WORKING PAPERS

N° 1531

November 2025

# "Does a veil of ignorance trigger the inner Kantian in us?"

Ingela Alger and José Ignacio Rivero-Wildemauwe

Toulouse School of Economics

# Does a veil of ignorance trigger the inner Kantian in us? *

Ingela Alger† and José Ignacio Rivero-Wildemauwe‡

15th November 2025

**Abstract:** Theoretical work shows that in social interactions Kantian (or universalization) concerns sometimes yield starkly different behavioral predictions than other-regarding motives. We hypothesize that in interactions with different roles, Kantian concerns are awakened through awareness of a possible role reversal. An experiment varying the salience of role uncertainty is conducted to test this hypothesis, and to disentangle Kantian concerns from other-regard. Participants decide whether to "sell a lemon" to a willing buyer (akin to a Dictator Game with taking), either behind the veil of ignorance as to the role distribution, or after simply being informed of their role as "seller". Based on the old hypothesis that markets promote selfishness, we also vary the wording used to describe decisions: we implement a Market frame and a Neutral frame. Role uncertainty salience does promote pro-social behaviors (in both frames), and selfish behavior is enhanced under the Market frame. Structural estimates of the preference parameters indicate that this is driven by a reduced other-regard under the Market frame, as moral concerns are of roughly the same intensity under both frames.

**JEL codes**: C91, D01, D91

**Keywords:** Kantian concerns, social preferences, market framing, lemons, experiment

*I.A. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 789111 - ERC EvolvingEconomics), as well as IAST funding from the French National Research Agency (ANR) under grant ANR-17-EURE-0010 (Investissements d'Avenir program). We are grateful to Lony Bessagnet, Jean-François Bonnefon, Eric Danan, Pierre Fleckinger, Peter Hammond, Astrid Hopfensitz, Nicolas Jacquemet, Pau Juan-Bartroli, Fabrice Le Lec, Thierry Magnac, Pauline Morault, Esteban Muñoz-Sobrado, Sébastien Pouget, Régis Renault, Angelo Secchi, Sigrid Suetens, Ao Wang and Boris van Leeuwen for helpful discussions, and to Stéphane Cézera for invaluable assistance in the lab. We thank audiences at BSE Summer Forum, ESA (Helsinki), FUR (Brisbane), SAET (Ischia), SEET (Dijon), SPUDM (Lucca), and Paris School of Economics Summer School, and at seminars at Universidad Católica del Uruguay, CY Cergy Paris University, Duke University, IAST, ITAM, Max Planck Institute, Toulouse School of Economics, Universidad de la República, University of Gothenburg, and Uruguayan Central Bank for feedback.

†Toulouse School of Economics, CNRS, IAST, and CEPR. ingela.alger@tse-fr.eu

‡Department of Economics and Business, Universidad Católica del Uruguay, Montevideo, Uruguay. joseignacio.rivero@ucu.edu.uy

# 1 Introduction

Economists are increasingly interested in how preferences come about, acknowledging that a host of factors – ranging from social ones (Bowles (1998); Heckman & Corbin (2016)) to physical ones such as resource constraints in the brain (Glimcher (2022); Engelmann, Meyer, Ruff, & Fehr (2019)) – likely play a role. Several theories have been proposed (see, e.g., Akerlof & Kranton (2000); Bernheim, Braghieri, Martínez-Marquina, & Zuckerman (2021); Nagler (2023); Boissonnet, Ghersengorin, & Gleyze (2023); Pivato (2023)). One strand of the theoretical literature focuses on the intra-generational transmission of preferences, the question being which preferences, if any, are favored.[1] In this literature, a sharp prediction was obtained by Alger, Weibull, & Lehmann (2020) for preferences that guide behavior in strategic interactions involving small material payoffs: evolutionary forces should favor preferences that combine material self-interest, a Kantian moral concern, and an other-regarding concern.[2] The other-regarding concern is consequentialistic: it makes the individual weigh the effect of each strategy on the material payoffs of others. By contrast, the Kantian concern makes the individual evaluate each strategy in the counterfactual scenario where this strategy would be universalized to the others. The two concerns sometimes yield the same behavioral predictions, but they can also differ starkly depending on the context (Alger & Weibull (2013, 2017); Sarkisian (2021); Eichner & Pethig (2022); Muñoz-Sobrado (2022); Laslier (2023); Rivero-Wildemauwe (2023, 2025); Juan-Bartroli & Karagözoğlu (2024)). In light of these results, it is important to evaluate the empirical relevance of Kantian concerns in different kinds of interactions. We conduct a laboratory experiment to contribute to this quest. A first objective of the experiment is thus to disentangle the role of Kantian concerns from other-regarding (or distributional) preferences, and to evaluate their relative importance.

A second objective of our experiment is to evaluate if preferences appear to depend on context. This is in line with the theoretical model by Alger et al. (2020), which predicts

[1] For recent surveys, see Alger & Weibull (2019), Alger (2023), and Bisin & Verdier (2023). Such transmission can be cultural and/or biological. There is evidence of biological transmission of trust, risk attitudes and time preferences based on twin studies; see, e.g., Kettlewell & Tymula (2021) and Kettlewell & Tymula (2024).

[2] When preferences are expressed with reproductive success rather than trivial material payoffs as arguments, Alger, Weibull, & Lehmann (2020) predict that preferences are expected to combine self-interest with a Kantian concern, confirming the result of Alger & Weibull (2013).

that evolution may have led to different weights attached to the three concerns (self-interest, other-regard, and the Kantian moral concern), depending on exactly how preferences are transmitted (for example, genetic transmission may give rise to sharply different weights than cultural transmission). Because the vast majority of economic experiments use the same, neutral, wording to describe tasks presented to subjects, we include such wording here. However, since interactions with such wording were not part of our evolutionary past, we cannot deliver any specific prediction for the preference weights. Inspired by the idea that market interactions may promote material self-interest, which has been present in the economics literature for decades (Hirschman (1982); Bowles (1998)) for half of the decisions we describe the task as a market transaction. We test whether the intensity of Kantian concerns and other regard is different when subjects make decisions in interactions described using "market" and neutral wording.

Subjects in our experiment face a series of anonymous binary choices, each of which affects the payoff of a randomly drawn other subject, hence avoiding that decisions be influenced by strategic or repeated interaction considerations, (Roth & Schoumaker, 1983; Andreoni & Miller, 1993), and social image or reciprocity concerns (Bénabou & Tirole, 2006; D. K. Levine, 1998; Charness & Rabin, 2002). In each binary choice one of the options – the *selfish option* – entails a gain for self and a loss for the other, compared to the other option – the *status quo* option. In other words, each decision situation is a Dictator Game, with taking rather than giving (Dreber, Ellingsen, Johannesson, & Rand, 2013). The choice thus resembles that of a seller deciding on whether to sell a lemon to a willing buyer, where the selfish option amounts to trading an object of low quality at such a price that she gets better off but the buyer gets worse off, while the *status quo* option amounts to the seller keeping the object.[3]

Since the selfish option increases the payoff gap between the decision-maker and the passive subject, the *status quo* may be selected due to the other-regarding concern in the form of altruism or inequity aversion, formalized as a weight $\beta$ on the payoff gap (Becker, 1974; Fehr & Schmidt, 1999). However, the *status quo* option could also be chosen as the result of the Kantian concern, which makes the subject evaluate each decision in the light of what her payoff would be, if – hypothetically – the subject's own strategy was universalized

[3]While it would be even more realistic to let the buyer decide whether or not to buy, arguably the dictator situation captures the instant where the seller and buyer have already agreed on a price, and the seller can still decide to keep the defective object rather than selling it.

if the roles were reversed.

Our key hypothesis is that the attention paid to the counterfactual universalization scenario may depend on the salience of the possibility of role reversal. Therefore, the experimental design varies this salience in order to assess the presence of Kantian moral motivations. Specifically, for some decisions the subjects are told that they stand an equal chance of being the decision-maker and the passive subject. They are thus put behind a *veil of ignorance* (VOI). For the remaining decisions, subjects are simply told that they are the decision-makers (non-VOI). In reality, however, both situations are the same, since upon entering the experiment any participant effectively stands an equal chance of being the decision-maker and the passive subject in any given match. Hence, the only difference is that we make this role uncertainty explicit in VOI decisions.

Our theory predicts that if subjects are driven by some Kantian concern and if they are not fully aware of the arbitrariness in the role distribution in the non-VOI decisions, they are more likely to sell (i.e., choose the *selfish option*) in non-VOI decisions than in VOI decisions. We test this hypothesis, and we also structurally estimate the contribution of the other-regarding and the Kantian motivation to the subjects' decisions.

To achieve our second objective, we let half of the decisions be taken in a standard *neutral frame* and the other half in a *market frame*, keeping the payoff consequences identical under both frames. Only the wording used to describe the situation differs. Specifically, in the market frame subjects get the following information: "As you can see on the decision screen above, the Buyer would be better off if you chose Not Sell, while you are better off if you Sell. Think of this as representing a situation in which the good that you sell has a defect which makes the Buyer enjoy owning the good less than you do." The wording was chosen to mimic a situation in which a seller can sell a "lemon" to an uninformed buyer (who in the experiment is passive). In the neutral frame the options are simply referred to as X and Y, and no context is given.

We find that, in line with our theoretical prediction, subjects select the selfish option significantly more often in non-VOI than in VOI decision situations. We further find that, on average, subjects are significantly more prone to selecting the selfish option in the market frame compared to the neutral frame. Comparing the effect of the VOI condition on the propensity to select the selfish decision between the market frame and the neutral frame, the magnitude is larger in the market frame in absolute terms, but not in relative terms. The effects of the market frame and of the VOI treatment are larger in magnitude in the within-subjects regressions, suggesting significant heterogeneity across subjects.

To obtain a better understanding of the potential motivations driving behavior, we also structurally estimate the preference parameters $\beta$ and $\kappa$, in line with several earlier experimental studies (e.g., Fisman, Kariv, & Markovits (2007); Bruhin, Fehr, & Schunk (2019); van Leeuwen & Alger (2024)). A novelty of our study is that we obtain estimates of the preference parameters in two different frames: the market frame and the neutral frame. Overall, we find notable differences between the market and neutral frames in terms of aheadness aversion, with participants in the market frame exhibiting significantly lower levels of other-regard. In contrast, Kantian morality estimates are positive and statistically significant in both frames and remain relatively stable across them. This is in line with our treatment effect estimates which suggest that the VOI condition does not have a larger relative impact when decisions are described using market rather than neutral wording. Our results thus suggest that behavioral differences between both frames are mostly attributable to different degrees of aheadness aversion.

Our study contributes to three strands of the literature. First, it complements a recent set of experiments that have sought to detect moral concerns as drivers of behavior, where moral concerns are distinct from distributional preferences (Capraro & Rand, 2018; Bursztyn, Fiorin, Gottlieb, & Kanz, 2019; Miettinen, Kosfeld, Fehr, & Weibull, 2020; Chen & Schonger, 2022; Feess, Kerzenmacher, & Timofeyev, 2022; Bénabou, Falk, Henkel, & Tirole, 2023; Bénabou, Falk, & Henkel, 2024; van Leeuwen & Alger, 2024). The most closely related experiments are those by Miettinen, Kosfeld, Fehr, & Weibull (2020) and van Leeuwen & Alger (2024), who also posit a utility function with a Kantian moral concern à la *Homo moralis* (Alger & Weibull, 2013), and who also seek to disentangle this concern from distributional preferences. Our study differs from both of them in two important ways. First, while they only use neutral wording, our experimental design allows us to test whether preferences differ between the neutral and the market frame. Second, they rely exclusively on what we call VOI decisions; our comparison of decisions in VOI and non-VOI decisions gives an indication about the extent to which the explicit mention of the role uncertainty helps trigger the Kantian moral concern.

This last point brings us to the second literature to which our study contributes: that which examines if decisions depend on whether they are taken after or before subjects learn the role distribution. Several studies find that decisions differ significantly, while other studies find mixed or no effects (see the surveys by Brandts & Charness (2011) and

Grech, Nax, & Soos (2022)). In the studies where an effect appears, decisions taken behind the veil of ignorance tend to be more pro-social on average (Sutter & Weck-Hannemann, 2003; Iriberri & Rey-Biel, 2011; Huang, Greene, & Bazerman, 2019; S. Levine, Kleiman-Weiner, Schulz, Tenenbaum, & Cushman, 2020; García-Pola, Iriberri, & Kovářík, 2020; Ortiz-Riomalo, Koessler, & Engel, 2021; Herne, Hietanen, Lappalainen, & Palosaari, 2022; Mesa-Vázquez, Rodriguez-Lara, & Urbano, 2021). Our findings are in line with this empirical regularity. But while previous studies have sought to explain the greater pro-sociality in decisions taken under role uncertainty by referring mainly to aspects of the decision-making process (e.g., number of decisions taken, complexity of the information presented, likelihood that emotions are triggered by the task), our experiment relies on a preference-based theory which predicts and thus explains it. According to our theory, the explicit mention of role uncertainty awakens a Kantian moral concern, which triggers pro-social behavior in our dictator game design.[4]

Finally, our study adds insights into an old question: do market interactions render people more selfish (Hirschman, 1982; Bowles, 1998)? Some authors have relied on lab-in-the-field experiments to compare the behavior of individuals with varying degrees of exposure to market interactions (e.g., Henrich et al. (2005, 2016) and Agneman & Chevrot-Bianco (2022)), while others have used laboratory or online experiments to examine how variation in the competition between subjects or other institutional factors affects behaviors (e.g., Cabrales, Miniaci, Piovesan, & Ponti (2010); Falk & Szech (2013); Huck, Lünser, & Tyran (2016); Sutter, Huber, Kirchler, Stefan, & Walzl (2020); Engel & Szech (2020); Dufwenberg, Johansson-Stenman, Kirchler, Lindner, & Schwaiger (2022); Bartling & Özdemir (2023); Byambadalai, Ma, & Wiesen (2023)). Our experiment explores a different possibility, by comparing behavior when the task is described as the sale of a lemon (recall the wording above), to behavior when it is described in neutral terms, the objective being to quantify the effect of the market frame (for a recent discussion of the use of frames in economic experiments, see Alekseev, Charness, & Gneezy (2017)). Our hypothesis is close to that put forward elsewhere (e.g., Bowles (1998); Kirman & Teschl (2010)): preferences may be endogenous and thus shaped differently by market interactions than by, say, interactions with friends and family. Some experiments lend support to the importance of frames, for example the fa-

[4]We discuss alternative explanations in Section 5, such as *ex ante* fairness concerns (Saito (2013)), the uncertainty as to whether the subject's decision will be carried out (Exley (2016); Chen & Schonger (2022)), and magical thinking ( Shafir & Tversky (1992); Daley & Sadowski (2017)).

mous study that showed that a Prisoner's dilemma generates more cooperation if it is labeled as "the community game" than if it is called "the Wall street game" (Liberman, Samuels, & Ross, 2004) (see also (Kay & Ross, 2003)). Other studies, however, found small or no effects (Dreber, Ellingsen, Johannesson, & Rand, 2013; Dufwenberg, Gächter, & Hennig-Schmidt, 2011). By contrast to other studies that use some kind of market-related framing (e.g., Hoffman, McCabe, Shachat, & Smith (1994); Dufwenberg, Gächter, & Hennig-Schmidt (2011); Thöni & Gächter (2015)), our instructions to the participants explicitly describe an action as the sale of a lemon, and we compare behavior under this wording to behavior under neutral wording. Furthermore, our design allows us to compare the estimated preference parameters between the neutral and the market frames.

