[Preliminary Version, Please Do Not Distribute!] First Show Me the Money: Moral Wiggle Room Reverted

Homayoon Moradi*
Alexander Nesterov†

March 9, 2017

Abstract

We study the dictator game with hidden payoffs that can be voluntarily revealed at zero cost, as in Dana et al. (2007), and vary the type of hidden information and the timing of revelation. In all treatments the dictator makes a choice between two options: a more selfish option and a more pro-social option. The distribution of payoffs for these options is known, but the actual payoffs can be hidden and then revealed by dictators at zero cost at a particular time depending on the treatment. In the *moral wiggle room* treatment, as in Dana et al. (2007), the dictator observes her actual payoff before deciding whether to reveal the receiver's payoff. In the *reversed order treatment*, the dictator observes her own actual payoff after deciding whether to reveal the receiver's payoff. We find that among dictators, 58% choose the selfish option in the moral wiggle room treatment while in the reversed order treatment only 25% do so. In the *self-revelation treatment,* the dictator first observes the receiver's payoff before she can reveal her own payoff. We observe that 34% of dictators choose not to reveal their own payoff, among them 85% choose the pro-social option. These results contradict the explanation in Dana et al. (2007): the observed discrepancies cannot be explained by image concerns. We conjecture that these results can be explained by the fact that the dictators are primed by the information they observe first and, independently, in that the dictators seek to avoid the painful decision between the two options and prefer to remain partly ignorant.

Keywords: Motivated Beliefs, Priming, Fairness, Experiments, Dictator games.

*WZB, Reichpietschufer 50, 10785, Berlin, Germany; e-mail: moradi@wzb.eu
††WZB, Reichpietschufer 50, 10785, Berlin, Germany; e-mail: alexander.nesterov@wzb.eu

1 Introduction

Jason Dana et al. (hereafter, DWK) (2007) provide insightful experimental evidence that the charitable giving in dictator games becomes weaker with the introduction of "moral wiggle room" – ostensible leeway for selfish behavior. The authors conclude that it may not so much be a preference for fair and altruistic outcome *per se* that is the main driver of altruistic behavior, but rather image concern based on a desire to not appear (instead of to not be) unfair, either to oneself or to others. We challenge this conclusion.

The first question we ask is whether moral wiggle room is robust to the order of information revelation. We argue that moral wiggle room depends on how the environment, specifically the payoff information, primes subjects in such a way that image concern alone does not suffice to explains it. The set up of moral wiggle room may primes subjects' preferences for fair and altruistic outcome. Our result shows that decreased charitable givings of moral wiggle room effect fades away in a similar treatment where subjects are not primed by the observation of their own payoff while facing moral wiggle room. To check the robustness of moral wiggle room to the order effect, we compare moral wiggle room treatment in DWK with a reversed order variation, in which dictators decide whether to reveal the receivers' payoff before they see their own payoff (or both payoffs in case they decided to reveal). If the only factor affecting the choosing of selfish actions is the moral wiggle room, we should not see any difference between the moral wiggle room treatment and the reversed order. Assuming that dictators can anticipate the consequences of choosing whether to reveal, they would choose to reveal at the same rate in both treatments. On the contrary, we find that among dictators, 58% choose a selfish option in the moral wiggle room treatment while in the reversed order treatment only 25% do so.

Our second question is whether moral wiggle room can be used selflessly. We investigate further how the underlying environment can affect moral wiggle room to be used as an excuse for selfish uses and even pro-social uses. We also show that there can be moral wiggle room for the self, where subjects stay ignorant of their own payoff for the benefit of others as a form of selfless self-deception.To further investigate the moral wiggle room concept, we provide a situation where subjects can use moral wiggle room in favor of pro-social giving and self loss. In the self-revelation treatment, dictators first observe the payoff of the receiver and then decide whether to reveal their own payoff. We observe that 28% of dictators choose not to reveal their own payoff, among them 80% choose the pro-social option.

Our paper has an implication for the design of information structure for cases with a conflict of interest. In various real-life situations informed experts make suggestions to uninformed customers. If the information gap between the expert and the customer is small, then this suggestion is a mere advice. If the information gap is large then this suggestion becomes a binding choice on behalf of the customer, as in the dictator game. This is, for example, the case

for experts of high specialization like doctors, financial advisers, headhunters and lawyers. When making this binding choice on behalf of the customer, the expert faces several incentives which are not necessarily aligned: the benefit for the customer (quality or fitness of the product) and the benefit for the expert (commission paid for a specific product compared to other products). As in our setting, the information about the quality of the product and the attached commission can be initially hidden and revealed by the expert (e.g., for a doctor prescribing a drug, both his patient diagnosis and the drug producer incentive program might be relevant and initially unknown). Our results suggest that if the information about the commission is revealed first, then it might have a detrimental effect on the expert's incentives to learn the information on the quality of the product and, as a consequence, may lead to poor advice.

A related result is found in Gneezy et al. (2015): an adviser recommending one of two financial lotteries, one of which is incentivized by a commission, can be affected by the choice of time when the commission is revealed. For the customer, the comparison between lotteries is subjective: one has a higher mean and variance than the other. They find that the experts become biased towards the incentivized lottery, but, surprisingly, not biased if they learn about the commission only after they have observed the two lotteries. The authors call this behavior "motivated self-deception" as if experts deceive themselves into believing that the incentivized lottery is better for the customer. They write: "Importantly, this bias in judgment occurs only when judgment is subjective and individuals are able to convince themselves that their behavior is ethical." In contrast to this statement, our findings suggest that the expert is not necessarily deceiving herself when her decision is affected by the order of information revelation. Indeed, in our setting the customer's (receiver) preferences may be hidden, but they are not subjective. Additionally, the expert (dictator) is not surprised by the information about the commission as it is Gneezy et al. (2015), but knows that one option pays better than the other.

