | | |
| Lower bound | 16.59 (2.14) | 13.73 (1.98) | -17.24 | 21.21 (2.55) | 16.95 (1.84) | -20.08 |
| Upper bound | 71.63 (2.58) | 65.49 (2.93) | -8.57 | 65.79 (2.36) | 62.57 (2.73) | -4.89 |
| *Ultimatum Bargaining Game* | | | | | | |
| Lower bound | 19.57 (2.14) | 18.14 (2.06) | -7.31 | 20.92 (2.35) | 17.24 (1.89) | -17.59 |
| Upper bound | 66.52 (2.83) | 63.10 (2.47) | -5.14 | 65.76 (2.45) | 64.15 (2.98) | -2.45 |
| *Trust Game – low transfer* | | | | | | |
| Lower bound | 0.90 (0.12) | 1.05 (0.11) | +16.67 | 1.24 (0.14) | 0.98 (0.14) | -20.97 |
| Upper bound | 4.17 (0.30) | 4.34 (0.32) | +4.08 | 4.19 (0.25) | 4.02 (0.26) | -4.06 |
| *Trust Game – High transfer* | | | | | | |
| Lower bound | 3.93 (0.45) | 4.00 (0.34) | +1.78 | 4.83 (0.37) | 4.28 (0.30) | -11.38 |
| Upper bound | 10.63 (0.52) | 10.61 (0.40) | -0.19 | 11.07 (0.44) | 10.69 (0.33) | -3.43 |

Note: In the Dictator and Ultimatum Games, the tables display the principles of players in the position of player A; in the Trust Game, they display the principles in the position of player B. Standard errors are in parentheses. N=42.

Table 3: Transfers of senders, hypothetical and real, in Dictator game and Ultimatum game

| Transfer in points | Dictator game | | | | Ultimatum game | | | |
|---|---|---|---|---|---|---|---|---|
| | Hypothetical choices | | Actual choices | | Hypothetical choices | | Actual choices | |
| | # | % | # | % | # | % | # | % |
| 0 | 6 | 14.63 | 14 | 34.15 | 0 | 0 | 0 | 0 |
| 1 | 2 | 4.88 | 5 | 12.20 | 3 | 7.14 | 2 | 4.76 |
| 2 | 9 | 21.95 | 11 | 26.83 | 5 | 11.90 | 5 | 11.90 |
| 3 | 11 | 26.83 | 5 | 12.20 | 14 | 33.33 | 14 | 33.33 |
| 4 | 9 | 21.95 | 2 | 4.88 | 12 | 28.57 | 12 | 28.57 |
| 5 | 4 | 9.76 | 3 | 7.32 | 7 | 16.67 | 9 | 21.43 |
| 6 | 0 | 0 | 0 | 0 | 1 | 2.38 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 1 | 2.44 | 0 | 0 | 0 | 0 |
| Total | 41 | 100 | 41 | 100 | 42 | 100 | 42 | 100 |

Note : In the Dictator game, there are 41 observations instead of 42 because one of the dictators participated only in the second session; therefore, his decisions are not taken into account in the data analysis.


Table 4: Transfers back of receivers, hypothetical and real, in Trust games

| Transfer back in points | Low transfer of player A (1/5) | | | | High transfer of player A (4/5) | | | |
|---|---|---|---|---|---|---|---|---|
| | Hypothetical choices | | Actual choices | | Hypothetical choices | | Actual choices | |
| | # | % | # | % | # | % | # | % |
| 0 | 3 | 7.14 | 16 | 38.10 | 0 | 0 | 10 | 23.81 |
| 1 | 12 | 28.57 | 17 | 40.48 | 1 | 2.38 | 2 | 4.76 |
| 2 | 22 | 52.38 | 8 | 19.05 | 2 | 4.76 | 3 | 7.14 |
| 3 | 4 | 9.52 | 0 | 0 | 1 | 2.38 | 5 | 11.90 |
| 4 | 0 | 0 | 1 | 2.38 | 4 | 9.52 | 7 | 16.67 |
| 5 | 1 | 2.38 | 0 | 0 | 5 | 11.90 | 6 | 14.29 |
| 6 | - | - | - | - | 7 | 16.67 | 3 | 7.14 |
| 7 | - | - | - | - | 6 | 14.29 | 6 | 14.29 |
| 8 | - | - | - | - | 11 | 26.19 | 0 | 0 |
| 9 | - | - | - | - | 4 | 9.52 | 0 | 0 |
| 10 | - | - | - | - | 0 | 0 | 0 | 0 |
| 11 | - | - | - | - | 1 | 2.38 | 0 | 0 |
| Total | 42 | 100 | 42 | 100 | 42 | 100 | 42 | 100 |