The next section describes the experimental design and procedures, and in Section 3 we test whether the wording used to describe the task (market vs neutral and VOI vs non-VOI) affects behaviors. In Section 4 we report the results from the structural estimations. In Section 5 we discuss alternative explanations, and Section 6 concludes.

# 2 The experiment

## 2.1 Game protocols and preferences

In the experiment subjects are matched into pairs to play anonymous one-shot interactions. In each interaction one subject is assigned to the *Player 1* role and the other to the *Player 2* role, each role assignment being equally likely. Player 1 has to choose between two actions, call them $X$ and $Y$ for now. If she chooses $X$, both individuals obtain their initial endowment, denoted $e_1$ and $e_2$, respectively, while if she chooses $Y$, she gets $e_1 + G > e_1$ while Player 2 gets $e_2 - L < e_2$, where $e_1 > e_2$, $G > 0$, and $L > 0$. That is, Player 1 makes a net gain $G$ from choosing $Y$ rather than $X$, while this choice entails a net loss $L$ for Player 2. Note that whether she chooses $X$ or $Y$, Player 1 receives a higher payoff than Player 2. The game tree for any matched pair of subjects—here called $i$ and $j$—is depicted in Figure 1.

We use two frames in the experiment: the **Neutral frame** and the **Market frame**. In the latter, we replace *Player 1* by *Seller* and *Player 2* by *Buyer*, and we describe actions $X$ and $Y$ by *Not Sell* and *Sell*, respectively. Indeed, the payoff structure captures in a stylized manner the case where a seller must decide whether to sell a low-quality item to a buyer in exchange for the going (fixed) price in the market—a classic "lemons" situation, in which $Y$ amounts to

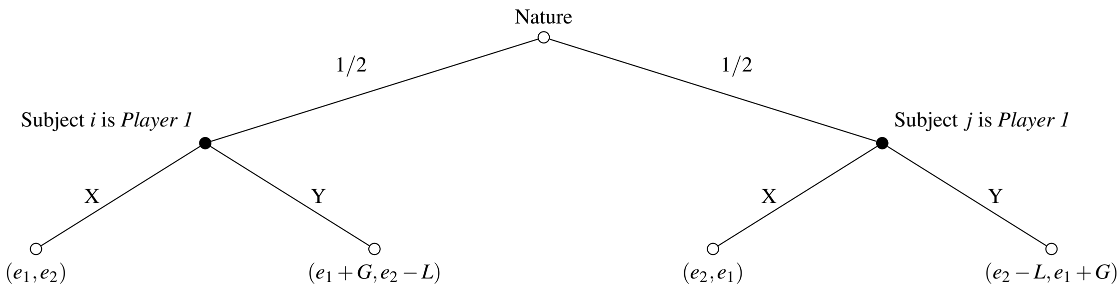



Figure 1: The game protocol used in the experiment. The payoff vectors state subject $i$'s and $j$'s payoff as the first and second component, respectively.

selling and $X$ to keeping the good. The Buyer has no say here: we adopt this simplification to concentrate on the Seller's willingness to sell a lemon. To facilitate the exposition, we henceforth use the term *Selfish option* to refer to *Sell* in the Market frame and action $Y$ in the Neutral frame, and the term *Status quo option* to refer to *Not Sell* in the Market frame and action $X$ in the Neutral frame. We will also refer to the two roles as the active and the passive roles.

In addition to the aforementioned frames, we vary the information that subjects receive about the decision situation. In the **VOI decision situation** subjects are asked to state their choice in the active role *before* learning the role distribution. Participants are told they have an equal chance of being cast in either role. Being thus explicitly informed about the whole game tree in Figure 2, they are led to reason behind the "Veil of Ignorance" (VOI) with respect to the role distribution. In the **non-VOI decision situation**, a subject is asked to state her decision *after* being informed of being in the active role and is therefore not behind the "Veil of Ignorance". Hence, even though the game actually being played between any given matched subject pair in the experiment is the one depicted in Figure 2 (since each subject stands an equal chance of being handed the active role), in a non-VOI decision situation a subject cast in the active role is given explicit information from us only about part of the game tree (the part highlighted in red in Figure 2, if $i$ is the active subject).

The decisions being anonymous and one-shot, the experimental design removes motivations such as social image concerns, repeated interaction, and reciprocity effects. Hence, a subject's decision should be driven by her intrinsic preferences and beliefs. We posit *social-Kantian preferences*, which combine material self-interest, other-regard, and a Kantian moral

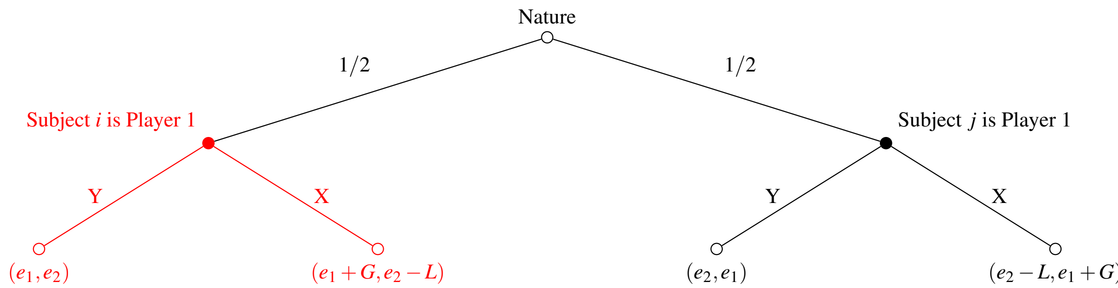



Figure 2: The matched subjects ($i$ and $j$) are informed about the whole game tree in a VOI decision, but only about the realized role distribution in a non-VOI decision (if the Player 1 role is assigned to subject $i$, she only receives information about the part highlighted in red).

concern. Formally, consider subjects $i$ and $j$ and denote their decisions by $x \in \{0,1\}$ and $y \in \{0,1\}$ respectively, where $x = 1$ if $i$ chooses the selfish option and $x = 0$ if she selects the *status quo* option when active, while $y = 1$ if $j$ selects the selfish option and $y = 0$ if she chooses the *status quo* option when active. We posit the following (expected) utility for $i$:[5]

$$\begin{aligned} U_i(x,y) &= (1-\kappa_i)\cdot\left[\frac{1}{2}\cdot(e_1+xG)+\frac{1}{2}\cdot(e_2-yL)\right] \\ &+ \kappa_i\cdot\left[\frac{1}{2}\cdot(e_1+xG)+\frac{1}{2}\cdot(e_2-xL)\right] \\ &- \frac{1}{2}\cdot\beta_i\cdot[e_1-e_2+x(G+L)] \\ &- \frac{1}{2}\cdot\alpha_i\cdot[e_1-e_2+y(G+L)] \end{aligned} \tag{1}$$

The term inside the square brackets in the first line is $i$'s expected material payoff: with probability $1/2$ $i$ gets the active role, and $i$'s decision $x$ determines her material payoff; with probability $1/2$ $j$ gets the active role, and $j$'s decision $y$ determines $i$'s material payoff. The term inside the square brackets in the second line captures $i$'s Kantian moral concern: it is the expected material payoff that $i$ (in fact, any subject in this interaction) would get if—hypothetically—the strategy $x$ was universalized. The parameter $\kappa_i \in [0,1]$ is $i$'s *degree of morality*.[6] The third line measures the effects on $i$'s utility of being materially ahead of

[5]This is the same as that posited in the main analysis of (van Leeuwen & Alger, 2024). Such a utility function, which describes the individual's preferences over (potentially trivial) material payoffs, has been shown to be favored by evolution by natural selection Alger et al. (2020).

[6]Mathematically, the Kantian moral concern in *Homo moralis* preferences has a similar effect as the false

the other when in the active role, while the last line measures the effects on $i$'s utility of being materially behind the other when in the passive role. The parameter $\beta_i$ represents $i$'s *aheadness aversion* (if $\beta_i > 0$) or *love for aheadness* (if $\beta_i < 0$), where aheadness means that $i$'s monetary payoff is larger than $j$'s. Similarly, the parameter $\alpha_i$ represents $i$'s *behindness aversion* (if $\alpha_i > 0$) or *love for behindness* (if $\alpha_i < 0$), where behindness means that $i$'s monetary payoff is smaller than $j$'s.

To state our main hypothesis, rewrite the expression in (1) as follows:

$$\begin{aligned} U_i(x,y) &= \frac{1}{2}\Big[e_1 + xG - \beta_i \cdot [e_1 - e_2 - x(G+L)]\Big] \\ &+ \frac{1}{2}\Big[e_2 - yL - \alpha_i \cdot [e_1 - e_2 + y(G+L)]\Big] - \frac{1}{2} \cdot \kappa_i \cdot (x-y)L. \end{aligned} \tag{2}$$

There are two conceptually different parts. The first two terms are *consequentialist* in nature: it is the expected material payoff, net of the psychological effect of inequality. The last term is *counterfactual*: independent of $j$'s actual action $y$ when in the active role (which from $i$'s point of view is a constant), if $\kappa_i > 0$ this term makes $i$ realize that if the action $x = 1$ was universalized to $j$, it would entail a reduction in her own material utility of $L$ when she is in the passive role. Our hypothesis is that subjects will pay more attention to this Kantian moral concern in VOI decision situations, in which the subject is made fully aware of the possibility that roles may be reversed. But consider now the same subject $i$ cast in the active role in a non-VOI decision situation, and assume that this subject is completely unaware of the fact that she could have been allocated to the passive role instead. Her utility then reduces to:

$$V_i(x,y) = e_1 + xG - \beta_i \cdot [e_1 - e_2 + x(G+L)]. \tag{3}$$

Comparing this with (1) or (2), we see that as long as $L > 0$, this subject may select a different $x$ in VOI and non-VOI decisions if $\kappa_i > 0$. Specifically, Kantian morality reduces the subject's willingness to select the selfish decision, since she is willing to select the selfish

belief that one's action affects the action of the opponent, a belief known as magical (or quasi-magical) thinking (Shafir & Tversky, 1992; Daley & Sadowski, 2017). Moreover, under the special case of *Homo moralis* preferences where the degree of morality $\kappa$ is equal to one, predictions based on Nash equilibrium play sometimes coincide with predictions based on Kantian equilibrium, an equilibrium concept introduced by Roemer (2010). The preference-based approach adopted here avoids reliance on false beliefs and allows for the adoption of the standard Nash equilibrium concept. For further discussion, we refer to Salonia (2024), who adopts an axiomatic approach to preferences for universalization.

decision if:

$$G \geq \beta_i(G+L) \tag{4}$$

in the non-VOI decision situation, but only if:

$$G - \kappa_i L \geq \beta_i(G+L) \tag{5}$$

in the VOI decision situation.

We are now in a position to state our main hypothesis: for any given preference type $(\beta_i, \kappa_i)$ and a given payoff vector $(e_1, e_2, e_1+G, e_2-L)$ a subject may *make a switch* from the selfish action in the non-VOI decision to the *status quo* act in the VOI decision situation. Indeed, conditions (4) and (5) imply that this happens if $i$'s aheadness aversion $\beta_i$ is not too pronounced,

$$\beta_i \leq \frac{G}{G+L} \equiv z, \tag{6}$$

and the Kantian moral concern $\kappa_i$ is sufficiently pronounced,

$$\kappa_i \geq \frac{G-\beta_i(G+L)}{L} = \frac{z-\beta_i}{1-z}. \tag{7}$$

With heterogeneity in the preference types, the hypothesis is that the average propensity to select the selfish action is lower in VOI than in non-VOI decision situations.

Figure 3 shows the threshold values $z$ and $\frac{z-\beta_i}{1-z}$ for two payoff configurations used in the experiment: the solid vertical and oblique lines show, respectively, the threshold values $z$ and $(z-\beta_i)/(1-z)$ for payoff configuration $(e_1, e_2, e_1+G, e_2-L) = (150, 100, 165, 90)$; the dashed lines for payoff configuration $(e_1, e_2, e_1+G, e_2-L) = (200, 190, 210, 100)$. To see how variation of the payoffs enables us to distinguish between different preference types, note that the two solid lines and the two dashed lines divide the $(\beta, \kappa)$-space into six regions. An individual with a preference type $(\beta_i, \kappa_i)$:

- in region A always selects the selfish option;
- in region B selects the selfish option under payoff $(150, 100, 165, 90)$, but makes a switch under payoff $(200, 190, 210, 100)$;
- in region C makes a switch under both payoffs;
- in region D selects the selfish option under payoff $(150, 100, 165, 90)$, but always the *status quo* option under payoff $(200, 190, 210, 100)$;
- in region E makes a switch under payoff $(150, 100, 165, 90)$, but always selects the *status quo* option under payoff $(200, 190, 210, 100)$;

- in region F always selects the *status quo* option.

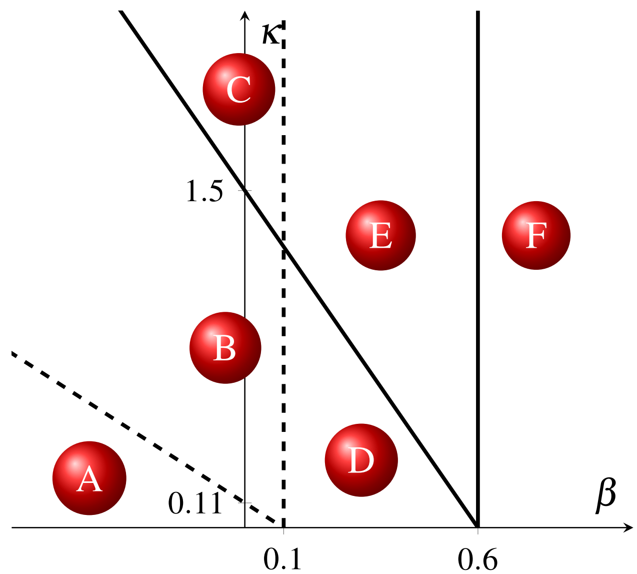



Figure 3: Threshold values $z$ and $(z-\beta_i)/(1-z)$ for payoffs $(e_1,e_2,e_1+G,e_2-L) = (150,100,165,90)$ (solid) and $(e_1,e_2,e_1+G,e_2-L) = (200,190,210,100)$ (dashed)

In the experiment we use 20 different payoffs $(e_1,e_2,e_1+G,e_2-L)$. Table 1 shows, for each of these 20 payoffs, the threshold values $z$ and $(z-\beta)/(1-z)$ (see (6) and (7)). Note that the 20 payoff configurations used in the experiment entail 20 different values of $z$. Let $Z$ denote the set of these 20 values. Figure 4 shows the lines $\kappa = (z-\beta)/(1-z)$ for the 20 values of $z$. Except for individuals with $\beta_i \geq 0.6$ (who always select the *status quo* option for all payoffs), and those with $\beta_i \leq 0.03$ and $\kappa_i \leq 0.03 - 1.03\beta_i$ (who always select the selfish option for all payoffs), these payoffs are expected to generate behavioral variation across payoffs and/or across the two conditions (VOI vs. non-VOI).[7]

The payoffs were chosen so as to maximize statistical power in the between-subject estimation of the VOI treatment's effects (conditional on the fact that we decided to have no

[7]Note that for all but two payoff configurations, the selfish option reduces the aggregate payoff; hence, selecting the *status quo* could be driven by a wish to maximize total payoff (Charness & Rabin (2002)). Such a wish is compatible with the posited utility function for $\beta_i = 1/2$. To see this, rewrite the expression in (3) as $(1-\beta_i)(e_1+xG)+\beta_i(e_2-xL)$.

more than 20 decisions in each sequence), where power computations were based on the results presented by van Leeuwen & Alger (2024). More precisely, we carried out 1000 simulations where in each one, the 109 subjects for which van Leeuwen & Alger (2024) estimate individual preference parameters were randomly assigned to either VOI (55 participants) or non-VOI (54 people). Then, we computed their decisions based on the $(\beta, \kappa)$-estimates obtained by van Leeuwen & Alger (2024), assuming that subjects in the non-VOI treatment where completely unaware of the possibility of role reversal. Finally, we estimated the effect of the VOI treatment on the decision to sell, through the specifications presented in Section 3. The results of this exercise indicate that the chosen payoffs generate behavioral variations that our main specification is able to detect 75% of the time (at a 5% confidence level).