2 Model

We first introduce our model of priming and then the model of image concerns that we suppose DWK had in mind to explain their findings.

Dictator game with hidden payoffs

We use a simple dual-self model of altruistic behavior. The utility of the dictator includes both his own payoff X and the receiver's payoff Y: $U = X + \beta Y$, where $\beta \geq 0$ is the dictator's level of altruism. The dictator's payoff $X \in \{\underline{X}, \overline{X}\}$ is always weakly higher than the receiver's payoffs $Y \in \{\underline{Y}, \overline{Y}\}$: $\underline{Y} \leq \overline{Y} \leq \underline{X} \leq \overline{X}\}$.[1]

The timing of the game is the following:

[1] Because our set of possible payoffs is binary for each agent, and because the dictator is at least as well off as the receiver, the simple utility function we use is consistent with a number of models of social preferences and altruism, including Levine (1998), Fehr and Schmidt (1999), Bolton and Ockenfels (2000) and Charness and Rabin (2002).

stage 1 The dictator observes the set of possible games and the probabilities with which each of them is played,

stage 2 The dictator decides whether to reveal the missing information about the game that he is playing,

stage 3 The dictator chooses between two options A and B, then the dictator and the receiver receive the realized payoffs.

Priming and inconsistency

Now we describe the priming mechanism: the way in which the observed information about the payoffs affects the revelation decisions of the dictator. We assume that the dictator is collecting information or ignoring information so as to support his favorable option at the last stage of the game. For example, a more altruistic dictator reveals the receiver's payoffs and possibly does not reveal his own payoffs in order to make a more altruistic choice at the last stage of the game. To be more specific, let us consider the dictator's decisions in a reversed order.

At stage 3, the dictator observes the available information and chooses one of the two options: A or B.

At stage 2, the (possibly primed) dictator decides whether to reveal the missing information about the payoffs. At this point this missing information is relevant for the dictator in two respects: it can be beneficial if his interests coincide with the dictator at stage 3, but it can also be used against him in case their interests differ.

At stage 1, the dictator is primed by the information he receives about the payoffs: knowing his own payoffs makes him less altruistic, while knowing the receipient's payoffs makes him more altruistic. Knowing both payoffs does not leave him primed. This is reflected in the temporary change of β: it becomes either $\beta - \pi^S$ if he is primed to be selfish, or it becomes $\beta + \pi^A$ if he is primed to be "fair," where $\pi^S, \pi^A \in [0, \beta]$. (Here, we assume that the magnitude of priming parameters π^S and π^A are the same for all agents, although they differ in their level of altruism β).

Next we solve the model for the dictator game with complete information and for the three treatments and formulate our hypotheses.

Image concerns

Now we present a simple model of image concerns. The utility function of the dictator is the same as before, but this time there is an additional cost c:

$$U_{IC} = X + \beta Y - c,$$

with which the dictator is punished when facing two options – a more selfish option and a more fair option – and for choosing the selfish option. This is the case only for the dictator game with complete information (as in the baseline

dictator game and also in other treatments when all information is revealed by the dictator). Alternatively, in case one option remains uncertain, the cost is zero. (Again, we assume here that dictators vary in β but have a constant image concerns cost c.)

Model predictions

Next we list the games used in our experiment and what each of the two models predicts for these games. The game trees for these games can be found in the Appendix.

Baseline: Dictator Game

In the baseline treatment, the dictator observes both payoffs: $X(A) = \overline{X}, X(B) = \underline{X}$, $Y(A) = \underline{Y}, Y(B) = \overline{Y}$ (conflicting payoffs). The dictator chooses the "fair" outcome B if his level of altruism is high enough:

$$\beta \geq \beta^* = \frac{\overline{X} - \underline{X}}{\overline{Y} - \underline{Y}}, \tag{2.1}$$

otherwise he chooses the "selfish" outcome A.

In the image concerns model, the dictator chooses the "fair" outcome B if his β exceeds the threshold value $\beta^*_{IC} < \beta^*$:

$$\beta \geq \beta^*_{IC} = \frac{\overline{X} - \underline{X} - c}{\overline{Y} - \underline{Y}}. \tag{2.2}$$

Moral wiggle room

At stage 1, the dictator observes his own payoffs $X(A) = \overline{X}, X(B) = \underline{X}$ and the distribution over the receipient's payoffs:

$$\begin{cases} Y(A) = \overline{Y}, Y(B) = \underline{Y} & \text{(aligned payoffs) with probability } p = 1/2, \\ Y(A) = \underline{Y}, Y(B) = \overline{Y} & \text{(conflicting payoffs) with probability } p = 1/2. \end{cases}$$

At stage 2, the dictator has the altruism level $(\beta - \pi^S)$ and chooses whether to reveal. If he does not reveal, at stage 3 he chooses A with the expected payoffs $\{\overline{X}, \frac{1}{2}(\overline{Y} + \underline{Y})\}$. If he reveals, at stage 3 he chooses B if $\beta \geq \beta^*$ given by 2.1, or chooses A otherwise. Therefore, depending on β, there are three optimal strategies:

$$\begin{cases} \text{if } \beta < \beta^* & \text{then he is indifferent between revealing or not, and always chooses } A, \\ \text{if } \beta^* \leq \beta < \beta^* + \pi^S & \text{then he does not reveal, and then chooses } A, \\ \text{if } \beta \geq \beta^* + \pi^S & \text{then he reveals, and chooses } B. \end{cases}$$

In the image concerns model, only those dictators who are willing to choose the "fair" outcome will reveal. As a result, no dictator suffers from the image concerns costs c; the dictators with $\beta \geq \beta^*$ will reveal and choose "fair," others do not reveal and choose the "selfish" option.