Table 5: Comparison between principles in sessions 1 and 2, hypothetical and actual transfers, by game and by category of actual transfers

| Percentage of variation | Actual transfer – principle stated in session 1 (1) | Principle stated in session 2 – principle in session 1 (2) | Actual – hypothetical transfers (3) |
|---|---|---|---|
| *Dictator game* | | | |
| Below median (N=19) | -88.97 (<.001***) | -34.65 (.051**) | -86.49 (<.001***) |
| Median (N=11) | -19.71 (.123) | -46.73 (.011**) | -40.55 (.001***) |
| Above median (N=11) | 162.85 (.002***) | 17.72 (.220) | 31.43 (.081*) |
| Total (N=41) | -19.10 (.157) | -28.18 (.016**) | -33.06 (.003***) |
| *Utimatum game* | | | |
| Below median (N=21) | 24.14 (.019**) | -18.59 (.032**) | -14.03 (.041**) |
| Median (N=12) | 46.79 (.007***) | -15.41 (.437) | 6.67 (.194) |
| Above median (N=9) | 50.01 (.006***) | 15.77 (.137) | 25.00 (.009***) |
| Total (N=42) | 38.67 (<.001***) | -3.68 (.284) | 2.07 (.345) |
| *Trust game – Low transfer* | | | |
| Below median (N=16) | -100 (<.001***) | -6.00 (.359) | -100.00 (<.001***) |
| Median (N=17) | -10.71 (.248) | -10.71 (.272) | -41.52 (<.001***) |
| Above median (N=9) | 42.31 (.098*) | 14.10 (.279) | -13.28 (.272) |
| Total (N=42) | -24.79 (.045**) | -2.56 (.425) | -49.42 (<.001***) |
| *Trust game – High transfer* | | | |
| Below median (N=15) | -89.40 (<.001***) | -25.40 (.033**) | -89.25 (<.001***) |
| Median (N=5) | -13.04 (.265) | -26.09 (.088*) | -39.39 (.009***) |
| Above median (N=22) | 18.66 (.014**) | 0 (-) | -13.11 (<.001***) |
| Total (N=42) | -22.18 (.014**) | -11.48 (.033**) | -37.30 (<.001***) |