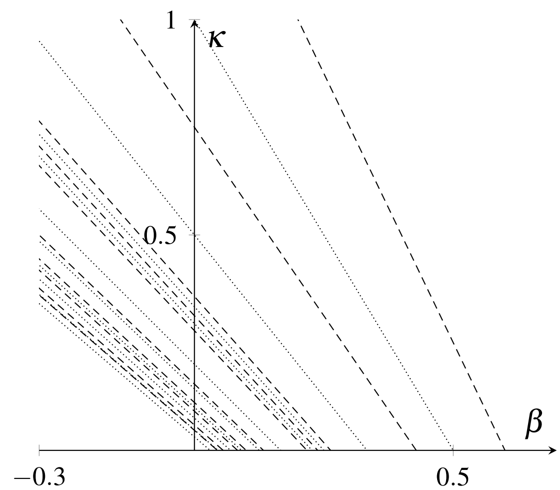



Figure 4: Each line represents the threshold value $\bar{\kappa}(\beta)$ for one of the 20 payoffs.

In the analysis above we assumed that in a non-VOI decision situation a subject is completely unaware of the complete game tree. However, some subjects may understand that they could have been in their counterpart's place with some non-null probability. The following equation shows the utility of a subject $i$ who assigns a probability $\hat{p}_i$ that she was cast

Table 1: Payoff combinations $(e_1, e_2, e_1+G, e_2-L)$ (first four columns) and (rounded) threshold values for $\beta$ and $\kappa$ as per equations (6) and (7) (last two columns)

| Payoff number ($s$) | $e_1$ | $e_2$ | $e_1+G$ | $e_2-L$ | $z=G/(G+L)$ | $(z-\beta_i)/(1-z)$ |
|---|---|---|---|---|---|---|
| 1 | 150 | 100 | 165 | 90 | 0.6 | $1.5-\beta_i\cdot 2.5$ |
| 2 | 150 | 100 | 160 | 90 | 0.5 | $1-\beta_i\cdot 2$ |
| 3 | 150 | 100 | 165 | 80 | 0.43 | $0.75-\beta_i\cdot 1.75$ |
| 4 | 150 | 100 | 165 | 70 | 0.33 | $0.5-\beta_i\cdot 1.5$ |
| 5 | 250 | 240 | 300 | 100 | 0.26 | $0.36-\beta_i\cdot 1.36$ |
| 6 | 250 | 240 | 300 | 90 | 0.25 | $0.33-\beta_i\cdot 1.33$ |
| 7 | 250 | 240 | 300 | 80 | 0.24 | $0.31-\beta_i\cdot 1.31$ |
| 8 | 250 | 240 | 300 | 70 | 0.23 | $0.29-\beta_i\cdot 1.29$ |
| 9 | 250 | 240 | 300 | 60 | 0.22 | $0.28-\beta_i\cdot 1.28$ |
| 10 | 150 | 120 | 170 | 20 | 0.17 | $0.2-\beta_i\cdot 1.2$ |
| 11 | 200 | 190 | 220 | 60 | 0.13 | $0.15-\beta_i\cdot 1.15$ |
| 12 | 200 | 190 | 220 | 50 | 0.125 | $0.14-\beta_i\cdot 1.14$ |
| 13 | 200 | 190 | 210 | 100 | 0.1 | $0.11-\beta_i\cdot 1.11$ |
| 14 | 200 | 190 | 210 | 90 | 0.09 | $0.1-\beta_i\cdot 1.1$ |
| 15 | 200 | 190 | 210 | 80 | 0.08 | $0.09-\beta_i\cdot 1.09$ |
| 16 | 250 | 220 | 265 | 20 | 0.07 | $0.08-\beta_i\cdot 1.08$ |
| 17 | 250 | 220 | 260 | 50 | 0.06 | $0.06-\beta_i\cdot 1.06$ |
| 18 | 250 | 220 | 260 | 10 | 0.05 | $0.05-\beta_i\cdot 1.05$ |
| 19 | 250 | 220 | 260 | 5 | 0.04 | $0.05-\beta_i\cdot 1.05$ |
| 20 | 250 | 220 | 255 | 60 | 0.03 | $0.03-\beta_i\cdot 1.03$ |

in the Seller role.

$$\begin{aligned}\hat{U}_i(x,y) &= (1-\kappa_i)\cdot\Big[\hat{p}_i\cdot(e_1+xG)+(1-\hat{p}_i)\cdot(e_2-yL)\Big] \\ &+ \kappa_i\cdot\Big[\hat{p}_i\cdot(e_1+xG)+(1-\hat{p}_i)\cdot(e_2-xL)\Big] \\ &- \hat{p}_i\cdot\beta_i\cdot[e_1-e_2+x(G+L)] \\ &- (1-\hat{p}_i)\cdot\alpha_i\cdot[e_1-e_2+y(G+L)]\end{aligned} \tag{8}$$

The expressions in (1) and (3) correspond to the special cases $\hat{p}_i = 1/2$ and $\hat{p}_i = 1$, respectively. We will say that a subject with $\hat{p}_i = 1$ is *fully unaware*, that one with $\hat{p}_i \in (1/2, 1)$ is *partially aware*, and that one with $\hat{p}_i = 1/2$ is *fully aware* of the full game tree.

A partially aware subject switches from the selfish option in the non-VOI decision situation to the *status quo* option in the VOI decision situation if:

$$\beta_i < \frac{G - \frac{1-\hat{p}_i}{\hat{p}_i}\kappa_i L}{G+L} \tag{9}$$

and:

$$\kappa_i > \frac{G-\beta_i(G+L)}{L}. \tag{10}$$

The prediction that some subjects may switch from the selfish option in a non-VOI decision to the *status quo* option in a VOI decision thus holds as long as they hold some belief $\hat{p} > 1/2$. Indeed, a fully aware subject, with $\hat{p} = 1/2$, or a subject who holds a belief $\hat{p} < 1/2$ would never switch, since conditions (9) and (10) are incompatible for such beliefs. While the observation of such switches thus has some implications for the subjects' beliefs, it is not possible to estimate both the preference parameters $\beta_i$ and $\kappa_i$ and the beliefs with our data (from conditions (9) and (10) it is clear that there would be too many degrees of freedom). Moreover, we deliberately chose not to elicit the subjects' beliefs $\hat{p}$: eliciting them prior to the presentation of the decision situations might have affected the choices, and eliciting them afterwards would make no sense since the subjects would then have been exposed already to VOI decisions, in which they are told explicitly that both role distributions are equally likely. In view of this limitation, in the analysis below we will adopt the admittedly stark assumption that $\hat{p} = 1$ in all non-VOI decisions.[8]

[8]The reader should, however, not infer from this that we believe that $\kappa$ is not at work in real decisions with an asymmetric role distribution. More research will be needed to obtain both a theory of how inclined individuals are to adopt a Veil of Ignorance to such decisions, and a method to measure this empirically. Our objective here is more modest: we are simply aiming at finding out whether subjects' decisions are consistent with the proposed hypothesis that the Veil of Ignorance is more salient in VOI than in non-VOI decisions.

## 2.2 Treatments

Each subject is asked to state her choice in two sequences, both of which consist of the same set of 20 payoff configurations, listed in Table 1. Each sequence of 20 decisions is either entirely non-VOI or entirely VOI.

Clearly, to avoid raising awareness of the arbitrariness of the role distribution in the non-VOI decision situations, subjects should be presented a non-VOI sequence before being exposed to a VOI sequence. However, always letting a VOI sequence be preceded by a non-VOI one may lead to anchoring and/or fatigue effects. To deal with these issues, and to enable meaningful comparisons between the Neutral and the Market frames, we adopt the following four treatments:

1. **Neutral (N)**: Non-VOI + VOI — Neutral
2. **Market (M)**: Non-VOI + VOI — Market
3. **Mixed A (A)**: VOI Neutral + VOI Market
4. **Mixed B (B)**: VOI Market + VOI Neutral

These treatments allow us to carry out a large number of comparisons: (1) a within-subject comparison of decisions in the non-VOI and VOI sequences in the **Neutral** treatment; (2) a within-subject comparison of decisions in the non-VOI and VOI sequences in the **Market** treatment; (3) a between-subject comparison of decisions in the non-VOI sequence in the **Neutral** treatment and the first sequence in the **Mixed A** treatment; (4) a between-subject comparison of decisions in the non-VOI sequence in the **Market** treatment and the first sequence in the **Mixed B** treatment; and (5) within-subject comparisons of decisions in differently framed VOI sequences, using treatments **Mixed A** and **Mixed B**.

For further use below, let $D$ denote the set of eight decision sequences in our four treatments: $D = \{N1, N2, M1, M2, A1, A2, B1, B2\}$, where a 1 refers to the first sequence and a 2 to the second sequence of 20 decisions in the treatment.

## 2.3 Procedures and data

We conducted the experiment between November 2021 and March 2022 at the Toulouse School of Economics TSE Lab for Experimental Social Sciences. The software oTree was used to program the experiment, and participants were recruited via email using the Laboratory's participant pool (people who had signed up to be informed of laboratory experiments). Overall, we recruited 453 participants, all of whom participated in only one session. Tables A.1 and A.2 in Appendix A1 provide information on the total number of subjects in each treatment and associated summary statistics. The experiment (the design and the empirical analysis reported below) was pre-registered on aspredicted.org on November 21st 2021 (available at https://aspredicted.org/qxt2-mh8j.pdf), the first session taking place on November 22nd.

Each session was allocated to one treatment, and lasted between 25 and 45 minutes. After being randomized to the lab booths, participants read the instructions on their desktop computer screens, were allowed to ask questions privately, and completed a comprehension test. Then the two sequences of 20 decisions were presented, upon which the participants filled out a post-experiment questionnaire (with questions on sex, age, nationality, and field of study). The English version of the instructions are included in Appendix A4 (the experiment was conducted in French). In each decision situation the participants had to answer questions about the payoffs correctly before being allowed to state their decision (see an example of the two screens shown for one decision in Appendix A4.4).

The participants' payoffs were determined based on two randomly drawn matches, one from each sequence. Each unit of the payoffs used in the instructions (see Table 1) was converted to 2.5 eurocents. For VOI decisions, for the randomly drawn match the role distribution was determined randomly, and the payoffs were calculated based on the active subject's decision. For non-VOI decisions, for the randomly drawn match each subject received the payoff corresponding to their decision, and also the payoff corresponding to the other's decision (of which they were not aware during the experiment). Participants who answered all the questions received a show-up fee of 4 euros. Together with the payoffs obtained from the tasks, participants earned on average 15.5 euros. Participants privately received their payoffs in cash, upon which they left the premises.

Table 2 below and Figures A.1-A.3 in Appendix A1 give a first glimpse of the data, here pooled for all the $n = 453$ participants' decisions across all the treatments and sessions. In the figures we define one observation as the number of selfish options selected by one participant in one sequence of 20 decisions (there are thus two observations per participant). Figure A.1 shows the observations from the Neutral frame on the left and those from the Market frame on the right. The dashed (resp. solid) horizontal lines show the average (resp.

median) number of selfish choices. It appears that, overall, the participants were more prone to select the selfish option in the Market than in the Neutral frame. Figure A.2 compares the cumulative distributions of the observations in the VOI decision sequences (solid line, $n = 686$) and the observations in the non-VOI decision sequences (dashed line, $n = 220$). Participants were more prone to select the selfish option in the non-VOI than in the VOI decision situations. Figure A.3 confirms these two tendencies by splitting the observations into the four different VOI-frame combinations: Neutral VOI (black line), Market VOI (red line), Neutral non-VOI (blue line), and Market non-VOI (green line). Finally, Table 2 provides the summary statistics on the number of selfish decisions: in the first two rows for the Neutral and the Market frames, followed by two rows for the VOI and non-VOI wordings, and finally the four different combinations. We see that the differences in the total number of selfish decisions between the Neutral and the Market frame, and between the VOI and non-VOI wordings, are significantly different from zero.

Table 2: Total number of selfish decisions per sequence of 20 decisions, by frame

| Frame | N | Mean | Median | Min | Max | Q1 | Q3 |
|---|---|---|---|---|---|---|---|
| Neutral | 449 | 7.08 | 6 | 0 | 20 | 2 | 10 |
| Market | 457 | 9.03 | 9 | 0 | 20 | 4 | 13 |
| t-test and Wilcoxon test p-values | | 0.00 | 0.00 | | | | |
| VOI | 686 | 7.35 | 7 | 0 | 20 | 3 | 10 |
| non-VOI | 220 | 10.30 | 10 | 0 | 20 | 6 | 15 |
| t-test and Wilcoxon test p-values | | 0.00 | 0.00 | | | | |
| VOI and Neutral | 341 | 6.52 | 6 | 0 | 20 | 2 | 9 |
| VOI and Market | 345 | 8.17 | 8 | 0 | 20 | 4 | 11 |
| non-VOI and Neutral | 108 | 8.86 | 8 | 0 | 20 | 4 | 13 |
| non-VOI and Market | 112 | 11.7 | 11 | 0 | 20 | 8 | 17 |

# 3 Hypothesis testing

In this section we report regression estimates that test whether the the VOI treatment and the Market frame have significant effects on behavior in the experiment.

### 3.1 Does the VOI treatment significantly affect behavior?

To evaluate the effects of the VOI treatment, we estimate linear models of the form:

$$x_{isd} = \gamma_0 + \gamma_1 V_{isd} + \gamma_2 s + \varepsilon_{isd}, \tag{11}$$

where $x_{isd} \in \{0,1\}$ is subject $i$'s decision in payoff configuration $s \in \{1,\ldots,20\}$ (recall Table 1) and decision sequence $d \in \{N1, N2, M1, M2, A1, A2, B1, B2\}$; $x_{isd} = 1$ if she chooses the selfish option and $x_{isd} = 0$ otherwise. The variable $V_{isd}$ is a dummy equal to 1 if the decision is made under the VOI condition, and 0 otherwise. The error term $\varepsilon_{isd}$ is assumed to have zero mean and to be uncorrelated with both the effect of the VOI condition and and the payoff fixed effect. Our null hypothesis is then that $\gamma_1 = 0$. The results, reported in Table 3, show that this hypothesis is clearly rejected, for both within- and between-subject analyses, and that the effect of VOI is large.