Reversed order

In this treatment the dictator first observes the distribution over his own and the receipient's payoffs:

$$\begin{cases} X(A) = \overline{X}, X(B) = \underline{X} & \text{with probability } p = 1/2, \\ X(A) = \underline{X}, X(B) = \overline{X} & \text{with probability } p = 1/2. \end{cases}$$

$$\begin{cases} Y(A) = \overline{Y}, Y(B) = \underline{Y} & \text{with probability } p = 1/2, \\ Y(A) = \underline{Y}, Y(B) = \overline{Y} & \text{with probability } p = 1/2. \end{cases}$$

Then the dictator decides whether to reveal the receipient's payoffs or not. Then, independent of the previous decision, the dictator observes his own payoffs. After that the dictator chooses between A and B.

Since the dictator is not primed with any information at stage 1, his predicted behavior is as follows:

$$\begin{cases} \text{if } \beta < \beta^* & \text{then he is indifferent between revealing or not, and always chooses the selfish outcome,} \\ \text{if } \beta \geq \beta^* & \text{then he reveals, and chooses the altruistic outcome.} \end{cases}$$

In the image concerns model, as in the previous treatment, only those dictators who choose the "fair" outcome will reveal. Again, the dictators with $\beta \geq \beta^*$ will reveal and choose "fair," others do not reveal and choose the "selfish" option.

Self-revelation

In this treatment the dictator first observes the receipient's payoffs $Y(A) = \underline{Y}, Y(B) = \overline{Y}$ and the distribution over his own payoffs:

$$\begin{cases} X(A) = \overline{X}, X(B) = \underline{X} & \text{(aligned payoffs) with probability } p = 1/2, \\ X(A) = \underline{X}, X(B) = \overline{X} & \text{(conflicting payoffs) with probability } p = 1/2. \end{cases}$$

At stage 2, the dictator has the altruism level $(\beta + \pi^A)$ and chooses whether to reveal. If he does not reveal, at stage 3 he chooses B with the expected payoffs $\{\frac{1}{2}(\overline{X} + \underline{X}), \overline{Y})\}$. If he reveals, at stage 3 he chooses A if $\beta < \beta^*_{BL}$ or chooses A otherwise. Therefore, depending on β, there are three optimal strategies:

$$\begin{cases} \text{if } \beta < \beta^*\text{-}\pi^A & \text{he reveals and chooses } A, \\ \text{if } \beta^* - \pi^A \leq \beta < \beta^* & \text{he does not reveal, and then chooses } B, \\ \text{if } \beta \geq \beta^* & \text{he is indifferent between revealing or not, and chooses } B. \end{cases}$$

In the image concerns model, the prediction is the same as in the baseline dictator game: the dictators with $\beta \geq \beta^*_{IC}$ reveal and choose the "selfish" outcome, others are indifferent between revealing or not since they always choose the "fair" outcome.

Hypotheses

Now, based on the predictions of the two models, we formulate the hypotheses which we test in the experiment.

First we determine the threshold level β^* at which the dictators are indifferent between the two options in the baseline dictator game (as the priming model suggests) or in the game with hidden payoffs (as the image concern model suggests). To compare this level with other studies we use the Fehr-Schmidt utility function $U_{FS} = X - \beta_{FS}(X - Y)$, then the threshold value is $\beta^*_{FS} = \frac{(\overline{X} - \underline{X})}{(\overline{X} - \underline{X} + \overline{Y} - \underline{Y})}$. For the payoffs used in the experiment, $\beta^*_{FS} = 0.2$. Blanco et al. (2011) observed that approximately $25 - 30\%$ of subjects have $\beta < 0.2$, which is very much consistent with the original assumptions in Fehr and Schmidt (1999) and also the results in the current paper.[2]

The first hypothesis is formulated using the image concerns model. It states that the image concerns alone can explain the behavior of the dictators..

H1: $c \gg 0$ – "making the revelation of receiver's payoffs voluntary affects the dictator's decision."

The hypothesis H1 can be rejected by comparing the moral wiggle room treatment and the reversed order treatment.

The next two hypotheses are formulated using the priming model and state that the order of information is irrelevant.

H2: $\pi^S = 0$ – "showing the dictator his own payoffs first does not affect his decision."

The hypothesis H2 can be rejected by comparing the two treatments with hidden receiver's payoffs: where the dictator is primed to be selfish (moral wiggle room) and where the dictators are not primed (the reversed order treatment).

H3: $\pi^A = 0$ – showing the dictator the receiver's payoffs first does not affect his decision

The hypothesis H3 can be rejected by comparing the treatment where the dictator is primed to be altruistic (self-revealation) to one of the treatments where the dictators are not primed.

3 Experimental Design

3.1 The Setting

We study variations of a binary dictator game in which the dictator is uncertain about the payoff of the receiver but can reveal it at no cost. We implemented

[2] A more recent study by Yang et al. (2016) finds a proportion of subjects with $\beta < 0.2$ to be almost twice as high. This might be due to a different design: subjects responded using the strategy method – they made their decisions without knowing the resulting game and their final role (dictator or receiver).

Player X's choices			
	A	X:6	Y:1
	B	X:5	Y:5

Figure 3.1: Baseline dictator game treatment

Player X's choices				
	A	X:6	Y:?	Reveal
	B	X:5	Y:?	

Figure 3.2: Interface for moral wiggle room treatment

the experimental protocol as following. In a simple form binary dictator game, each dictator ("Player X") makes a decision between choosing a selfish option A with €6 payoff for herself and €1 for the receiver (Player Y), or an altruistic option B with €5 payoff for both as depicted in figure 3.1 (we chose neutral symbols in the experiment).