Note: The categories below median, median and above median refer to the actual transfers made in session 2. *** indicate significance at the 1% level, ** at the 5% level, and * at the 10% level.
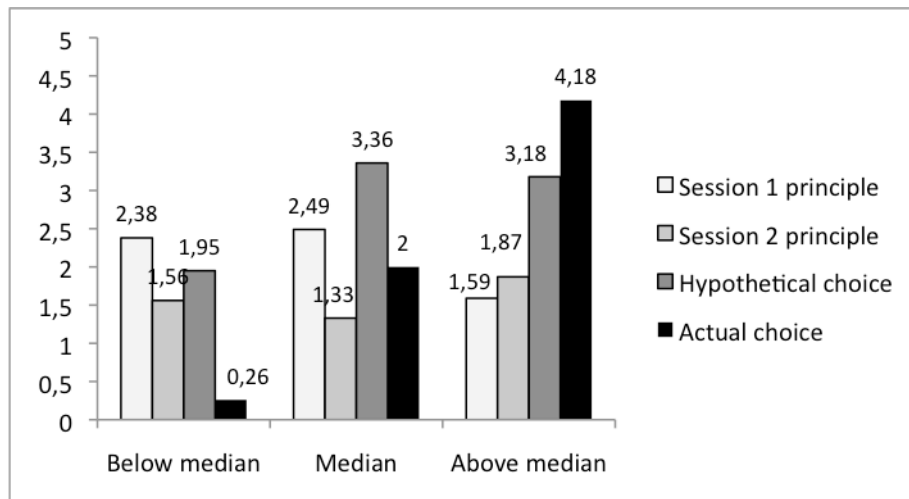
Figure 3. Comparison between principles in sessions 1 and 2, hypothetical and real transfers, by category of actual transfer in the Dictator game (N = 41)
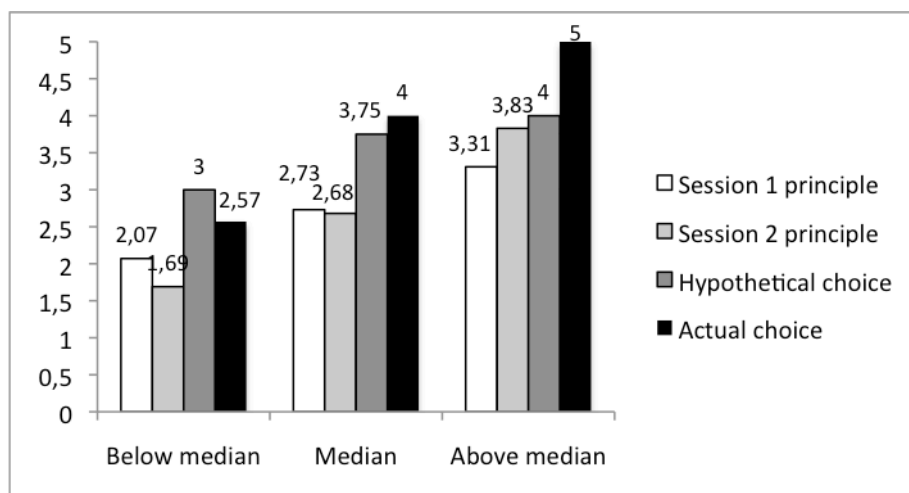


Figure 4. Comparison between principles in sessions 1 and 2, hypothetical and real transfers, by category of actual transfer in the Ultimatum game (N = 42)
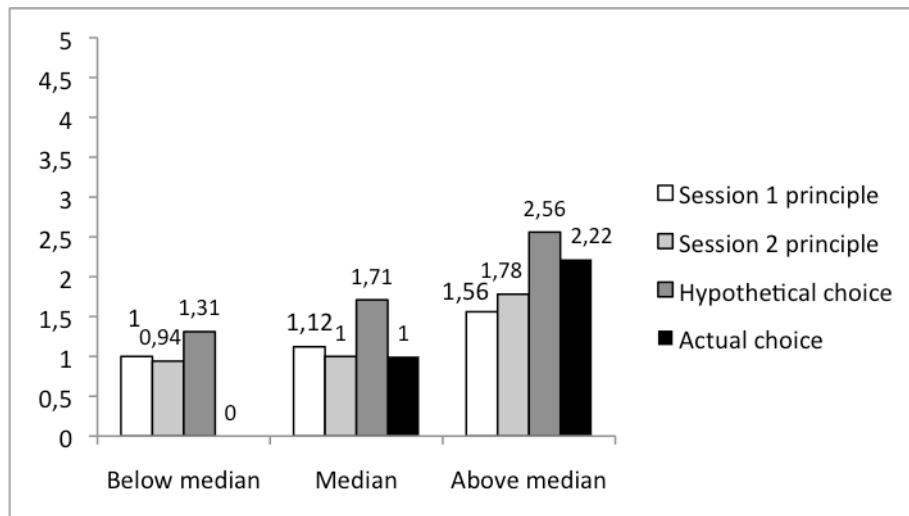
Figure 5. Comparison between principles in sessions 1 and 2, hypothetical and real transfers, by category of actual transfer in the Trust game with a low transfer (N = 42)
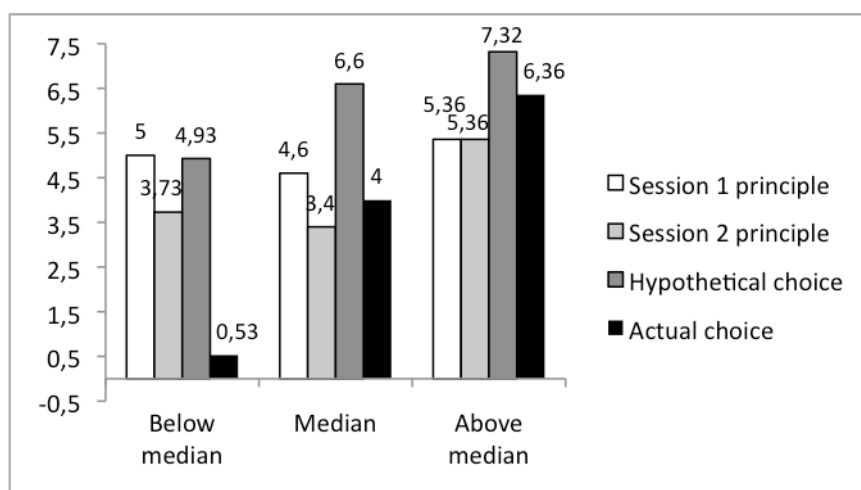