Payoff configuration fixed effects are included in all regressions. For within-subject comparisons, we also include subject fixed effects, leveraging the panel structure of the data. This controls for all stable individual traits and the variation across payoff structures. In between-subject comparisons, we account for observable heterogeneity by instead controlling for socio-demographic characteristics.

Beginning with the neutral frame, the first and second columns of Table 3 present the within-subject ($d \in$ {N1, N2}) and between-subject ($d \in$ {N1, A1}) comparisons, respectively. In the non-VOI condition, the selfish option is chosen in 44.3% of decisions. Relative to this baseline, the VOI treatment reduces the share of selfish decisions decisions by 15.28 percentage points in the within-subject comparison — a sizable and highly significant effect.[9] The between-subjects effect is smaller, at 7.28 percentage points. One possible explanation for the difference is heterogeneity across subjects: the between-subjects effect is only weakly significant, and the model's fit is noticeably worse.

Turning to the market frame, the third and fourth columns of Table 3 report the within- and between-subject comparisons ($d \in$ {M1, M2} and $d \in$ {M1, B1}, respectively). In the non-VOI condition, the selfish option is chosen in 58.4% of decisions — 14.1 percentage points higher than in the neutral frame. The VOI treatment reduces the share of selfish decisions by 18.66 percentage points in the within-subject comparison, and by 18.79 percentage points in the between-subject comparison. Both effects are highly significant. The

[9] This suggests that the within-subject design does not suffer strongly from a desire to act consistently across VOI and non-VOI situations with identical payoffs.

Table 3: Effect of VOI on Selfish. Dependent variable: *Selfish decision* (1 if subject chose the selfish action, 0 otherwise)

| | Neutral – Within (N1–N2) | Neutral – Bet. (N1–A1) | Market – Within (M1–M2) | Market – Bet. (M1–B1) |
|---|---|---|---|---|
| VOI | -0.1528*** | -0.0728* | -0.1866*** | -0.1879*** |
| | (0.0294) | (0.0403) | (0.0349) | (0.0402) |
| *Socio-dem. controls* | | | | |
| | — | Yes | — | Yes |
| *Fixed-effects* | | | | |
| Subject | Yes | — | Yes | — |
| Payoff config. ($s$) | Yes | Yes | Yes | Yes |
| Observations | 4,320 | 4,580 | 4,480 | 4,480 |
| Mean selfish (VOI = "No") | 0.443 | 0.443 | 0.584 | 0.584 |
| # Subjects | 108 | 229 | 112 | 224 |
| # Subjects (VOI = "No") | 108 | 108 | 112 | 112 |
| # Subjects (VOI = "Yes") | 108 | 121 | 112 | 112 |
| *Fit statistics* | | | | |
| $R^2$ | 0.334 | 0.060 | 0.376 | 0.141 |
| Within $R^2$ | 0.036 | 0.010 | 0.053 | 0.062 |

*Note:* Standard errors in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors are clustered at the subject and payoff configuration level. Socio-demographic controls are: nationality, gender, couple status, previous experience in experiments and attendance to a private school.

within-subject effect is only moderately stronger than in the neutral frame (a difference of 3.4 percentage points), and the similarity between the within- and between-subject effects suggests that heterogeneity across subjects plays a smaller role in the market frame. Overall, the VOI treatment has a large and consistent effect on the propensity to sell.

The results reported in Table 3 are robust to replacing the payoff configuration fixed effects $s \in \{1, \ldots, 20\}$ in the regression by the continuous regressor $z = G/(G+L)$, which captures the relative "advantageousness" of selfish behavior. In the non-VOI condition, the selfish action is chosen when $\beta_i \leq z$ (recall (6)), so a higher $z$ increases the share of individuals for whom this inequality holds. In the VOI condition, the selfish action is chosen when $\kappa_i < \frac{z-\beta_i}{1-z}$ (recall (7)). This threshold is increasing in $z$ for $\beta_i < 1$. As a consequence, if a significant share of individuals are not extremely aheadness averse ($\beta_i > 1$), we should expect more selfish decisions for payoff configurations with a larger $z$. Empirically observing a positive coefficient on $z$ is therefore consistent with the model's predictions and suggests that most individuals fall within the moderate range of aheadness aversion. The results are reported in Tables A.5-A.8 in Appendix A1. These regressions yield very similar estimates and statistical significance as the ones reported in Table 3. Tables A.5-A.8 also show the results of alternative specifications where (*i*) we exclude the fixed effects and payoff-configuration controls as well as assume iid-distributed errors; and (*ii*) assume that errors are correlated only across subjects or only across payoff configurations rather than across both. Our results hold throughout.

Finally, we check the robustness of our findings to functional form assumptions, by estimating logit models, allowing for nonlinear effects of the regressors on the probability of selfish choice, as detailed in our pre-registration. In Table A.9, we report the corresponding Average Marginal Effects (AMEs) with bootstrapped standard errors. The direction and significance of the estimated effects are identical to those in Table 3, and the AMEs are quantitatively very similar to the linear probability coefficients. Tables A.10-A.13 in Appendix A1 present the logit regressions corresponding to the linear ones found in Tables A.5-A.8. The findings remain consistent across all specifications and the logits' AME's are in all cases quantitatively very close to the linear models' coefficients.

## 3.2 Does the Market frame significantly affect behavior?

To assess the effects of the Market frame we estimate linear models of the form:

$$x_{isd} = \lambda_0 + \lambda_1 M_{isd} + \lambda_2 s + \upsilon_{isd}, \tag{12}$$

where $M_{isd}$ is a dummy variable that takes value 1 if the frame is Market and 0 if the frame is Neutral; and $\upsilon_{isd}$ is a mean-zero random variable assumed to be uncorrelated with the frame effects and with the payoff fixed effects. Our null hypothesis is then that $\lambda_1 = 0$. Table 4 presents the estimated effect of the Market frame on the likelihood of choosing the selfish option across four experimental contrasts: within-subject comparisons under VOI (first and second columns), and between-subject comparisons under non-VOI (third column) and VOI (fourth column). The null hypothesis is rejected in all cases but one.

Table 4: Effect of Market Frame on Selfish. Dependent variable: *Selfish (1 if subject chose the selfish action, 0 otherwise)*

| | VOI – Within (A1–A2) | VOI – Within (B1–B2) | nVOI – Bet. (N1–M1) | VOI – Bet. (A1–B1) |
|---|---|---|---|---|
| Market | 0.0674** | 0.0728** | 0.1425*** | 0.0488 |
| | (0.0291) | (0.0295) | (0.0403) | (0.0387) |
| *Socio-dem. controls* | | | | |
| | — | — | Yes | Yes |
| *Fixed-effects* | | | | |
| Subject | Yes | Yes | — | — |
| Payoff config. ($s$) | Yes | Yes | Yes | Yes |
| Observations | 4,840 | 4,480 | 4,400 | 4,660 |
| Mean Selfish (Neutral Frame) | 0.369 | 0.328 | 0.443 | 0.369 |
| # Subjects | 121 | 112 | 220 | 233 |
| # Subjects (Neutral Frame) | 121 | 112 | 108 | 121 |
| # Subjects (Market Frame) | 121 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| $R^2$ | 0.347 | 0.331 | 0.103 | 0.082 |
| Within $R^2$ | 0.007 | 0.008 | 0.03 | 0.02 |

*Note:* Standard errors in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors are clustered at the subject and payoff configuration level. Socio-demographic controls are: nationality, gender, couple status, previous experience in experiments and attendance to a private school.

The strongest and most robust effect appears in the between-subject comparison of non-VOI decisions (N1 vs. M1). Here, the Market frame increases the share of selfish choices by 14.3 percentage points (from 44.3% to 58.6%). This effect is statistically significant at the 1% level.

In VOI decisions, the framing effect is smaller in magnitude. Within-subject comparisons from Treatments A and B (columns 1 and 2) show that the Market frame increases selfish choices by 6.7 and 7.3 percentage points, respectively. In Treatment A, the share of selfish choices increases from 36.9% in the Neutral frame to 43.6% in the Market frame; in Treatment B, it increases from 32.8% to 40.1%. These effects are statistically significant at the 5% level. Importantly, Treatment A presents the Neutral frame first and the Market frame second, while Treatment B reverses the order. The similarity of effects across the two treatments suggests that the observed difference is not due to decision order but to the frame itself.

By contrast, the between-subject comparison in VOI decisions (A1 vs. B1) yields a smaller point estimate of 4.9 percentage points, which is not statistically significant in the specification shown.[10] In this group, the share of selfish choices under the Neutral frame was 36.9%, rising to 41.8% under the Market frame. This result reflects both the smaller size of the framing effect under VOI conditions—as evidenced by the within-subject comparisons in the first two columns—and the lower precision associated with between-subject designs, which are more vulnerable to individual-level heterogeneity.

Taken together, the results provide robust evidence that describing the decision as a market transaction increases the likelihood of selfish behavior. The effect is strongest in non-VOI conditions, but is also present and statistically significant in VOI decisions when the frame varies within subjects.

Like we did for the effects of VOI, robustness checks are included in Appendix A1. For linear regressions using $z$ instead of $s$ as the regressor controlling for payoff effects, with alternative standard error clustering levels, see Tables A.14-A.17. The logistic regressions corresponding to the linear regressions in Table 4 and Tables A.14-A.17 can be found in Table A.18 and Tables A.19-A.22, respectively. The results remain robust across these alternative specifications, and the AMEs computed using the logit coefficients are very similar in magnitude to the linear regressions' coefficients.

[10] In the same model, omitting the clustering of standard errors yields statistically significant results. These differences arise purely from the choice of inference method, not from changes in model specification. We report here the most conservative estimates—those clustering standard errors at the subject and payoff configuration levels; see Table A.17 in Appendix A1 for alternative clustering approaches.

## 3.3 Interaction effects

The results reported above show that VOI has a larger absolute impact on the likelihood of choosing the selfish option when decisions are framed as market transactions. However, from the descriptive statistics we also know that subjects are on average more willing to choose the selfish option in the Market than in the Neutral frame. To test whether the relative effect of VOI differs across frames, we estimate a pooled regression that includes an interaction between VOI and Market frame:

$$x_{isd} = \gamma_0 + \gamma_1 V_{isd} + \gamma_2 M_{isd} + \gamma_3 (V_{isd} \times M_{isd}) + \gamma_4 s + \varepsilon_{isd}. \tag{13}$$

Here, $\gamma_3$ captures the differential effect of VOI under the Market frame relative to the Neutral frame. Our null hypothesis is that $\gamma_3 = 0$, i.e., that the relative impact of VOI is the same whether the frame is Neutral or Market,

Note that to properly identify all the coefficients, we need to use the observations from all the treatments, since each coefficient is identified from a distinct source of variation. The coefficient on $V_{isd}$ ($\gamma_1$) captures the regression-weighted average effect of VOI in the Neutral frame. It is estimated from the within-subject contrast N1–N2, as well as between-subject comparisons of N1 with A1 and B2 (while holding payoff structure constant). This means that $\gamma_1$ reflects the average effect of VOI when the frame is held constant (Neutral), combining all available within- and between-subject variation under that condition. The Market frame coefficient $\gamma_2$ is identified purely from between-subject differences under non-VOI conditions, comparing decisions in N1 and M1 (again, net of payoff configuration). The interaction coefficient $\gamma_3$ tests whether the VOI effect differs under the Market frame; it is identified by comparing how the effect of VOI changes when moving from Neutral to Market conditions. This includes variation across all treatments that combine VOI and framing: A1, A2, B1, B2, N1, N2, and M1.

The results, presented in Table 5, show that the interaction coefficient $\gamma_3$ is not statistically significant. This implies that, despite a stronger absolute effect of VOI in the Market frame, the relative effect of VOI is statistically indistinguishable across frames. Importantly, the main effects of VOI and Market frame ($\gamma_1$ and $\gamma_2$) remain statistically significant and consistent in magnitude with those obtained in the more focused within- and between-subject comparisons discussed earlier. This confirms that the previously documented behavioral responses to VOI and framing are robust to the inclusion of all available variation in the full sample.

Table 5: Effect of VOI and Market on Selfish. Dependent variable: *Selfish decision* (1 if subject chose the selfish action, 0 otherwise)

| | (1) | (2) |
|---|---|---|
| VOI | -0.1127** | -0.1096** |
| | (0.0306) | (0.0306) |
| Market Frame | 0.1413** | 0.1491** |
| | (0.0405) | (0.0404) |
| VOI $\times$ Market Frame | -0.0596 | -0.0649 |
| | (0.0383) | (0.0383) |
| *Socio-dem. controls* | No | Yes |
| *Fixed-effects* | Payoff config. | Payoff config. |
| Observations | 18,120 | 18,120 |
| *Fit statistics* | | |
| $R^2$ | 0.112 | 0.117 |
| Within $R^2$ | 0.028 | 0.034 |

*Note:* Standard errors in parentheses. ** p < 0.05. All models include fixed effects for payoff configuration and cluster standard errors at both the subject and payoff level. Socio-demographic controls are: nationality, gender, couple status, prior experience in experiments, and attendance to a private school.

Table A.23 in Appendix A1 reports the AMEs for the logit version of this model. The sign and significance of the effects match the linear specification, but their magnitudes differ more than in the models that focused separately on VOI or framing. In those cases, effects were identified from balanced groups of similar size and with comparable proportions of selfish choices—around 40–45%—so the logistic scaling was nearly uniform. In the interaction model, by contrast, the contrasts underlying $\gamma_1$ and $\gamma_2$ differ both in size and baseline frequencies: 341 participants with sell shares between 33% and 44% contribute to the VOI comparisons, whereas the Market effect relies on smaller non-VOI groups—108 Neutral

(44.3%) and 112 Market (58.6%) participants. Because these contrasts are drawn from parts of the logistic curve with different slopes, the VOI AMEs appears slightly larger and the Market AMEs slightly smaller than their linear counterparts, even though both specifications point to the same behavioral pattern.

# 4 Structural estimation of $\beta$ and $\kappa$

Besides the reduced-form estimates of treatment effects specified above, we structurally estimate the aheadness aversion and Kantian morality parameters $\beta$ and $\kappa$.

## 4.1 Method

We carry out the analysis using the data obtained from subjects exposed to both non-VOI and VOI sequences. We first report the estimates of the preference parameters, $(\hat{\beta}, \hat{\kappa})$, and the choice sensitivity, $\hat{\sigma}$, under the hypothesis that the observed data pooled over all the subjects (in a given treatment) had emanated from a representative agent.

For non-VOI decisions we assume that any subject is fully unaware of the possible role reversal, i.e., we posit $\hat{p} = 1$ (see (8)). By contrast, for VOI decisions we assume that any subject is fully aware of the arbitrariness of the role distribution (by positing $\hat{p} = 1/2$), since we explicitly inform the subjects about it. These assumptions are extreme, and we do not claim that they are in line with the subjects' subjective beliefs about the role distribution. However, the advantage is that they imply a conservative estimate of $\hat{\kappa}$. Indeed, under the assumption that $\hat{p} = 1$ in non-VOI decisions, only a positive $\hat{\beta}$ can explain why a subject would refrain from selecting the selfish option. Hence, for any subject who is in fact fully or partly aware of the role reversal in the non-VOI decisions (i.e., with $\hat{p} = 1 \in [1/2, 1)$) and who refrains from the selfish option in such a decision due to a combination of aheadness attitude and Kantian morality (see (9)), our assumption that $\hat{\kappa}$ cannot be at work leads to over-estimation of $\hat{\beta}$ and under-estimation of $\hat{\kappa}$.