In other variations, player X is ignorant about the precise payoffs but can learn it at no cost. She knows that either the payoffs are conflicting, as in the dictator game above, or payoffs are aligned where there is a choice that maximizes both of them with equal chances. True payoffs would not be revealed publicly, but Player X could reveal them by clicking a button. Player X's decision of whether to reveal would be kept private from Player Y. All subjects also received a €5 show up fee.

The experiment has four treatments. In the baseline dictator game treatment, dictators play a simple binary dictator game. Each dictator decides between option A with payoff of €6 for herself and €1 for the receiver or option B with payoff of €5 for both. In moral wiggle room treatments, a dictator chooses whether to reveal the receiver's payoff after observing her own payoff, and then chooses between A and B, as depicted in Figure 3.2. The reversed order treatment each dictators choose whether to reveal the receiver's payoff before observing her own payoff. Upon deciding to reveal, the receiver's payoff is either 1 in A and 5 in B or reversed with equal chances; and similarly, the dictator's payoff is either 6 in A and 5 in B or reversed with equal chances. Independent of the revelation choice, the dictator will learn her payoff in the next screen. In self-revelation treatment, the dictator first observes the receiver payoff's and is uncertain about her own payoff, in which she can reveal it at zero cost by clicking a button. The dictator's payoff is either 6 in A and 5 in B or reversed with equal chances.

If the only factor affecting choosing the selfish actions is the moral wiggle room, we should not see any difference between the moral wiggle room treatment and the reversed order. Assuming that dictators can anticipate the consequence of choosing whether to reveal, they would choose to reveal at the same rate in both treatments. But if observing the the self-payoff primes dictators to abstain from revealing the receiver's payoff more compare to not observing the self payoff, dictators should reveal and choose the fair option more in the reversed order treatment compare to moral wiggle room treatment.

In the self-revelation treatment, dictators would all choose to reveal if they

Player X's choices			
	A	Y:?	Reveal Y
	B	Y:?	

Figure 3.3: Interface for reverse order treatment

	Revealed			*Not Revealed*		
Player X's choices	A	X:6	Y:1	Player X's choices	A	X:6
	B	X:5	Y:5		B	X:5

Table 1: Interface for reverse order treatment after revelation choices

did not see any harm in learning their own payoff. But if learning their own payoff make dictators worse off then the rate of choosing not to reveal should differ from zero.

3.2 Procedure

We conducted the experiment at the Experimental Laboratory at the Technical University of Berlin from July 2016 to January 2017. Randomization across the three treatments occurred at the participant level. Sessions were gender balanced. Most of the subjects were undergraduate students from the Technical University of Berlin. We did not exclude subjects based on college major or past participation in other experiments. The only exclusion was that subjects must have not participated in this experiment before. Sessions were computerized using a program done with z-Tree (Fischbacher 2007). All experimental sessions were run by at least 16 subjects present.

At the beginning of the experiments subjects drew a token that assigned their role and their seat. Subjects learned that they would be playing with another person in the room with whom they will be matched randomly and anonymously. They also learned that the decision of the dictators (Player X) would determine the amount of the final payments for all the members of this group. Instructions were read aloud.

After receiving instructions describing a generic payoff table, subjects completed a short quiz to make sure that they had understood. Subjects then learned the actual payoffs for the experiment and any other necessary information to describe their particular experiment. We conducted three treatments,

	Left			*Right*		
Player X's choices	A	X:6	Y:1	A	X:6	Y:5
	B	X:5	Y:5	B	X:5	Y:1

Player X's choices				
	A	X:?	Y:1	Reveal
	B	X:?	Y:5	

Figure 3.4: Interface for self-revelation treatment

using a total of 628 students with an average payoff of €9.80 euro for about 20 minutes.

We conducted 28 sessions. In every treatment we conducted four sessions for aligned payoffs and four sessions for conflicting payoffs. Subjects learned that the actual payoff set would not be revealed publicly whereas dictators could learn about that by clicking a button. Subjects understood that dictators choice of revelation would remain unknown to the receiver.

4 Results

Figure 4.1 shows that not observing the own payoffs before deciding to reveal, reverts selfish choices in the moral wiggle room treatment. In the dictator game treatment 29% of dictators choose the selfish option. Consistent with DWK, the rate of selfish choices in the moral wiggle room treatment increases significantly to 58% [$\chi^2(1) = 7$, p = 0.01]. The rate of selfish choices, however, decreases significantly to 33% in the reversed order treatment [$\chi^2(1) = 4$, p = 0.04]. The result in figure 4.1 are therefore inconsistent with hypothesis H1, that getting primed with self-payoff has no effect. The same result allows us to reject hypothesis H2 for the priming model: priming with own payoffs is significant.

Figure 4.1 also shows that not observing the own payoffs before deciding to reveal also increases the revelation choices. In moral wiggle room 66% of dictators choose to reveal the true payoff. The revelation rate increases significantly to 85% in the reversed order treatment [$\chi^2(1) = 7$, p = 0.008]. Higher rate of revealing in the reversed order leads to a reverted rate of selfish choices in the moral wiggle room since there is no significant difference on the rate of selfish choices when they choose to reveal between treatments. We provide the detailed results of treatments in the Appendix.

There are two possible explanations for the result. One possible explanation for the result might be that dictators cannot strategically anticipate that revealing the receiver's payoff can be costly for them since they cannot use their ignorance to choose the selfish option. The reversed order treatment therefore makes it more difficult for subjects to anticipate the consequences of their revelation. We believe this cannot be the case since we check subjects' answers to a quiz that confirms their understanding of the direct consequence of revealing and not revealing in both treatments.