Figure 6. Comparison between principles in sessions 1 and 2, hypothetical and real transfers, by category of actual transfer in the Trust game with a high transfer (N = 42)

**Appendix B. Instructions for the first session of the experiment**

We thank you for participating in this experiment that consists of two sessions. We remind you that you have committed to participate in the two sessions.

During these two sessions, you will be able to earn money. The amount of your earnings depends on your decisions and on the decisions of other participants in this experiment. Your earnings during these two sessions will be added up and paid to you at the end of the second session. This means that the earnings you will make during today's session will be paid to you at the end of the second session.

Throughout the two sessions, we will use points, with the following conversion rate between points and Euros:

1 point = 2 Euros

You will be paid in cash and in private in a separate room by somebody who is not aware of the content of this experiment. No other participant will be informed on your individual payoffs. Your answers will be always kept anonymous and confidential. You will never have to enter your name in the computer.

You have been given a tag indicating the name of a computer and a password. This password is strictly confidential and personal. <u>Do not forget to keep this tag with you and to bring it back with you to be allowed to participate in the second session</u>. If you lose your tag, you will not be allowed to participate in the second session and thus, you will not be paid at all.

Today, you will receive the instructions for the first session only and you will earn 8 Euros. This amount does not depend on your decisions. Please note that the content of the second session will not be affected by your decisions in today's session.

Today, the experiment consists of four independent parts. You have received a set of instructions for the first part. You will receive other sets of instructions once this part will have been completed.

**Part 1**

We ask you to answer to questions related to a scenario. This scenario is the following:

*Imagine that a participant A receives 10 points that he can share with a participant B. A keeps for himself the points he has not transferred to B. B has no decision to make.*

*In this scenario, A earns: 10 points – the amount transferred to B.*

*B earns: the amount transferred by A.*

We ask you to imagine first that you are <u>the participant A</u> and we ask you to answer the following questions:

1) What do <u>you</u> consider as being <u>fair</u> shares between A and B?

A horizontal bar will appear on your screen, together with two cursors, as indicated in the two following examples chosen at random.
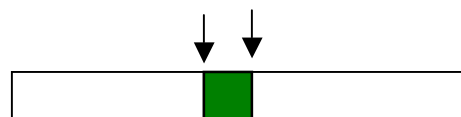
\* You move the left cursor to indicate the share of the amount transferred from A to B from which you consider the shares as being fair.

\* You move the right cursor to indicate the share of the amount transferred from A to B up to which you consider the shares as being fair.

The dark inside area so defined indicates all the shares that you consider as being fair.

*In the first example, all the shares in which A transfers to B at least 28% and at most 89% of the received amount are considered as being fair. In other words, all the shares in which A keeps for himself a maximum of 72% and a minimum of 11% of the received amount are considered as being fair.*

*In the second example, all the shares in which A transfers to B at least 43% and at most 53% of the received amount are considered as being fair. In other words, all the shares in which A keeps for himself a maximum of 57% and a minimum of 47% of the received amount are considered as being fair.*

Next, you validate your answer by pressing the OK button. Once you have pressed this button, the following question will appear on your screen.