We use a standard random utility model, by assuming that the representative agent's true utility from using pure strategy $x$ is a random variable of the form:

$$\tilde{U}(x; \beta, \kappa, \sigma, z) = 2U(x; \beta, \kappa, z) + \varepsilon_x \tag{14}$$

in a VOI decision situation with payoff configuration $z \in Z$ and:

$$\tilde{V}(x;\beta,\sigma,z) = V(x;\beta,z) + \varepsilon_x \tag{15}$$

in a non-VOI decision situation with payoff configuration $z \in Z$, where the function $U(x;\beta,\kappa,z)$ is that specified in (1), the function $V(x;\beta,z)$ is that specified in (3), and $\varepsilon_x$ represents randomness in the utility evaluation, assumed to be independent across all the decisions.

This random variable is taken to follow a type 1 extreme value distribution with scale parameter $1/\sigma$. The factor 2 in equation (14) avoids that the magnitude of noise relative to $z$, $\beta$, and $\kappa$ be twice as large for VOI than for non-VOI decisions (recall the factor 1/2 in (1)).

In a VOI decision situation, the selfish option is selected if:

$$\tilde{U}(1;\beta,\kappa,\sigma,z) \geq \tilde{U}(0;\beta,\kappa,\sigma,z), \tag{16}$$

or:

$$2U(1;\beta,\kappa,z) - 2U(0;\beta,\kappa,z) \geq \varepsilon_0 - \varepsilon_1, \tag{17}$$

which reduces to:

$$G - \kappa L - \beta(G+L) \geq \varepsilon_0 - \varepsilon_1. \tag{18}$$

Likewise, in a non-VOI decision situation, the selfish option is selected if:

$$\tilde{V}(1;\beta,\sigma,z) \geq \tilde{V}(0;\beta,\sigma,z), \tag{19}$$

or:

$$V(1;\beta,z) - V(0;\beta,z) \geq \varepsilon_0 - \varepsilon_1, \tag{20}$$

which reduces to:

$$G - \beta(G+L) \geq \varepsilon_0 - \varepsilon_1. \tag{21}$$

Letting $\nu$ be a dummy variable that takes the value 1 in a VOI decision situation and the value 0 in a non-VOI decision situation, it follows that the probability that the selfish option is selected in payoff configuration $z$ equals (McFadden, 1974):

$$\frac{1}{1+\exp\Big[-\sigma\big[\nu\,[(1-\beta)G-(\beta+\kappa)L]+(1-\nu)\,[(1-\beta)G-\beta L]\big]\Big]}, \tag{22}$$

where $\sigma$ is the representative agent's choice sensitivity with respect to differences in deterministic utility. If $\sigma$ is close to 0, the agent chooses either option with a probability close to 1/2, independent of the difference in deterministic utilities. In contrast, arbitrarily large values of $\sigma$ indicate that the probability of choosing the option that results in the higher deterministic utility approaches 1. Noting that the expression in (22) can be written as:

$$H\big(\sigma\big[\nu\,[(1-\beta)G-(\beta+\kappa)L]+(1-\nu)\,[(1-\beta)G-\beta L]\big]\big),$$

where $H:\mathbb{R}\to(0,1)$ is the logistic function, the probability density function can be written:

$$f(\mathbf{x},\beta,\kappa,\sigma)=\prod_{i\in I}\prod_{z_i\in Z}H\left(\sigma\left(U(x^{(z_i)}=1;\beta,\kappa,z_i)-U(x^{(z_i)}=0;\beta,\kappa,z_i)\right)\right)^{\mathbb{1}\{x^{(z_i)}=1\}}$$
$$\times\left[1-H\left(\sigma\left(U(x^{(z_i)}=1;\beta,\kappa,z)-U(x^{(z_i)}=0;\beta,\kappa,z_i)\right)\right)\right]^{1-\mathbb{1}\{x^{(z_i)}=1\}}, \tag{23}$$

where $I$ is the set of subjects, and $z_i$ refers to the situation in which subject $i$ faced payoff $z$: this notation addresses the fact that here all the subjects' decisions are pooled and treated as if taken by one single individual. In the expression, $\mathbf{x}$ is the vector of observed choices, and $\mathbb{1}\{x^{(z_i)}=1\}$ is an indicator function that takes the value 1 if $x^{(z_i)}=1$ and 0 if $x^{(z_i)}=0$. Using (23), we obtain the log-likelihood:

$$\ln(L(\beta,\kappa,\sigma;\mathbf{x}))=\log(f(\mathbf{x},\beta,\kappa,\sigma)), \tag{24}$$

and we will report the vector of estimates $(\hat{\beta},\hat{\kappa},\hat{\sigma})$ that maximizes this log-likelihood.

Second, we use the finite-mixture approach used by Bruhin, Fehr, & Schunk (2019) and van Leeuwen & Alger (2024) to evaluate whether the data is reasonably described by two estimated preference types, and if so, whether the Kantian moral concern appears relevant for both. For each type $k=\{1,2\}$, we estimate the preference parameters $(\beta_k,\kappa_k)$ and the noise parameter $\sigma_k$. The log-likelihood is then given by:

$$\ln(L(\boldsymbol{\beta},\boldsymbol{\kappa},\boldsymbol{\sigma};\mathbf{x}))=\ln\left(\sum_{k=1}^{2}\phi_k\cdot f(\beta_k,\kappa_k,\sigma_k;\mathbf{x})\right), \tag{25}$$

where $\phi_k$ is the share of subjects of type $k$, estimated together with the preference and noise parameters. While this approach could be used for estimation of more than two preference types, we do not report such estimates, because our experimental design is better suited for estimation at the aggregate level (i.e., treating the observations as if they emanated from one or a small number of preference types) than for estimation of each individual's preference type $(\beta_i,\kappa_i)$ (we refer to Appendix A3 for a detailed explanation).

All structural estimations of the preference parameters $\beta$ and $\kappa$ are conducted both for the Neutral frame and for the Market frame, to examine whether any differences appear. Furthermore, we carry out all our estimation procedures for the full samples and for two subsets of the latter, which we call “Core 1” and “Core 2”. Our Core 1 samples consist of the 98 subjects in the Neutral treatment and 100 subjects in the Market treatment who made at least one decision switch between non-VOI and VOI.

Recalling that our theoretical model predicts that a subject should be less inclined to select the selfish decision in VOI situations, we say that a subject switches in the expected direction if she selects the selfish option under non-VOI and the *status quo* option under VOI, and in the unexpected direction if she selects the *status quo* option under non-VOI and the selfish option under VOI. Thus, for our Core 2 samples we remove subjects who made at least as many unexpected as expected switches from the Core 1 samples. This leaves us with 68 subjects in the Neutral treatment and 79 subjects in the Market treatment. In the following section, all reported estimates are obtained using the Core 1 samples, while results with the full and Core 2 samples can be found in Appendix A1.

## 4.2 Results

### 4.2.1 Representative agent estimates

Table 6 shows the representative agent estimates of the preference parameters $\beta$ and $\kappa$, as well as that of the choice sensitivity parameter $\sigma$ for our Core 1 sample for the Neutral frame (first column) and the Market frame (second column).

Both the aheadness aversion and the Kantian morality estimates significantly differ from zero and are positive. They are also both higher in the Neutral than in the Market frame: $\beta = 0.203$ and $\kappa = 0.223$ in the Neutral frame while $\beta = 0.106$ and $\kappa = 0.204$ in the Market frame.[11] The absolute difference is thus larger for the aheadness aversion estimates than for the Kantian morality estimates, and statistically different only for the former, as shown in the third column. Both the Log likelihood ratio (LR) test and the Bayesian Information Criterion

[11] The estimates in van Leeuwen & Alger (2024), reported in their Table 4, show a higher $\beta$ and a lower $\kappa$ than our estimates (for the Neutral frame, which is the most relevant since they also use neutral wording). This could be due to a number of factors, including fundamental differences between the student populations in France and the Netherlands, and systematic differences in preferences for different game classes. This question is left for future research.

Table 6: Estimated preferences for the representative agent in Neutral and Market frames — Core 1 sample

| | Neutral | Market | $H_0$: Neutral = Market$^+$ |
|---|---|---|---|
| $\beta$: Aheadness aversion | 0.203*** | 0.106*** | 0.0075 |
| | (0.030) | (0.021) | |
| $\kappa$: Degree of morality | 0.223*** | 0.204*** | 0.8170 |
| | (0.069) | (0.037) | |
| $\sigma$: Choice sensitivity | 0.026*** | 0.038*** | 0.3313 |
| | (0.008) | (0.009) | |
| LR test ($H_0$: $\beta = \kappa = 0$) | $p < 0.001$ | $p < 0.001$ | |
| Pseudo-$R^2$ | 0.190 | 0.200 | |
| BIC (Full model) | 4,874.68 | 4,993.97 | |
| BIC (Null: $\beta = \kappa = 0$) | 5,997.16 | 6,219.01 | |
| Number of observations | 3,920 | 4,000 | |
| Number of subjects | 98 | 100 | |
| Log likelihood (Full) | −2,424.93 | −2,484.55 | |

Notes: Clustered SE's (subject level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
$^+$ $p$-value of two-sided $z$-test for equality of coefficients across frames.

(BIC) show that the posited model outperforms a model assuming $\beta = \kappa = 0$.

The estimates based on the full sample and the Core 2 samples, reported in Tables A.24-A.25 in Appendix A1, are qualitatively similar, except that for the Core 2 sample, for which the estimated aheadness aversion parameter is not significantly different under the Market than under the Neutral frame.[12]

Finally, to assess the model's out-of-sample predictive power in the Core 1 sample, we performed a three-fold cross-validation using log-loss as the scoring metric. In each iteration, we randomly selected two-thirds of the subjects to estimate the model and computed

[12]This result is partly driven by the clustering of the standard errors at the individual level, as they generate wider confidence intervals compared to non-clustered standard errors.

the log-loss of the remaining third using those parameter estimates. This procedure was applied to both the full specification and a restricted model imposing $\beta = \kappa = 0$, and we then averaged, across the three iterations, the difference in log-loss between the restricted and full models. The full model consistently outperformed the restricted one, reducing average log-loss from approximately 0.69 to 0.62—an improvement of about ten percent in predictive accuracy. This indicates that incorporating the moral preference parameters ($\beta$, $\kappa$) yields systematically better predictions of selfish choice behavior. The procedure is detailed in Appendix A2, where we include Table A.28 with the results of the cross validation exercise.

#### 4.2.2 Finite-mixture estimates

Table 7 presents the finite mixture estimates using a two-type model (recall (25)), as well as the shares of subjects who are classified either as Type 1 or as Type 2, depending on the type that is more aligned with their choices.[13] It is worth underscoring that both types in either frame show some degree of Kantian morality, which is in line with the significant VOI treatment effects presented in Section 3.

Starting with the Neutral frame, Type 1, which accounts for 55.4% of the sample, displays aheadness aversion that is almost twice that of the representative agent (0.327 compared to 0.203), but a Kantian morality that is only about half as large as the representative agent's (0.116 compared to 0.223). By contrast, Type 2 (with a share of 44.6%) seems to be indifferent to advantageous inequality (notice that the -0.065 estimate is not significant at any of the usual confidence levels) and exhibits a much larger Kantian morality estimate than the representative agent (0.342 versus 0.223).

Turning now to the Market frame, a similar pattern emerges. We again see a type (Type 1) who is about twice as aheadness averse as the representative agent –0.203 compared with 0.106– but less morally concerned, with a Kantian morality estimate of 0.153 (versus 0.204 for the representative agent). In addition, we estimate that 56% of subjects in the Market frame belong to Type 1. Type 2 appears to not display aheadness liking nor aversion (the negative $\beta$ estimate of -0.143 is once more not statistically significant at the usual levels). Meanwhile, the Kantian morality estimate of this second type is larger than that of the representative agent (0.325 compared to 0.204). Here, 44% of the subjects are classified as this latter type.

[13]The estimates based on the full sample and the Core 2 samples are reported in Tables A.26-A.27 in Appendix A1.

Table 7: Two-type Finite Mixture Model estimates — Core 1 sample

| | Neutral | | Market | |
|---|---|---|---|---|
| | Type 1 | Type 2 | Type 1 | Type 2 |
| $\beta$: Aheadness aversion | 0.327*** | −0.065 | 0.203*** | −0.143 |
| | (0.082) | (0.258) | (0.015) | (0.117) |
| $\kappa$: Degree of morality | 0.116** | 0.342* | 0.153*** | 0.325*** |
| | (0.049) | (0.199) | (0.023) | (0.072) |
| $\sigma$: Choice sensitivity | 0.046*** | 0.025** | 0.068*** | 0.030*** |
| | (0.009) | (0.012) | (0.013) | (0.010) |
| Share | 0.554*** | 0.446*** | 0.560*** | 0.440*** |
| | (0.173) | (0.173) | (0.056) | (0.056) |
| Observations | 3,920 | | 4,000 | |
| Number of subjects | 98 | | 100 | |
| Log likelihood | −2,155.56 | | −2,193.21 | |

Notes: Clustered SE's (individual level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The estimated parameter values described above allow us to endogenously classify each subject into the preference type that best fits her observed behavior. More precisely, using Bayes' rule, a subject $i$'s posterior probability of belonging to Type $k = 1,2$ is given by:

$$\tau_{ik} = \frac{\hat{\pi}_k f(\hat{\beta}_k, \hat{\kappa}_k, \hat{\sigma}_k; \mathbf{x}_i)}{\sum_{m=1}^{2} \hat{\pi}_m f(\hat{\beta}_m, \hat{\kappa}_m, \hat{\sigma}_m; \mathbf{x}_i)}. \tag{26}$$

These probabilities indicate the type with which the participant's behavior is the most compatible. A model that manages to capture the preferences and share of the underlying types should give way to most posterior probabilities being very close to 0 or to 1. Moreover, the share of subjects classified as belonging to a given type according to these probabilities should be similar to that same type's share estimated by the model. Figure 5 shows that this is indeed the case here.