Another possible explanation is that observing the own payoff primes subjects to be more concerned about their own earning and less concerned about the receiver's payoff. The idea of getting primed by the initial information is the gist of our simple dual self model.

Figure 4.2 shows that in the self-revelation treatment, a significant number of dictators choose not to reveal their own payoff. Even though there is no significant difference of altruistic choices between the dictator game and the self-revelation treatment, 34% of subjects choose to not reveal their own payoff. From the 34% of dictators who would rather stay ignorant of their own payoff, 85% choose the altruistic option. In the self-revelation treatment, therefore,

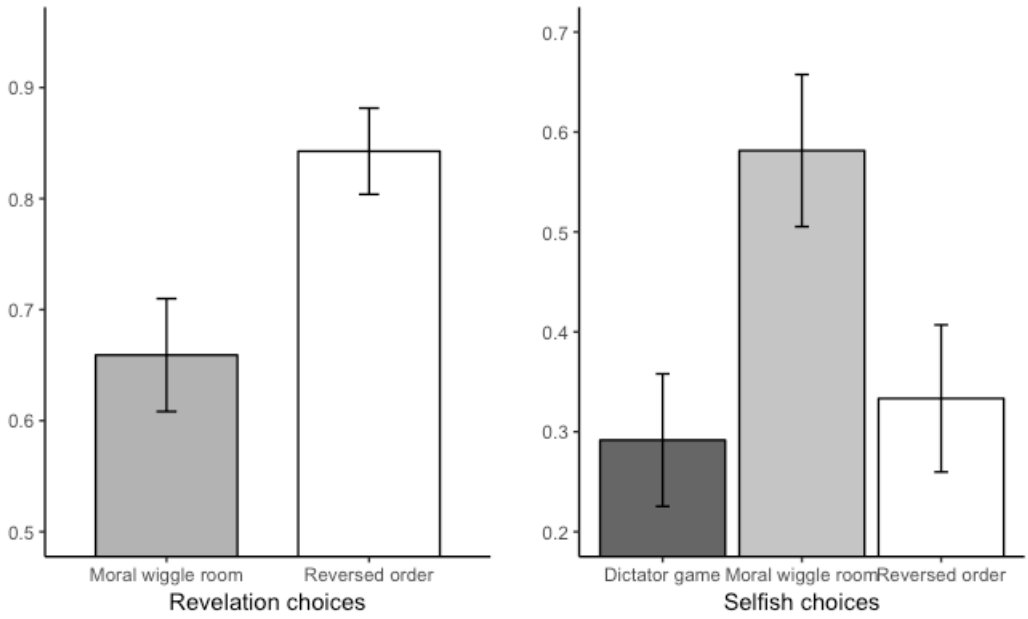

Figure 4.1: Fraction of dictators choices in the reversed order, moral wiggle room, and dictator game treatments.
The error bars represent $+/-1$ standard error.

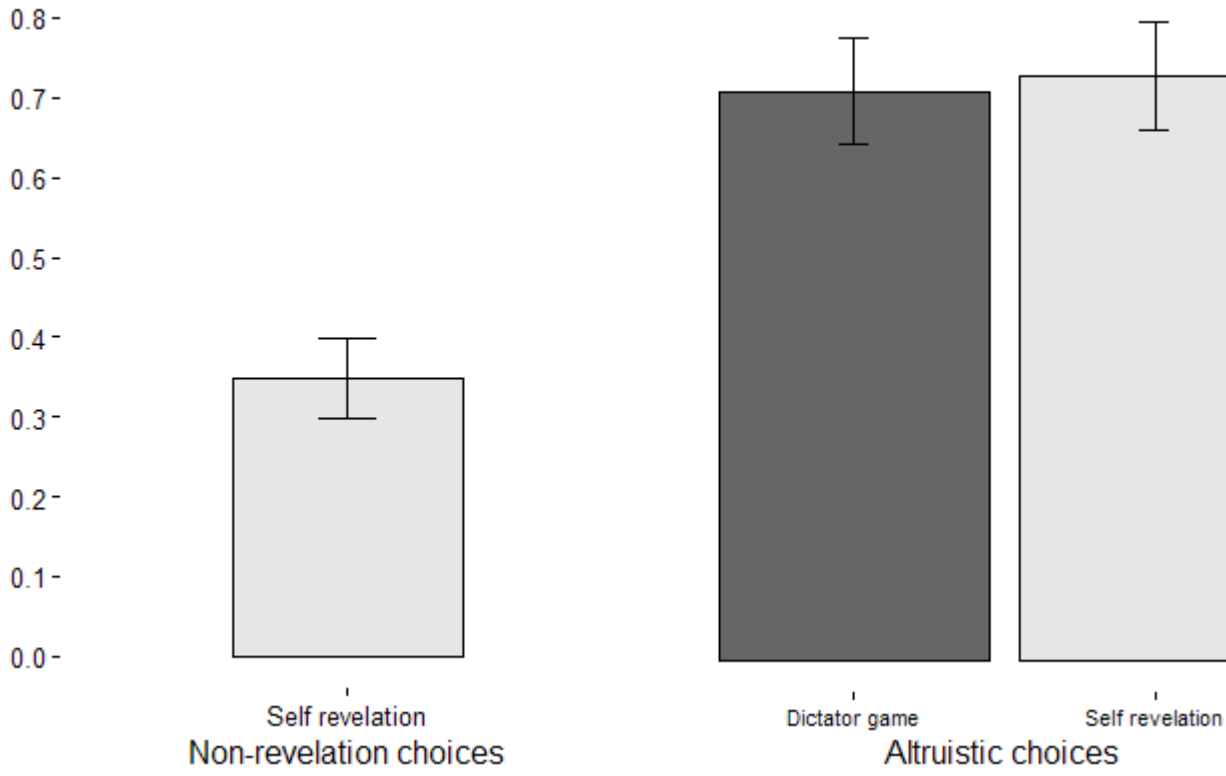

Figure 4.2: Fraction of dictators choices in the self-revelation and dictator game treatments.
. The error bars represent $+/-1$ standard error.