2) What do you consider as being unfair shares between A and B?

Here too, a horizontal bar with two cursors will appear on your screen, as indicated in the two following examples chosen at random.

I consider as being unfair all the shares in which A transfers to B less than 8% or more than 83% of the amount.

I consider as being unfair all the shares in which A transfers to B less than 34% or more than 64% of the amount.

\* You move the left cursor to indicate the share of the amount transferred from A to B below which you consider the shares as being unfair.

\* You move the right cursor to indicate the share of the amount transferred from A to B beyond which you consider the shares as being unfair.

The two outside dark areas so defined indicate the set of shares that you consider as being unfair. Then, you validate your choices by pressing the OK button.

*In the first example, all the shares in which A transfers to B less than 8% and those in which he transfers more than 83% of the received amount are considered as being unfair. In other words, all the shares in which A keeps for himself more than 92% and those in which he keeps for himself less than 17% of the received amount are considered as being unfair.*

*In the second example, all the shares in which A transfers to B less than 34% and those in which he transfers more than 64% of the received amount are considered as being unfair. In other words, all the shares in which A keeps for himself more than 66% and those in which he keeps for himself less than 36% of the received amount are considered as being unfair.*

Next, the following question will appear on your screen.

3) What do you think most people consider as being fair shares between A and B?

4) What do you think <u>most people</u> consider as being <u>unfair</u> shares between A and B?

Then, we will ask you the following questions:

5) If <u>you</u> could decide on the share between you and another participant, which amount would you decide to transfer to him?

6) If <u>the other participants</u> in today's session could decide on the share, which amount do you think would be transferred by others on average?

After you have responded to these questions, we will ask you to imagine that you are <u>the participant B</u>. Then, you will answer to the same questions regarding the definition of the fair and unfair shares from your own point of view, and then from the point of view of most people, according to you.


If you have any question regarding these instructions, please raise your hand. We will answer to your questions in private. During the session, communication between participants is strictly forbidden.

Before we start, please answer to the understanding questionnaire. We will check your answers individually. Then, you will enter your password and part 1 will begin.


**Part 2**

(*These instructions were distributed after the completion of part 1*)

We present you with a second scenario.

*Imagine that a participant A receives 10 points that he can share with a participant B. A keeps for himself the points he has not transferred to B.*

*B can accept or reject A's offer. If B rejects A's offer, both A and B earn 0 (all the points are cancelled out). If B accepts A's offer, B earns the amount transferred by A and A keeps for himself the difference between the 10 points he has received and the amount transferred to B.*

*In this scenario, if his offer is accepted, A earns: 10 points – the amount transferred to B.*

*if his offer is rejected, A earns: 0 point.*

*if he has accepted A's offer, B earns: the amount transferred by A*

*if he has rejected A's offer, B earns: 0 point.*

We ask you to imagine first that you are <u>the participant A</u> and we ask you to answer the following questions:

1) What do <u>you</u> consider as being <u>fair</u> shares between A and B?

2) What do <u>you</u> consider as being <u>unfair</u> shares between A and B?

3) What do you think <u>most people</u> consider as being <u>fair</u> shares between A and B?

4) What do you think <u>most people</u> consider as being <u>unfair</u> shares between A and B?

5) If <u>you</u> could decide on the share between you and another participant, which amount would you decide to transfer to him?

6) If <u>the other participants</u> in today's session could decide on the share, which amount do you think would be transferred by others on average?

Then, we will ask you to imagine that you are <u>the participant B</u>. Then, you will answer to the same questions regarding the definition of the fair and unfair shares from your own point of view, and then from the point of view of most people, according to you.

If you have any question regarding these instructions, please raise your hand. We will answer to your questions in private. Before we start, please answer to the understanding questionnaire. We will check your answers individually.

(*The following instructions were distributed after the completion of part 2*)

The following instructions are for part 3 and part 4.

**Part 3**

We present you with a new scenario that takes two versions (scenario 3 and scenario 3 bis).