According to our theory, subjects who are classified as Type 1 should display behavior that is more consistent across the non-VOI and VOI-treatments, as indicated by the relatively low value of $\kappa$. By contrast, those who are classified as Type 2 ought to exhibit less consis-

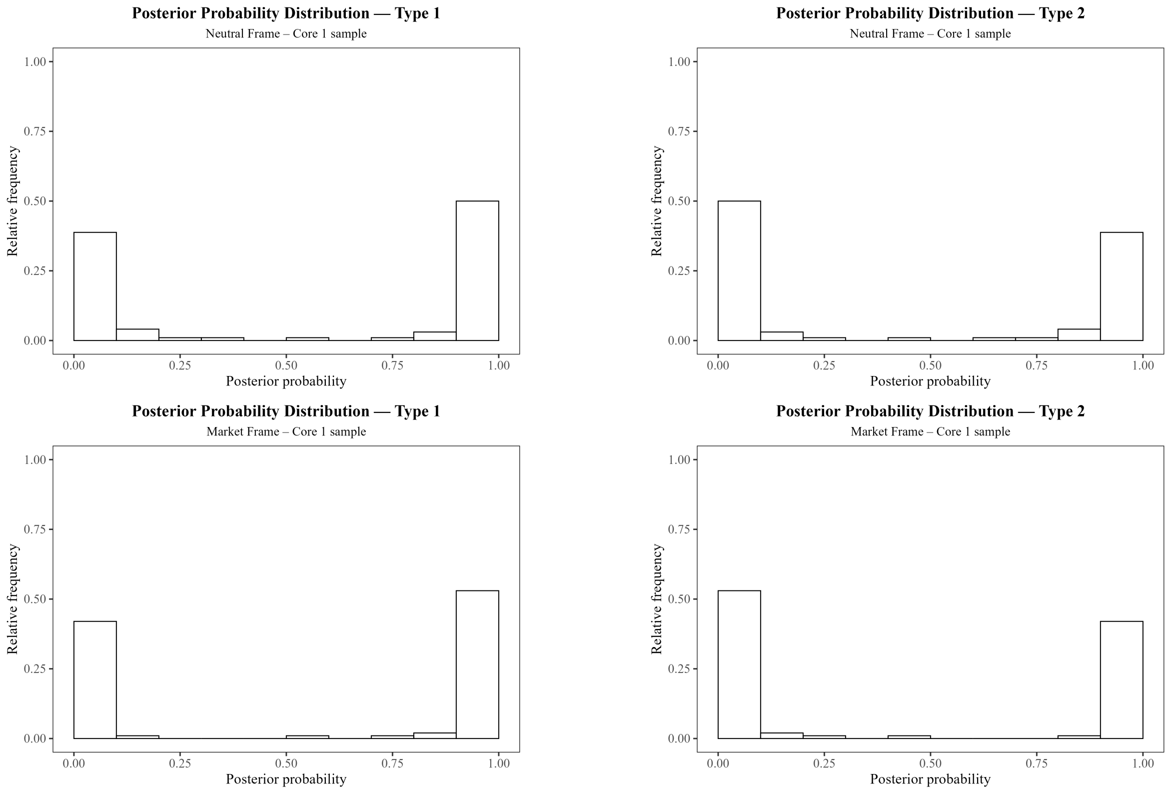
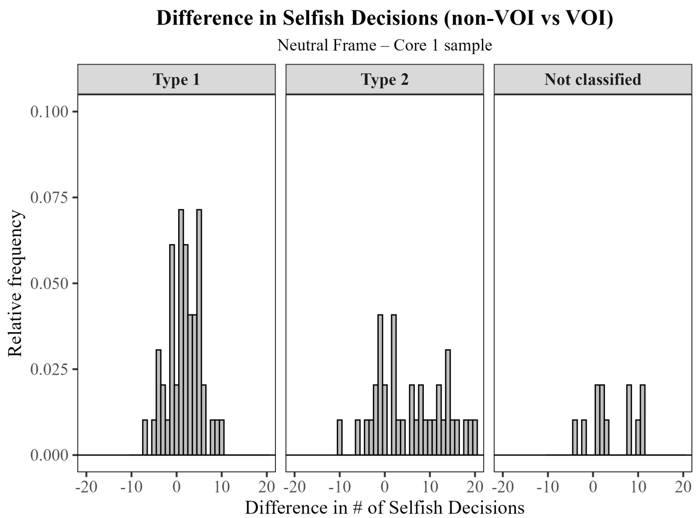


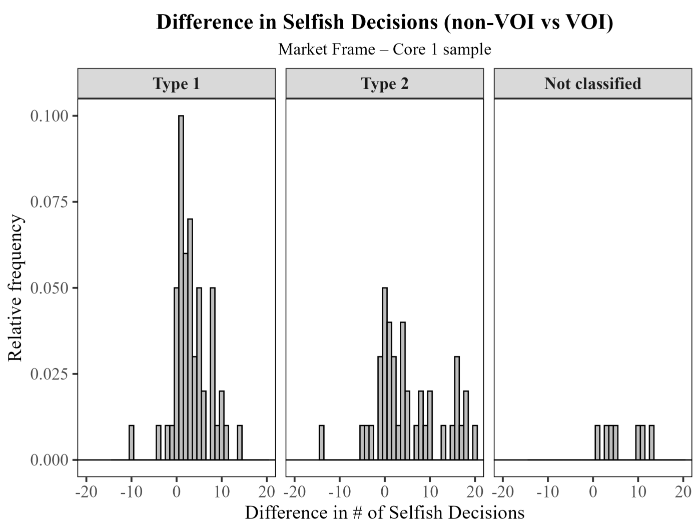

Figure 5: Distribution of posterior probabilities of individual type-membership in Neutral (first row) and Market frames (second row). Core 1 sample estimates.

tent behavior across the non-VOI and VOI-treatments, as suggested by the high value of $\kappa$. This is clearly seen in Figure 6, where we show that the difference in the number of selfish actions in the non-VOI and VOI treatments is more concentrated around zero for participants classified as Type 1.[14]

[14] We refer to Figures A.4-A.7 in Appendix A1 for the corresponding analysis for the full and Core 2 samples.




(a) Neutral frame




(b) Market frame

Figure 6: Difference in selfish actions non-VOI versus VOI. Core 1 sample. Subjects classified as Type 1 if $\tau_1 > 0.95$ or Type 2 if $\tau_1 < 0.05$

# 5 Alternative hypotheses

In this section we discuss possible alternative explanations for the effects of the VOI treatment.

## 5.1 *Ex ante* fairness concerns

We have followed Fehr & Schmidt (1999) by positing that subjects care about the realized differences between own and other's material payoff (i.e., $e_1 - e_2 + x(G+L)$ if active and $e_2 - e_1 + y(G+L)$ if passive), both in non-VOI and VOI decisions. Alternatively, one could assume that in VOI decisions they care about the difference between own and other's expected material payoff, $(x-y)(G+L)$ (Saito (2013)). Would such preferences lead to the same predictions as above?

To answer this question, assume that $i$'s expected utility in a non-VOI decision is still the expression in (3), while her expected utility in a VOI decision reflects *ex ante* fairness concerns. While with our utility specification the individual's belief about the opponent's action is irrelevant, this is not the case with *ex ante* fairness concerns. Letting $\hat{y}$ denote $i$'s belief about $j$'s decision, suppose that her utility in a VOI decision is

$$\begin{aligned} U_i(x,\hat{y}) &= \frac{1}{2}\cdot(e_1+xG)+\frac{1}{2}\cdot(e_2-\hat{y}L) \\ &- \frac{1}{2}\cdot\beta_i\cdot\max\{0,(x-\hat{y})(G+L)] \\ &- \frac{1}{2}\cdot\alpha_i\cdot\max\{0,(\hat{y}-x)(G+L)]. \end{aligned} \tag{27}$$

It is straightforward to see that $i$ would select the *status quo* option in the VOI decision in the following cases only (for conciseness we omit the knife-edge cases where the preference parameters are just equal to the threshold values):

1. she believes that $j$ would select the selfish option ($\hat{y}=1$) and exhibits a sufficiently strong love towards being behind, $\alpha_i < -G/(G+L)$;

2. she believes that $j$ would select the *status quo* option ($\hat{y}=0$) and is sufficiently averse towards being ahead, $\beta_i > G/(G+L)$.

Recalling that $i$ selects the selfish option in a non-VOI decision only if not strongly averse towards being ahead, $\beta_i < G/(G+L)$ (see (6)), we see that this condition contradicts the condition on $\beta$ in case 2. Hence, $i$ switches from the selfish option in a non-VOI decision to the *status quo* decision in the payoff-equivalent VOI decision if and only if $i$ is: (a) not too strongly averse towards being ahead ($\beta_i < G/(G+L)$), (b) pessimistic about $j$'s action ($\hat{y} = 1$), and (c) sufficiently loving towards being behind the other materially ($\alpha_i < -G/(G+L)$). In sum, a subject's preferences would need to conform with the combined requirement that $\alpha < 0$ and $|\alpha| > \beta$. While we cannot test the relevance of this reasoning, since we did not elicit the subjects' beliefs, the literature indicates that behindness loving is rare. Moreover, when it does appear to be supported by the data, estimates of $\alpha$ and $\beta$ contradict the combined requirement that $\alpha < 0$ and $|\alpha| > \beta$ (Bruhin et al. (2019); van Leeuwen & Alger (2024)).

Furthermore, with this specification there could be switches in the opposite direction: an individual who is aheadness averse enough ($\beta_i > G/(G+L)$) to select the *status quo* option in the non-VOI decision, and who believes the other will select the selfish option if active ($\hat{y} = 1$), would select the selfish option in the VOI decision as long as she does not exhibit a strong enough love towards being behind ($\alpha_i > -G/(G+L)$). While we do observe some switches in this direction, most of the switches are in the direction expected under our specification; the ratio of expected vs unexpected switches was 2.8 under the Neutral frame and 4 under the Market frame.

Note, finally, that if an individual is subject to a false consensus effect (Ross et al., 1977; Butler et al., 2015), whereby she tends to believe that the opponent will select the same action as her, then she would always select the selfish option in a VOI decision, since the fairness concern terms are both nil if $\hat{y} = x$.

## 5.2 Uncertainty about the behavioral consequence

Compared to non-VOI, the VOI treatment introduces uncertainty as to whether the subject's decision will be carried out or not. Chen & Schonger (2022) conduct an experiment in which they vary this uncertainty, by shredding the envelope containing the decision with some probability; they show that the average donation amount in dictator games with giving is higher is decreasing in the probability that the decision is carried out. Their interpretation is that as the probability that the decision is carried out decreases, it becomes cheaper to achieve deontological motivations.

A key difference is thus that we render the arbitrariness of the role distribution salient, thus making universalization concerns natural; they render the uncertainty of the decision execution itself salient, without any reference to role distribution.[15]

Another key difference is that we hypothesize that the utility term corresponding to the Kantian concern depends on the material payoffs, while for Chen & Schonger (2022) the deontological motivation is a black box. The fact that our estimates of $\kappa$ are significantly different from 0 suggest that the data supports our hypothesis.

### 5.3 Magical thinking

Our experimental design does not allow to rule out magical thinking as a potential driver of behavior Shafir & Tversky (1992); Daley & Sadowski (2017), according to which subjects would have a tendency to falsely believe that their own action influences the action of others. However, the findings of van Leeuwen & Alger (2024) suggest that such false beliefs cannot be the sole explanation; their experiment involved only VOI decisions in sequential games, and they estimate the preference parameters controlling for the subjects' beliefs about the opponent's strategy choice, and they still find that including Kantian concerns enhances the explanatory power.

## 6 Concluding remarks

In order to issue effective and desirable policy recommendations, economists need to achieve a realistic understanding of the motivations of human behavior (Bowles, 2016; Duflo, 2017). Indeed, insights about the functioning of markets and institutions derived in theoretical models where humans are driven solely by material self-interest may be misleading. We carried out an experiment aimed at testing whether subjects' decisions are compatible with preferences combining Kantian moral motivations and other-regarding concerns. Our design allows us to disentangle the influence of Kantian and other-regarding concerns in a setting that captures the essentials of a decision sitting at the origin of several market inefficiency results: that of selling a lemon – i.e., a good with hidden quality issues whose purchase actually makes the buyer worse off. Our experiment further allows us to structurally estimate preference parameters, and to test whether the obtained estimates differ between decisions

[15] Accordingly, in our design there are only two probabilities that the decision will be carried out – 1 and 1/2; in their experiment, they included more than two probabilities of decision execution.

taken in a frame that describes the situation as the sale of a lemon – the market frame – and those taken in a standard neutral frame.

The subjects in our experiment were on average much more prone to "sell lemons" in the market frame than they were to select the payoff-equivalent decisions in the neutral frame. The propensity to sell lemons was, however, strongly mitigated when the instructions made explicit reference to the arbitrariness in the role distribution. This result is in line with our theoretical model, according to which such instructions are expected to awaken or strengthen subjects' Kantian moral concerns. As per our structural estimations of this Kantian moral concern and the aversion towards being ahead materially compared to the other subject in a match, the difference in behavior between the market and the neutral frame was driven mainly by a weaker aheadness aversion in the market frame.[16]

Our results are in line with those of other studies that have also shown that preferences that combine a Kantian moral concern with other-regard significantly enhances the explanatory power compared to preferences without the Kantian moral concern (Miettinen, Kosfeld, Fehr, & Weibull, 2020; van Leeuwen & Alger, 2024). Like van Leeuwen & Alger (2024), who also estimate aheadness aversion and Kantian moral preference parameters, we find estimates which indicate that some individuals appear to be driven by a combination of altruism (a positive $\beta$) and a Kantian moral concern, and all the estimates of the Kantian moral concern parameter are positive and significantly so. Interestingly, under the Market frame some individuals appear to be driven by a combination of spite (a negative $\beta$) and a Kantian moral concern. This suggests that in our experiment the Market frame may have triggered more competitive preferences.

Our study is the first to explore – and find support for – the hypothesis that the Kantian moral concern is (at least somewhat) muted in situations where role uncertainty is not made explicit. If externally valid, this result would mean that the mere mention of role uncertainty in situations outside of the experimental laboratory could be used to awaken Kantian moral concerns, and thus trigger behavioral changes. In this line, Bowles (2016, 2023) suggests that indeed, in order to improve the functioning of markets, policies ought to combine both incentives and *moral messages*, leveraging synergies between the two (Kranton, 2019). Field experiments might be appropriate to examine this question in the future.

[16]Note, however, that this discrepancy could also be driven by differences between the frames in the subjective beliefs that the role distribution could have differed from the realized one in the non-VOI condition (recall $\hat{p}_i$) in (8). This hypothesis is left for future research.

Our results further strongly suggest that both frames in our experiment matter. On average, decisions are more selfish under the Market than under the Neutral frame, and the difference is large. However, the VOI wording also has a stronger dampening effect on selfish actions under the Market frame. Accordingly, the estimates of the preferences based on subjects' decisions in market interactions differ from those based on decisions in interactions with similar payoff consequences but presented with a neutral frame. The difference is particularly marked for subjects' aheadness aversion, thus indicating that other regard is the main driving force behind the difference in behavior across frames.

The detected disparities in preferences across different social situations would be in line with the theory of preference evolution, which predicts that as long as humans in our evolutionary past could discriminate between classes of interactions (say, market vs non-market), preferences would be expected to depend on the specifics of each class of interaction (for surveys, see Alger & Weibull (2019) and Alger (2023)). As such, they strengthen the more general case made elsewhere for further research on how the language used to describe a task matters for subjects' decisions (Alekseev, Charness, & Gneezy, 2017; Capraro, Halpern, & Perc, 2024).

As a final note, it is worth mentioning that recent theoretical studies highlight the role that Kantian moral concerns can play in mitigating inefficiencies in equilibrium outcomes in bilateral trade situations plagued by information asymmetries (Rivero-Wildemauwe, 2023, 2025). While our study focuses solely on the behavior of potential "lemon" sellers, by shutting down the buyers' decisions, these works find that morality decreases or even eliminates the exchange of "lemons" *in equilibrium*. Moreover, they document that altruism has effects that go in the same direction, although they are generally weaker. It would thus be worth extending our present work to capture interactions between both sides of the market. With both sellers and buyers being active, beliefs should also play a role; the evidence that preferences may be affected by framing can help explain why beliefs would also be affected by framing (Dufwenberg, Gächter, & Hennig-Schmidt, 2011).