29% of subjects would rather to stay ignorant of their own payoff while they choose an altruistic option. These results do not allow us to reject hypothesis H3 that there was any significant priming with the receiver's payoff.

We imagine three possibilities to explain the not revealing own payoff result. First, a dictator might choose to not reveal her own payoff as a selfless self-deception. She knows that revealing her own payoff might tempt her to not choose altruistically. By not revealing, therefore, she deceives herself into choose an altruistic option. Second, for one who is certain about choosing the altruistic option independent of whether payoffs are conflicting as in the baseline or aligned, choosing the altruistic option might make him appear more altruistic rather than observing that payoffs are aligned where now the altruistic choice is also a selfish choice too. This explanation is in line with DWK, that dictators do not like to appear selfish. Third, primed with the receiver's payoff, a dictator

is perhaps less concerned with his payoff since she knows that there is chance of receiving conflicting information from revealing her own payoff. Inline with the dual self model, the dictator prefers to not reveal and to choose the altruistic option.

5 Conclusion

DWK suggests that altruistic behavior is not so much driven by the preferences *per se*, but rather by image concerns and the desire to not appear selfish. In the present experiments we allow such strategic ignorance but we vary two thing. First we vary the order of choices to stay ignorant and the initial priming information. Second we vary the source of priming information regarding self versus others payoff. We are thus able to provide a pure test for both for robustness of moral wiggle room effect in two different dimentions. It turns out that image concerns does not seem to be the reason behind the manipulability of behavior in the dictator game. A simple model of dual self can rationalize most of the data.

With free access to the full information, the order of receiving the information plays a significant role in the specific context of the dictator game. The dictator game resembling various daily life situations, raises the open question of whether the trivial change of order impacts on relevant decision makings.

References

1. Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. Games and Economic Behavior, 72(2), 321-338.

2. Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. American economic review, 166-193.

3. Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. Quarterly journal of Economics, 817-869.

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

5. Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10(2), 171–178.

6. Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. Quarterly journal of Economics, 817-868. ISO 690

7. Gneezy, U., Saccardo, S., Serra-Garcia, M., & van Veldhuizen (2015), R. Motivated self-deception, identity, and unethical behavior. *Mimeo.*

8. Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. Review of economic dynamics, 1(3), 593-622.

9. Yang, Y., Onderstal, S., & Schram, A. (2016). Inequity aversion revisited. Journal of Economic Psychology, 54, 1-16.

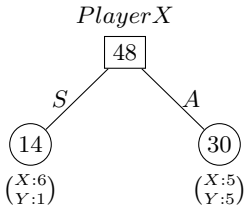

Figure 6.1: Dictator game
Notes: Player X observes all payoffs and chooses the action: either Selfish (S), i.e, maximizing his own payoff, or Altruistic (A), i.e, maximizing Player Y's payoff. The number inside each node shows the number of subjects.

6 Appendix

6.1 Game trees with results

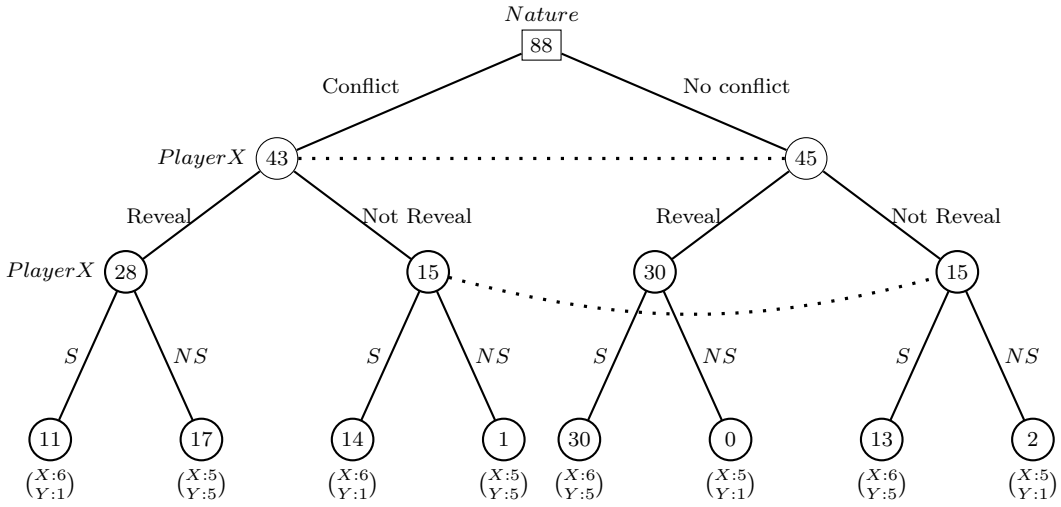

Figure 6.2: Moral wiggle room
Notes: First Nature chooses the payoff structure, then Player X observes his own payoffs and chooses whether to reveal Player Y's payoffs, and then chooses the action: either Selfish (S), i.e, maximizing his own payoff, or Not Selfish (NS). The number inside each node shows the number of subjects.