*Imagine that two participants, A and B, receive 5 points each. A can send to participant B a certain amount, in between 0 and 5 points. A keeps for himself the points he has not transferred to B. Each point sent to B is tripled. Then, B can send points back to A and he keeps for himself the points he has not sent back. In this scenario, A earns: 5 points – the amount transferred to B + the amount sent back by B. B earns: 5 points + 3 times the amount transferred by A - the amount sent back to A.*

**Scenario 3. Imagine that A transfers 1 point to B and keeps 4 points for himself. Thus, B receives 3 points in addition to his initial 5 points. B can send back to A between 0 and 8 points.**

We ask you to imagine first that you are <u>the participant B</u> and we ask you to answer the same questions as in the two previous scenarios regarding the definition of the fair and unfair amounts sent back by B to A, from your own point of view, and then from the point of view of most people, according to you. Then, we will ask you to imagine that you are <u>the participant A</u>. You will again answer to these questions about the definition of the fair and unfair amounts sent back to A.

**Scenario 3 bis. Next, imagine that A transfers 4 points to B and keeps 1 point for himself. This, B receives 12 points in addition to his initial 5 points. B can send back to A between 0 and 17 points.**

We ask you to answer to the same questions as for the scenario 3, first in imagining that you are the participant B and then as the participant A.

**Part 4**

This part will automatically start once part 3 has been completed. It consists of a questionnaire including 60 statements. For each statement, you must choose among five possible options the option that corresponds the most to your opinion.

Choose **SD** (Strongly Disagree) if the statement is absolutely wrong or if you strongly disagree.

Choose **D** (Disagree) if the statement is rather wrong or if you disagree.

Choose **N** (Neutral) if the statement is equally wrong or true or if you cannot choose or if you have no opinion.

Choose **A** (Agree) if the statement is rather true or if you agree.

Choose **SA** (Strongly agree) if the statement is absolutely true or if you strongly agree.

There is no "good" or "bad" answers. The aim of this questionnaire will be reached if you describe yourself and if you express your opinion as exactly as possible.

----

**End of the session**

Once you have completed part 4, a few last questions will appear on your computer screen; then you will receive a message allowing you to leave the lab. Between the beginning of this questionnaire in part 4 and the moment you will be allowed to leave the lab, 10 minutes minimum will have elapsed.

**Last, please give us back the instructions and do not forget to take your tag with your password to be allowed to participate in the second session.**

We thank you for not communicating with anyone about the questions and your answers in this session. **It is indeed very important that nobody else is aware of your answers in this session.**

----

Please read again these instructions. If you have any question, please raise your hand. Before we start part 3, please answer to the understanding questionnaire.

----

**Appendix C. Instructions for the second session**


We thank you for participating in this second session that consists of four parts. Today, you will be able to earn additional payoffs. The amount of your earnings depends on your decisions and on the decisions of other participants. Your earnings will be calculated in points, with the following conversion rate between points and Euros:

1 point = 2 Euros

The payoffs earned in each of the two sessions will be added up and converted into Euros at the end of this session. The total amount of your earnings will remain confidential. We remind you that you will be paid in cash and in private in a separate room by somebody who is not aware of the content of this experiment.

The identity of the participants with whom you will interact during this session will never be communicated to you. Your answers will be kept anonymous and confidential. For this reason, we ask you not to communicate your choices to anybody during or after the experiment.

One of the three first parts of this session will be randomly drawn at the end of the session and you will be paid what you have actually earned in this part. The random draw is independent for each participant. Moreover, you will receive €8 for your participation in today's session that will be added up to the €8 that you have already earned last week.

Before we start the first part, we ask you to enter your password in your computer and to answer to a preliminary question.

<div align="center">

**Part 1**

</div>

You are randomly matched with another participant. One of you is the "participant A", the other is the "participant B". The assignment of roles is random.

**If you are a participant B**, you have no other decision to make during this part. What occurs depends only on participant A.

**If you are a participant A**, you receive 10 points. You must decide on the amount, between 0 and 10 points, that you are willing to transfer to the participant B and you keep the rest for yourself. Once you have made your choice, you press the OK button to validate your choice.

You make your decision once.