# References

Agneman, G., & Chevrot-Bianco, E. (2022). Market participation and moral decision-making: Experimental evidence from Greenland. *Economic Journal*, *133*(650), 537-581.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and Identity. *The Quarterly Journal of Economics*, *115*(3), 715–753.

Alekseev, A., Charness, G., & Gneezy, U. (2017). Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior & Organization*, *134*, 48-59.

Alger, I. (2023). Evolutionarily stable preferences. *Philosophical Transactions B*, *378*, 20210505.

Alger, I., & Weibull, J. W. (2013). Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica*, *81*(6), 2269–2302.

Alger, I., & Weibull, J. W. (2017). Strategic behavior of moralists and altruists. *Games*, *8*(3).

Alger, I., & Weibull, J. W. (2019). Evolutionary models of preference formation. *Annual Review of Economics*, *11*, 329–354.

Alger, I., Weibull, J. W., & Lehmann, L. (2020). Evolution of preferences in structured populations: Genes, guns, and culture. *Journal of Economic Theory*, *185*, 104951.

Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *Economic Journal*, *103*(418), 570–585.

Bartling, B., & Özdemir, Y. (2023). The limits to moral erosion in markets: Social norms and the replacement excuse. *Games and Economic Behavior*, *138*, 143–160.

Becker, G. S. (1974). A theory of social interactions. *Journal of Political Economy*, *82*(6), 1063–1093.

Bénabou, R., Falk, A., & Henkel, L. (2024). *Ends versus means: Kantians, utilitarians, and moral decisions* (NBER Working Paper 32073).

Bénabou, R., Falk, A., Henkel, L., & Tirole, J. (2023). *Eliciting moral preferences: Theory and experiment* (CRC TR 224 Discussion Paper Series).

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*(5), 1652-1678.

Bernheim, B. D., Braghieri, L., Martínez-Marquina, A., & Zuckerman, D. (2021). A theory of chosen preferences. *American Economic Review*, *111*(2), 720–54.

Bisin, A., & Verdier, T. (2023). Advances in the economic theory of cultural transmission [Journal Article]. *Annual Review of Economics*, *15*(Volume 15, 2023), 63-89.

Boissonnet, N., Ghersengorin, A., & Gleyze, S. (2023). Revealed deliberate preference change. *Games and Economic Behavior*, *142*, 357-367.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, *36*(1), 75–111.

Bowles, S. (2016). *The moral economy – why good incentives are no substitute for good citizens*. New Haven, CT: Yale University Press.

Bowles, S. (2023). Moral economics. *Fiscal Studies*, *44*(2), 151-160.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, *14*(3), 375–398.

Bruhin, A., Fehr, E., & Schunk, D. (2019). The many faces of human sociality: Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association*, *17*(4), 1025–1069.

Bursztyn, L., Fiorin, S., Gottlieb, D., & Kanz, M. (2019). Moral incentives in credit card debt repayment: Evidence from a field experiment. *Journal of Political Economy*, *127*(4), 1641–1683.

Butler, J. V., Giuliano, P., & Guiso, L. (2015). Trust, values, and false consensus. *International Economic Review*, *56*(3), 889-915.

Byambadalai, U., Ma, C. t. A., & Wiesen, D. (2023). Changing preferences: An experiment and estimation of market-incentive effects on altruism. *Journal of Health Economics*, *92*, 102808.

Cabrales, A., Miniaci, R., Piovesan, M., & Ponti, G. (2010). Social preferences and strategic uncertainty: An experiment on markets and contracts. *American Economic Review*, *100*(5), 2261–2278.

Capraro, V., Halpern, J. Y., & Perc, M. (2024, March). From outcome-based to language-based preferences. *Journal of Economic Literature*, *62*(1), 115-54.

Capraro, V., & Rand, D. G. (2018). Do the right thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. *Judgment and Decision Making*.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *117*(3), 817–869.

Chen, D. L., & Schonger, M. (2022). Social preferences or sacred values? Theory and evidence of deontological motivations. *Science Advances*, *8*(19).

Daley, B., & Sadowski, P. (2017). Magical thinking: A representation result. *Theoretical Economics*, *12*(2), 909–956.

Dreber, A., Ellingsen, T., Johannesson, M., & Rand, D. G. (2013). Do people care about social context? Framing effects in dictator games. *Experimental Economics*, *16*(3), 349–371.

Duflo, E. (2017). The Economist as Plumber. *American Economic Review*, *107*(5), 1–26.

Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, *73*(2), 459-478.

Dufwenberg, M., Johansson-Stenman, O., Kirchler, M., Lindner, F., & Schwaiger, R. (2022). Mean markets or kind commerce? *Journal of Public Economics*, *209*, 104648.

Eichner, T., & Pethig, R. (2022). Kantians defy the economists' mantra of uniform pigovian emissions taxes. *Ecological Economics*, *200*, 107514.

Engel, J., & Szech, N. (2020). A little good is good enough: Ethical consumption, cheap excuses, and moral self-licensing. *PLoS ONE*, *15*(1), 1–19.

Engelmann, J. B., Meyer, F., Ruff, C. C., & Fehr, E. (2019). The neural circuitry of affect-induced distortions of trust. *Science Advances*, *5*(3).

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *Review of Economic Studies*, *83*(2), 587–628.

Falk, A., & Szech, N. (2013). Morals and markets. *Science*, *340*(6133), 707-711.

Feess, E., Kerzenmacher, F., & Timofeyev, Y. (2022). Utilitarian or deontological models of moral behavior—What predicts morally questionable decisions? *European Economic Review*, *149*, 104264.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, *97*(5), 1858–1876.

García-Pola, B., Iriberri, N., & Kovářík, J. (2020). Hot versus cold behavior in centipede games. *Journal of the Economic Science Association*, *6*(2), 226–238.

Glimcher, P. W. (2022). Efficiently irrational: deciphering the riddle of human choice. *Trends in Cognitive Sciences*, *26*(8), 669-687.

Grech, P. D., Nax, H. H., & Soos, A. (2022). Incentivization matters: A meta-perspective on dictator games. *Journal of the Economic Science Association*, *8*(1), 34–44.

Heckman, J. J., & Corbin, C. O. (2016). Capabilities and skills. *Journal of Human Development and Capabilities*, *17*(3), 342–359.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., … Tracer, D. (2005). “Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(06), 795–855.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Bolyanatz, A., Cardenas, J. C., … Ziker, J. (2016). Markets , religion , community size , and the evolution of fairness and punishment. *Science*, *327*(5972), 1480–1484.

Herne, K., Hietanen, J. K., Lappalainen, O., & Palosaari, E. (2022). The influence of role awareness, empathy induction and trait empathy on dictator game giving. *PLoS ONE*, *17*(3), 1–19.

Hirschman, A. O. (1982). Rival Interpretations of Market Society: Civilizing, Destructive, or Feeble? *Journal of Economic Literature*, *20*, 1463–1484.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior*, *7*(3), 346–380.

Huang, K., Greene, J. D., & Bazerman, M. (2019). Veil-of-ignorance reasoning favors the greater good. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(48), 23989–23995.

Huck, S., Lünser, G. K., & Tyran, J.-R. (2016). Price competition and reputation in markets for experience goods: an experimental study. *RAND Journal of Economics*, *47*(1), 99–117.

Iriberri, N., & Rey-Biel, P. (2011). The role of role uncertainty in modified dictator games. *Experimental Economics*, *14*(2), 160–180.

Juan-Bartroli, P., & Karagözoğlu, E. (2024). Moral preferences in bargaining. *Economic Theory*. Retrieved from https://doi.org/10.1007/s00199-023-01544-7

Kay, A. C., & Ross, L. (2003). The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the prisoner's dilemma. *Journal of Experimental Social Psychology*, *39*(6), 634-643.

Kettlewell, N., & Tymula, A. (2021). The australian twins economic preferences survey. *Twin Research and Human Genetics*, *24*(6), 365–370.

Kettlewell, N., & Tymula, A. (2024). Heritability across different domains of trust. *Journal of Economic Behavior & Organization*, *219*, 549-563.

Kirman, A., & Teschl, M. (2010). Do markets foster selfishness? *Revue de philosophie économique*, *Vol. 11*(1), 113–140.

Kranton, R. (2019). The devil is in the details: Implications of Samuel Bowles's *The Moral Economy* for economics and policy research. *Journal of Economic Literature*, *57*(1), 147–160.

Laslier, J.-F. (2023). Universalization and altruism. *Social Choice and Welfare*, *60*, 579–594.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*(3), 593–622.

Levine, S., Kleiman-Weiner, M., Schulz, L., Tenenbaum, J., & Cushman, F. (2020). The logic of universalization guides moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(42), 26158–26169.

Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining Prisoner's Dilemma game moves. *Personality and Social Psychology Bulletin*, *30*(9), 1175–1185.

Mesa-Vázquez, E., Rodriguez-Lara, I., & Urbano, A. (2021). Standard vs random dictator games: On the effects of role uncertainty and framing on generosity. *Economics Letters*, *206*, 109981.

Miettinen, T., Kosfeld, M., Fehr, E., & Weibull, J. W. (2020). Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions. *Journal of Economic Behavior & Organization*, *173*, 1-25.

Muñoz-Sobrado, E. (2022). *Taxing moral agents* (Tech. Rep.). CESifo Working Paper No. 9867.

Nagler, M. G. (2023). Thoughts matter: a theory of motivated preference. *Theory and Decision*, *94*(2), 211–247.

Ortiz-Riomalo, J. F., Koessler, A. K., & Engel, S. (2021). Inducing perspective-taking for prosocial behaviour in natural resource management. *Journal of Environmental Economics and Management*, *110*, 102513.

Pivato, M. (2023). *Universal recursive preference structures* (Tech. Rep.). Université Paris 1 Panthéon-Sorbonne.

Rivero-Wildemauwe, J. I. (2023). *Moral motivations in sequential buyer-seller interactions with adverse selection* (Tech. Rep.). THEMA (THéorie Economique, Modélisation et Applications), CY Cergy Paris Université.

Rivero-Wildemauwe, J. I. (2025). *Trade among moral agents with information asymmetries* (Technical Report No. arXiv:2505.20551). arXiv. (Available at https://arxiv.org/abs/2505.20551)

Roemer, J. E. (2010). Kantian equilibrium. *Scandinavian Journal of Economics*, *112*(1), 1–24.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279-301.

Roth, B. A. E., & Schoumaker, F. (1983). Expectations and reputations in bargaining: An experimental study. *American Economic Review*, *73*(3), 362–372.

Saito, K. (2013). Social preferences under risk: Equality of opportunity versus equality of outcome. *American Economic Review*, *103*(7), 3084–3101.

Salonia, E. M. (2024). *A foundation for universalisation in games* (TSE Working Papers No. 1586). Toulouse School of Economics (TSE).

Sarkisian, R. (2021). Screening teams of moral and altruistic agents. *Games*, *12*(4).

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, *24*(4), 449–474.

Sutter, M., Huber, J., Kirchler, M., Stefan, M., & Walzl, M. (2020). Where to look for the morals in markets? *Experimental Economics*, *23*(1), 30–52.

Sutter, M., & Weck-Hannemann, H. (2003). Taxation and the veil of ignorance - A real effort experiment on the Laffer curve. *Public Choice*, *115*(1-2), 217–240.

Thöni, C., & Gächter, S. (2015). Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis. *Journal of Economic Psychology*, *48*, 72-88.

van Leeuwen, B., & Alger, I. (2024). Estimating Social Preferences and Kantian Morality in Strategic Interactions. *Journal of Political Economy Microeconomics*, *2*(4), 665–706.

# FOR ONLINE PUBLICATION

## Appendix A1 Additional tables and figures

Table A.1: Participants' socio-demographic characteristics

| Variable | N | Mean | Variable | N | Mean |
|---|---|---|---|---|---|
| **Age** | 453 | 21 | **Education** | 448 | |
| **Gender** | 453 | | ... Highschool | 37 | 8% |
| ... Female | 193 | 43% | ... Undergrad. | 272 | 61% |
| ... Male | 260 | 57% | ... Engineer | 5 | 1% |
| **Background** | 418 | | ... Master 1 | 68 | 15% |
| ... Economics | 339 | 81% | ... Master 2 | 44 | 10% |
| ... Sciences, Eng. | 10 | 2% | ... PhD | 3 | 1% |
| ... Political Science | 47 | 11% | ... Other | 19 | 4% |
| ... Languages | 8 | 2% | **Experience** | 453 | |
| ... Literature, Philo. | 11 | 3% | ... Yes | 140 | 31% |
| ... History, Geog. | 3 | 1% | ... No | 313 | 69% |
| **Nationality** | 453 | | **Couple** | 453 | |
| ... French | 362 | 80% | ... Yes | 66 | 15% |
| ... Other | 91 | 20% | ... No | 387 | 85% |

For Age: Std. Dev. = 3.9, Min = 18, Pctl. 25 = 19, Pctl. 75 = 22, Max = 54. Five participants did not report their education level, while 35 did not report their background.

Table A.2: Participants' socio-demographic characteristics

| Treatment | N | | M | | A | | B | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | N | Mean | N | Mean | N | Mean | N | Mean | Test |
| **Age** | 108 | 21.2 | 112 | 22 | 121 | 20.5 | 112 | 20.3 | F: 0.003 |
| **Gender** | 108 | | 112 | | 121 | | 112 | | $\chi^2$ : 0.29 |
| ... Female | 53 | 49% | 44 | 39% | 54 | 45% | 42 | 38% | |
| ... Male | 55 | 51% | 68 | 61% | 67 | 55% | 70 | 62% | |
| **Education** | 108 | | 112 | | 118 | | 110 | | $\chi^2$ : 0.12 |
| ... Highschool | 10 | 9% | 7 | 6% | 11 | 9% | 9 | 8% | |
| ... Undergrad. | 60 | 56% | 67 | 60% | 73 | 62% | 72 | 65% | |
| ... Engineer | 1 | 1% | 2 | 2% | 1 | 1% | 1 | 1% | |
| ... Master 1 | 25 | 23% | 18 | 16% | 17 | 14% | 8 | 7% | |
| ... Master 2 | 10 | 9% | 13 | 12% | 11 | 9% | 10 | 9% | |
| ... PhD | 1 | 1% | 0 | 0% | 2 | 2% | 0 | 0% | |
| ... Other | 1 | 1% | 5 | 4% | 3 | 3% | 10 | 9% | |
| **Background** | 103 | | 99 | | 111 | | 105 | | $\chi^2$ : 0.034 |
| ... Economics | 84 | 82% | 78 | 79% | 98 | 88% | 79 | 75% | |
| ... Sciences, Eng. | 3 | 3% | 2 | 2% | 4 | 4% | 1 | 1% | |
| ... Political Science | 9 | 9% | 16 | 16% | 7 | 6% | 15 | 14% | |
| ... Languages | 1 | 1% | 0 | 0% | 1 | 1% | 6 | 6% | |
| ... Literature, Philo. | 5 | 5% | 2 | 2% | 0 | 0% | 4 | 4% | |
| ... History, Geog. | 1 | 1% | 1 | 1% | 1 | 1% | 0 | 0% | |
| **Experience** | 108 | | 112 | | 121 | | 112 | | $\chi^2$ : 0.023 |
| ... Yes | 28 | 26% | 44 | 39% | 28 | 23% | 40 | 36% | |
| ... No | 80 | 74% | 68 | 61% | 93 | 77% | 72 | 64% | |
| **Couple** | 108 | | 112 | | 121 | | 112 | | $\chi^2$ : 0.551 |
| ... Yes | 15 | 14% | 13 | 12% | 22 | 18% | 16 | 14% | |
| ... No | 93 | 86% | 99 | 88% | 99 | 82% | 96 | 86% | |
| **Nationality** | 108 | | 112 | | 121 | | 112 | | $\chi^2$ : 0.818 |
| ... French | 83 | 77% | 90 | 80% | 99 | 82% | 90 | 80% | |
| ... Other | 25 | 23% | 22 | 20% | 22 | 18% | 22 | 20% | |

Standard deviations for Age are respectively: 3.6, 5.6, 2.9 and 2.4. The last column reports the p-value of an independence test between each variable and the treatment arm. An F test is used for the "Age" variable and "$\chi^2$" for the remaining ones. Five participants did not report their education level, while 35 did not report their background.