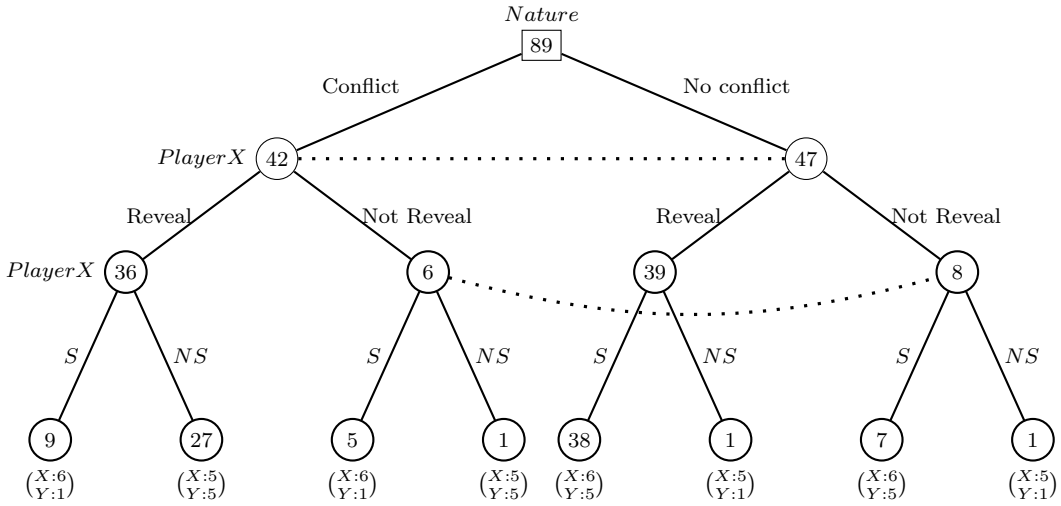

Figure 6.3: Reversed order
Notes: First Nature chooses the payoff structure, then Player X chooses whether to reveal Player Y's payoffs, then Player X observes his own payoff, and then chooses the action: either Fair (S), i.e, maximizing own payoff, or Not Selfish (NS). The number inside each node shows the number of subjects.

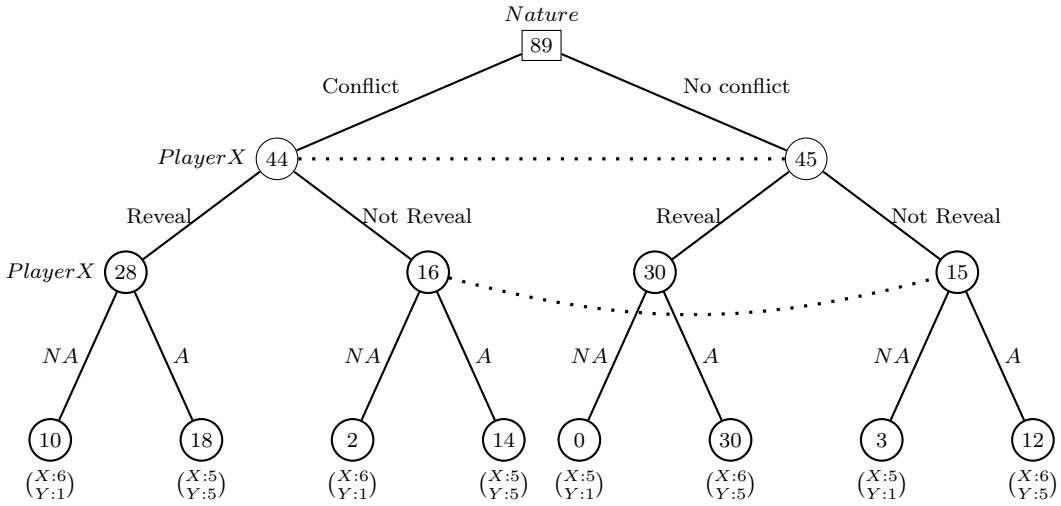

Figure 6.4: Self-revelation
Notes: First Nature chooses the payoff structure, then Player X observes Player Y's payoffs and chooses whether to reveal his own payoffs, and then chooses the action: either Altruistic (A), i.e, maximizing the Player Y's payoff, or Not Altruistic (NA). The number inside each node shows the number of subjects.

6.2 Instructions

All Conditions

This is an experiment in the economics of decision-making. You will be paid for your participation in the experiment. The exact amount you will be paid will depend on your and/or others' decisions. Your payment will consist of the amount you accumulate plus a €5 participation bonus. You will be paid privately in cash at the conclusion of the experiment. If you have a question during the experiment, raise your hand and an experimenter will assist you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. Please put away all outside materials (such as book bags, notebooks) before starting the experiment. Participants violating the rules will be asked to leave the experiment and will not be paid.

In this experiment, each of you will play a game with one other person in the room. Before playing, we will randomly match people into pairs. The grouping will be anonymous, meaning that no one will ever know which person in the room they played with. Each of you will be randomly assigned a role in this game. Your role will be player X or player Y. This role will also be kept anonymous. The difference between these roles will be described below. Thus, exactly one half of you will be a Player X and one half a Player Y. Also, each of you will be in a pair that includes exactly one of each of these types. The game your pair will play will be like the one pictured below. Player X will choose one

of two options: "A" or "B". Player Y will not make any choice. Both players will receive payments based on the choice of Player X. The numbers in the table are the payments players receive. The payments in this table were chosen only to demonstrate how the game works. In the actual game, the payments will be different. For example, if player X chooses "B", then we should look in the right square for the earnings. Here, Player X receives 3 euros and Player Y receives 4 euros. Note that player X's payment is in the lower-left corner of the square, player Y's payment is in the upper-right corner.