The payoffs are determined as follows:

> **The participant A** earns: 10 points – the amount transferred to B
>
> **The participant B** earns: the amount transferred from A

A feedback on the amount transferred by A to B will be given to B only at the end of the session. At the end of the session, you will also be informed on whether this part has been selected for payment in Euros.

Once A has made his decision, both A and B can see a question displayed on their computer screen. A correct answer to this question will allow you to earn €1 more.

We remind you that communication between participants is strictly forbidden. If after reading these instructions again you have any question, please raise your hand. We will answer to your questions in private. Please fill out the understanding questionnaire that has been distributed.


<div align="center">

**Part 2** (*The following instructions were distributed after the completion of part 1*)

</div>

In part 2, you are randomly matched with another participant. This co-participant is likely another person than in the previous part. One of you will be "participant A", the other one will be the "participant B". The assignment of roles is random and independent of the previous part.

<div align="center">

39

</div>

**If you are a participant A**, you receive 10 points. You make an offer to the participant B about the amount, between 0 and 10 points, that you are willing to transfer to him. Once you have made your decision, please press the OK button.

You make your decision once.

**If you are a participant B**, you decide on whether you accept or you reject the offer made by the participant A. However, you will not be informed immediately on the offer made by A. Your computer screen will display all the possible choices made by A and you will have to decide for each possible choice made by A if you accept or you reject it. Once you have made your series of decisions, please press the OK button.

What are the consequences of accepting or rejecting an offer?

If A's offer is rejected, both A and B earn 0 point.

If A's offer is accepted, B earns the amount transferred by A and A keeps for himself the difference between the 10 points he received initially and the amount transferred to B.

At the end of the session, we will match the amount actually offered by A to B and B's decision for this amount. Payoffs are calculated as follows.

---

**Participant A**:

If his offer is accepted, A earns: 10 points – the amount transferred to B

If his offer is rejected, A earns: 0 point.

**Participant B:**

If he has accepted A's offer, B earns: the amount transferred by A

If he has rejected A's offer, B earns: 0 point.

---

At the end of the session, we will inform B about the offer actually made by A and we will inform A about the decision of B for this offer. At the end of the session, you will also be informed on whether this part has been selected for payment in Euros.

Once A and B have made their decisions, both A and B can see a question displayed on their computer screen. A correct answer to this question will allow you to earn €1 more.

If after reading these instructions again you have any question, please raise your hand. We will answer to your questions in private. Please fill out the understanding questionnaire that has been distributed.

**Part 3** (*The following instructions were distributed after the completion of part 2*)

In part 3, you are randomly matched with another participant. This co-participant is likely another person than in the previous parts. One of you will be "participant A", the other one will be the "participant B". The assignment of roles is random and independent of the previous parts.

Both participants A and B receive an initial endowment of 5 points.

**If you are a participant A**, you send to the participant B an amount, comprised in between 0 and 5 points, included, that is taken out of your endowment. Once you have made your decision, please press the OK button.

You make your decision once.

Once you have validated your decision, each point sent to B is tripled.

**If you are a participant B**, you decide on how many points you want to send back to the participant A, between 0 and your total number of points available (i.e. 5points + the tripled amount of points sent by A).

However, you will not be informed immediately on the amount actually sent by A. Your computer screen will display all the possible choices made by A. Then, you will have to decide for each possible amount sent by A how

many points you want to send him.

This means that as a participant B, you must make several decisions regarding the amount you are willing to send back, one for each possible amount sent by A. Once you have made your series of decisions, please press the OK button.

At the end of the session, we will match the amount actually sent by A to B and the corresponding amount sent back by B. Payoffs are calculated as follows.

> **The participant A** earns:  5 points – the amount sent to B + the amount sent back by B
>
> **The participant B** earns: 5 points + 3 times the amount sent by A - the amount sent back to A

At the end of the session, we will inform B about the amount actually sent by A and we will inform A about the corresponding amount actually sent back by B. At the end of the session, you will also be informed on whether this part has been selected for payment in Euros.

Once A and B have made their decisions, both A and B can see a question displayed on their computer screen.  A correct answer to this question will allow you to earn €1 more. We consider an answer as being correct if is exact at 10%.