Player X's choices			
	A	X:1	Y:2
	B	X:3	Y:4

At this point, to make sure that everyone understands the game, please answer the following questions:

In this example, if Player X chooses "B" then:
Player X receives __
Player Y receives __
In this example, if Player X chooses "A" then:
Player X receives __
Player Y receives __
<answers read aloud>

[MORAL WIGGLE ROOM TREATMENT]

The actual game you will play will be one of the two pictured below. Note that both games are the same except that Player Y's payments are flipped between the two. Note that in both games, Player X gets his or her highest payment of €6 by choosing A. In the game on the left, this gives Player Y his or her lowest payment of €1. In the game on the right this gives Player Y his or her highest payment of €5. In both games, if Player X chooses B, he or she gets a lower payment of €5. In the game on the left, this gives Player Y the highest payment of €5. In the game on the right, this gives Player Y the lowest payment of €1.

	Left			*Right*		
Player X's choices	A	X:6	Y:1	A	X:6	Y:5
	B	X:5	Y:5	B	X:5	Y:1

You do not know which of the games you will be playing. However, note that for Player X, the payments will be identical. The only thing that differs is the payments for Player Y. The actual game you will play was determined by a coin flip before the experiment. However, we will not reveal publicly which game you are actually playing. Before playing, Player X can choose to find out which game is being played, if they want to do so, by clicking a button. This choice will be anonymous, thus Player Y will not know if X knows which game is being played. Player X is not required to find out and may choose not to do so. When the game ends, we will pay each player privately.

Player X's choices				
	A	X:6	Y:?	Reveal
	B	X:5	Y:?	

At this point, to make sure that everyone understands the game, please answer the following questions:

In both games, which action gives player X his or her highest payment of €6?__

If Player X chooses B, then Player Y receives __

1. €5
2. €1
3. either €5 or €1

[Reversed Order TREATMENT]

The actual game you will play will be one of the four pictured below. Note that TOP AND BOTTOM column games are the same except that Player X's payments are flipped between the two. Similarly, LEFT and RIGHT row games are the same except that Player Y's payments are flipped between the two.

Note that in games in TOP, Player X gets his or her highest payment of €6 by choosing A. In the TOP LEFT game, this gives Player Y's lowest payment of €1, and in the game TOP RIGHT, highest payment of €5. Note that in games in TOP, if Player X chooses B, he or she gets a lower payment of €5. In the game TOP LEFT, this gives Player Y the highest payment of €5, and In the TOP RIGHT game, the lowest payment of €1.

Note that in games in BOTTOM, Player X gets his or her highest payment of €6 by choosing B. In the BOTTOM LEFT game, this gives Player Y's highest payment of €5, and in the game BOTTOM RIGHT, lowest payment of €1. In games in BOTTOM, if Player X chooses A, he or she gets a lower payment of €5. In the game BOTTOM LEFT, this gives Player Y the lowest payment of €1, and In the BOTTOM RIGHT game, the highest payment of €5.

TopLeft

Player X's choices			
	A	X:6	Y:1
	B	X:5	Y:5

TopRight

A	X:6	Y:5
B	X:5	Y:1

BottomLeft

Player X's choices			
	A	X:5	Y:1
	B	X:6	Y:5

BottomRight

A	X:5	5
B	X:6	1

You do not know which of the games you will be playing. The actual game you will play was determined by two coin flips (one for TOP vs BOTTOM, and one for LEFT vs. RIGHT) before the experiment. However, we will not reveal publicly which game you are actually playing.

Before playing, Player X can choose to find out which games from LEFT and RIGHT is being played, if they want to do so, by clicking a "Reveal Player Y's Payoff" button. Note that for Player X, the payments will be identical. The only thing that differs is the payments for Player Y. This choice will be anonymous; thus Player Y will not know if X knows which game is being played. Player X is not required to find out and may choose not to do so by clicking on the "Continue" button. After deciding to reveal or not, Player X will be informed which game(s) from TOP and BOTTOM is being played. This is independent of his or her actions. When the game ends, we will pay each player privately.

Player X's choices		
	A	Y:?
	B	Y:?

Reveal Y

At this point, to make sure that everyone understands the game, please answer the following questions:

In TOP games, which action gives player X his or her highest payment of €6? __

In TOP games , if Player X chooses B, then Player Y receives __

1. €5
2. €1
3. either €5 or €1

When player X clicks on "Reveal" button, his final payoff table contains information about ...

1. Only player X's payoff
2. Only player Y's payoff
3. Both players' payoff

When player X does not click on "Reveal" button, his final payoff table contains information about ...

1. Only player X's payoff
2. Only player Y's payoff
3. Both players' payoff

[SELF-REVEALATION TREATMENT]

The actual game you will play will be one of the two pictured below. Note that both games are the same except that Player X's payments are flipped between the two. Note that Player X will get his or her highest payment of €6 by choosing A in game LEFT and choosing B in game RIGHT, otherwise Player X will get his or her lowest payoff of €5. Note that in both games, Player Y gets his or her highest payment of €5 if B is chosen and his or her lowest payment of €1 if A is chosen.

	Left			*Right*		
Player X's choices	A	X:6	Y:1	A	X:6	Y:5
	B	X:5	Y:5	B	X:5	Y:1

You do not know which of the games you will be playing. The only thing that differs is the payments for Player X. However, note that for Player Y, the payments will be identical.

The actual game you will play was determined by a coin flip before the experiment. However, we will not reveal publicly which game you are actually playing. Before playing, Player X can choose to find out which game is being

played, if they want to do so, by clicking a button. This choice will be anonymous, thus Player Y will not know if X knows which game is being played. Player X is not required to find out and may choose not to do so. When the game ends, we will pay each player privately.

Player X's choices				
	A	X:?	Y:1	Reveal
	B	X:?	Y:5	

At this point, to make sure that everyone understands the game, please answer the following questions:

In game LEFT, which action gives player X his or her highest payment of €6? _ _

In game RIGHT, which action gives player X his or her highest payment of €6? _ _

If Player X chooses B, then Player Y receives _ _

1. €5
2. €1
3. either €5 or €1