If after reading these instructions again you have any question, please raise your hand.  We will answer to your questions in private. Please fill out the understanding questionnaire that has been distributed.

-----

**Part 4** (*The following instructions were distributed after the completion of part 3*)

We will present three scenarios on your computer screen successively (one of which has two versions) in which a participant A and a participant B interact together. Each scenario replicates exactly the rules of each of the three previous parts. The scenario 1 corresponds to part 1, the scenario 2 corresponds to part 2 and the scenarios 3 and 3 bis correspond to part 3.

In each scenario, you are requested to imagine that you are the participant A and we ask you the following questions.

1) What do you consider as being fair shares between A and B or fair amounts sent back from B to A (according to the scenario)?

2) What do you consider as being unfair shares between A and B or unfair amounts sent back from B to A (according to the scenario)?

3) What do you think most people consider as being fair shares or fair amounts sent back?

4) What do you think most people consider as being unfair shares or unfair amounts sent back?

Then, we will ask you to imagine that you are the participant B and you will answer to the same questions. In addition, participants B in parts 1 and 2 and participants A in part 3 will have to answer to an additional question in each scenario.

To enter your answers about the fair shares or amounts sent back, a horizontal bar with two cursors will appear on your screen, as indicated in the two following random examples.
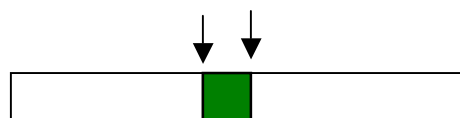
Left cursor        Right cursor                Left cursor   Right cursor

> I consider as being fair all the shares in which A transfers to B at least 28% and at most 89% of the amount.

> I consider as being fair all the shares in which A transfers to B at least 43% and at most 53% of the amount .

41

* You move the left cursor to indicate the share of the amount transferred from A to B or the amount sent back from B to A from which you consider the share or the amount sent back as being fair.

* You move the right cursor to indicate the share of the amount transferred from A to B or the amount sent back from B to A up to which you consider the share or the amount sent back as being fair.

The dark area so defined indicates all the shares or the amount sent back that you consider as being fair.

*In the first example, all the shares in which A transfers to B at least 28% and at most 89% of the received amount are considered as being fair. In other words, all the shares in which A keeps for himself a maximum of 72% and a minimum of 11% of the received amount are considered as being fair.*

*In the second example, all the shares in which A transfers to B at least 43% and at most 53% of the received amount are considered as being fair. In other words, all the shares in which A keeps for himself a maximum of 57% and a minimum of 47% of the received amount are considered as being fair.*

To enter your answers about the <u>unfair</u> shares or amounts sent back, you also use the horizontal bar with the two cursors, as indicated in the two following random examples.



| I consider as being unfair all the shares in which A transfers to B less than 8% or more than 83% of the amount. | I consider as being unfair all the shares in which A transfers to B less than 34% or more than 64% of the amount. |

* You move the left cursor to indicate the share of the amounts transferred from A to B or the amounts sent back by B to A below which you consider the shares or the amounts sent back as being unfair.

* You move the right cursor to indicate the share of the amounts transferred from A to B or the amounts sent back by B to A beyond which you consider the shares or the amounts sent back as being unfair.

The two outside dark areas so defined indicate the set of shares or amounts sent back that you consider as being unfair.

*In the first example, all the shares in which A transfers to B less than 8% and those in which he transfers more than 83% of the received amount are considered as being unfair. In other words, all the shares in which A keeps for himself more than 92% and those in which he keeps for himself less than 17% of the received amount are considered as being unfair.*

*In the second example, all the shares in which A transfers to B less than 34% and those in which he transfers more than 64% of the received amount are considered as being unfair. In other words, all the shares in which A keeps for himself more than 66% and those in which he keeps for himself less than 36% of the received amount are considered as being unfair.*

**End of the experiment**

**At the end of the fourth part,** we will give you a feedback on the actual choice of your co-participant in each of the first three parts and on your associated potential payoffs. Then, we will randomly draw the part that will be used for your actual payment.

After answering last questions, you will be invited to leave the room.