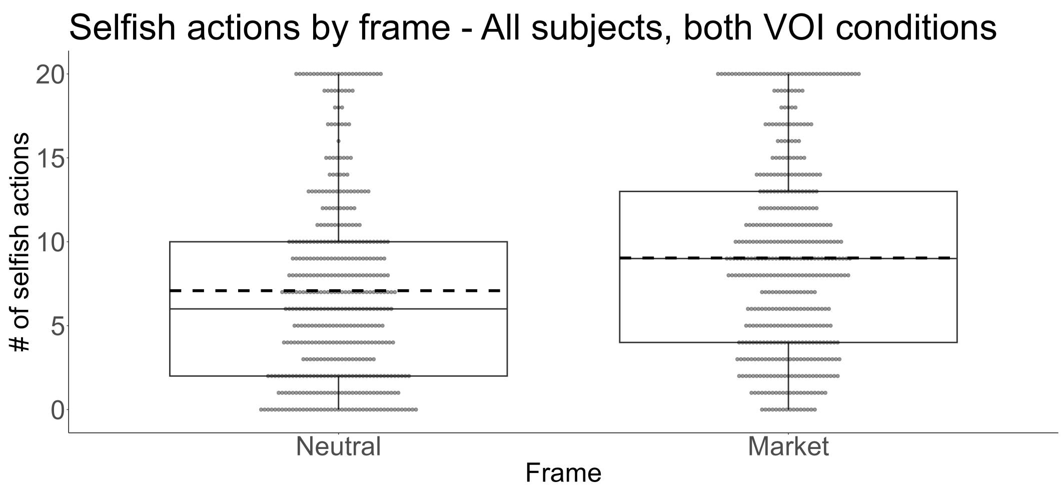



Figure A.1: Each dot represents one observation, defined as the total number of times that one subject selected the selfish option in one sequence of 20 decisions. Each subject was presented with two such sequences, hence each subject is represented by two observations. The left part shows the $n = 453$ observations collected under the Neutral frame in sequences $N1$, $N2$, $A1$, and $B2$, while the right part shows the $n = 453$ observations collected under the Market frame in sequences $M1$, $M2$, $A2$, and $B1$. The two-sided t-test for the difference in means results in a $p$-value of 0.0000004. The Wilcoxon rank-sum test for the difference in medians results in a $p$-value of 0.0000004.

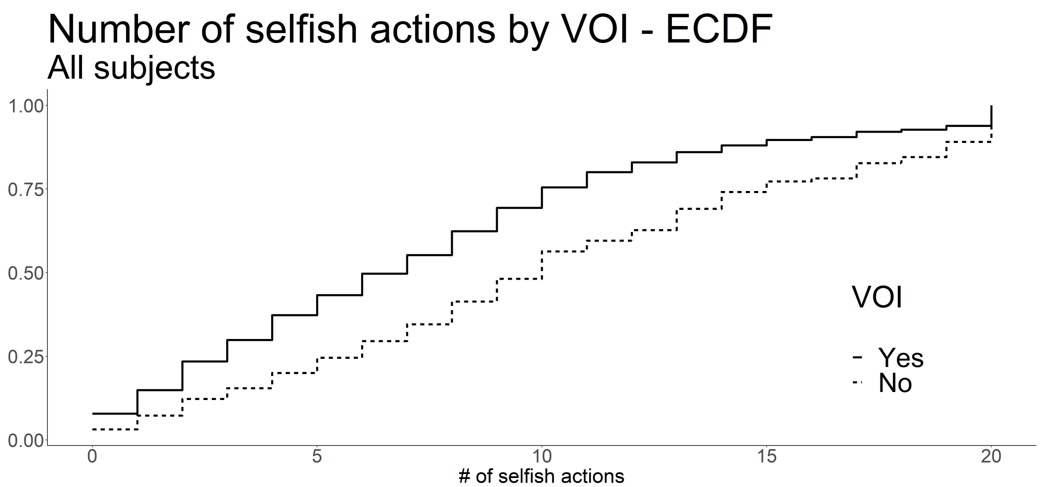



Figure A.2: Cumulative distributions of observations—defined as the total number of times that one subject selected the selfish option in one sequence of 20 decisions—for decisions collected in the VOI sequences (solid line, for the $n = 686$ observations in sequences $N2$, $M2$, $A1$, $A2$, $B1$, and $B2$) and those collected in the non-VOI sequences (dashed line, for the $n = 220$ observations collected in sequences $N1$ and $M1$). The Kolmogorov–Smirnov test's $D$ statistic was 0.21206, while the *p-value* was .0000006231.

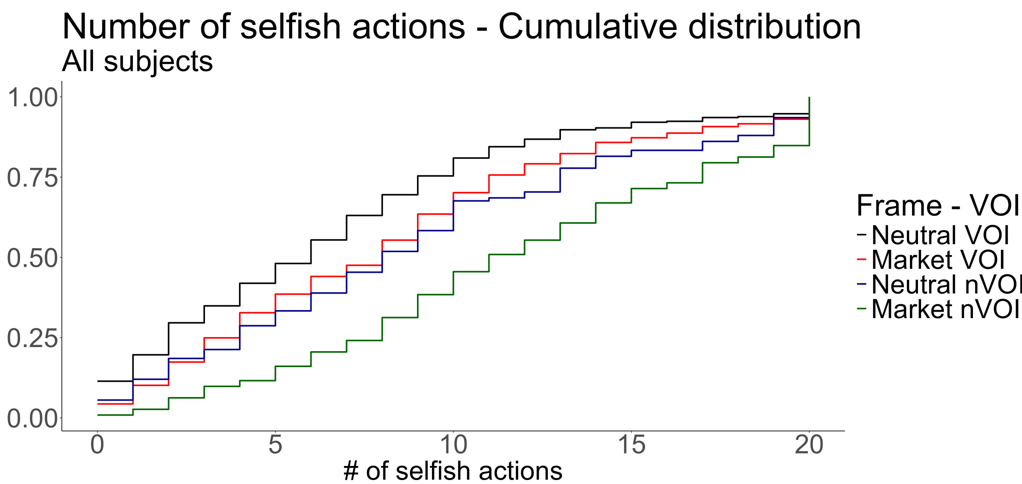



Figure A.3: Cumulative distributions of observations—defined as the total number of times that one subject selected the selfish option in one sequence of 20 decisions—for decisions collected in the Neutral VOI sequences $N2$, $A1$, and $B2$ (black line, $n = 341$), the Market VOI sequences $M2$, $A2$, and $B1$ (red line, $n = 345$), the Neutral non-VOI sequence $N1$ (blue line, $n = 108$), and the Market non-VOI sequences $M1$ (green line, $n = 112$). All pairwise Kolmogorov-Smirnoff tests result in $p$-values at or below 0.01, with the exception of Market VOI versus Neutral non-VOI, which gives a $p$-value of 0.6.

Table A.3: Share of subjects choosing selfish action - By Frame

| | n-VOI - between (N1-M1) | | VOI - between (A1-B1) | | VOI - within (A1-A2-B1-B2) | |
|---|---|---|---|---|---|---|
| Payoff | Neutral | Market | Neutral | Market | Neutral | Market |
| 1 | 0.69 | 0.84 | 0.59 | 0.80 | 0.65 | 0.77 |
| 2 | 0.69 | 0.82 | 0.55 | 0.58 | 0.60 | 0.65 |
| 3 | 0.64 | 0.74 | 0.54 | 0.56 | 0.62 | 0.66 |
| 4 | 0.49 | 0.72 | 0.48 | 0.49 | 0.50 | 0.50 |
| 5 | 0.53 | 0.72 | 0.45 | 0.61 | 0.45 | 0.57 |
| 6 | 0.53 | 0.67 | 0.42 | 0.44 | 0.40 | 0.48 |
| 7 | 0.45 | 0.66 | 0.37 | 0.47 | 0.34 | 0.48 |
| 8 | 0.51 | 0.60 | 0.36 | 0.34 | 0.31 | 0.39 |
| 9 | 0.52 | 0.73 | 0.38 | 0.53 | 0.35 | 0.52 |
| 10 | 0.39 | 0.60 | 0.31 | 0.25 | 0.23 | 0.32 |
| 11 | 0.38 | 0.58 | 0.29 | 0.30 | 0.27 | 0.33 |
| 12 | 0.36 | 0.50 | 0.32 | 0.32 | 0.26 | 0.34 |
| 13 | 0.42 | 0.46 | 0.32 | 0.42 | 0.36 | 0.40 |
| 14 | 0.34 | 0.39 | 0.24 | 0.29 | 0.25 | 0.31 |
| 15 | 0.31 | 0.54 | 0.36 | 0.30 | 0.28 | 0.30 |
| 16 | 0.36 | 0.46 | 0.30 | 0.21 | 0.23 | 0.24 |
| 17 | 0.24 | 0.39 | 0.24 | 0.34 | 0.19 | 0.33 |
| 18 | 0.44 | 0.60 | 0.25 | 0.24 | 0.19 | 0.26 |
| 19 | 0.29 | 0.35 | 0.35 | 0.27 | 0.26 | 0.29 |
| 20 | 0.29 | 0.30 | 0.26 | 0.24 | 0.21 | 0.25 |
| Mean | 0.44 | 0.58 | 0.37 | 0.40 | 0.35 | 0.42 |

Table A.4: Share of subjects choosing selfish action - By VOI

| | Neutral, bet. (N1-A1) | | Neutral, within (N1-N2) | | Market, within (M1-M2) | | Market, bet. (M1-B1) | |
|---|---|---|---|---|---|---|---|---|
| Payoff | non-VOI | VOI | non-VOI | VOI | non-VOI | VOI | non-VOI | VOI |
| 1 | 0.69 | 0.59 | 0.69 | 0.60 | 0.84 | 0.83 | 0.84 | 0.80 |
| 2 | 0.69 | 0.55 | 0.69 | 0.42 | 0.82 | 0.74 | 0.82 | 0.58 |
| 3 | 0.64 | 0.54 | 0.64 | 0.60 | 0.74 | 0.83 | 0.74 | 0.56 |
| 4 | 0.49 | 0.48 | 0.49 | 0.40 | 0.72 | 0.61 | 0.72 | 0.49 |
| 5 | 0.53 | 0.45 | 0.53 | 0.40 | 0.72 | 0.53 | 0.72 | 0.61 |
| 6 | 0.53 | 0.42 | 0.53 | 0.31 | 0.67 | 0.46 | 0.67 | 0.44 |
| 7 | 0.45 | 0.37 | 0.45 | 0.30 | 0.66 | 0.41 | 0.66 | 0.47 |
| 8 | 0.51 | 0.36 | 0.51 | 0.28 | 0.60 | 0.34 | 0.60 | 0.34 |
| 9 | 0.52 | 0.38 | 0.52 | 0.30 | 0.73 | 0.39 | 0.73 | 0.53 |
| 10 | 0.39 | 0.31 | 0.39 | 0.23 | 0.60 | 0.29 | 0.60 | 0.25 |
| 11 | 0.38 | 0.29 | 0.38 | 0.26 | 0.58 | 0.31 | 0.58 | 0.30 |
| 12 | 0.36 | 0.32 | 0.36 | 0.19 | 0.50 | 0.27 | 0.50 | 0.32 |
| 13 | 0.42 | 0.32 | 0.42 | 0.29 | 0.46 | 0.32 | 0.46 | 0.42 |
| 14 | 0.34 | 0.24 | 0.34 | 0.23 | 0.39 | 0.26 | 0.39 | 0.29 |
| 15 | 0.31 | 0.36 | 0.31 | 0.23 | 0.54 | 0.26 | 0.54 | 0.30 |
| 16 | 0.36 | 0.30 | 0.36 | 0.19 | 0.46 | 0.21 | 0.46 | 0.21 |
| 17 | 0.24 | 0.24 | 0.24 | 0.17 | 0.39 | 0.27 | 0.39 | 0.34 |
| 18 | 0.44 | 0.25 | 0.44 | 0.15 | 0.60 | 0.25 | 0.60 | 0.24 |
| 19 | 0.29 | 0.35 | 0.29 | 0.15 | 0.35 | 0.18 | 0.35 | 0.27 |
| 20 | 0.29 | 0.26 | 0.29 | 0.12 | 0.30 | 0.20 | 0.30 | 0.24 |
| Mean | 0.44 | 0.37 | 0.44 | 0.29 | 0.58 | 0.40 | 0.58 | 0.40 |

Table A.5: Linear regression of Selfish on VOI under Neutral Frame - within subjects (N1-N2). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise).*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VOI | -0.1528*** | -0.1528*** | -0.1528*** | -0.1528*** |
| | (0.0145) | (0.0140) | (0.0277) | (0.0294) |
| z | – | 0.7685*** | 0.7685*** | 0.7685*** |
| | – | (0.0447) | (0.0828) | (0.0931) |
| *Fixed-effects* | | | | |
| Subject | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Subject | Subject and Payoff |
| Observations | 4,320 | 4,320 | 4,320 | 4,320 |
| Mean Selfish (baseline) | 0.443 | 0.443 | 0.443 | 0.443 |
| # Subjects | 108 | 108 | 108 | 108 |
| # Subjects (baseline) | 108 | 108 | 108 | 108 |
| # Subjects (treated) | 108 | 108 | 108 | 108 |
| *Fit statistics* | | | | |
| $R^2$ | 0.025 | 0.087 | 0.329 | 0.329 |
| Within $R^2$ | — | — | 0.115 | 0.115 |

*Note:* Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered as reported in the table.

Table A.6: Linear regression of Selfish on VOI under Neutral Frame - between subjects (N1-A1). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise).*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VOI | -0.0745*** | -0.0728*** | -0.0728*** | -0.0728* |
| | (0.0145) | (0.0142) | (0.0145) | (0.0403) |
| z | – | 0.6780*** | 0.6780*** | 0.6780*** |
| | – | (0.0450) | (0.0445) | (0.0589) |
| *Socio-dem. controls* | | | | |
| | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Payoff | Subject and Payoff |
| Observations | 4,580 | 4,580 | 4,580 | 4,580 |
| Mean Selfish (baseline) | 0.443 | 0.443 | 0.443 | 0.443 |
| # Subjects | 229 | 229 | 229 | 229 |
| # Subjects (baseline) | 108 | 108 | 108 | 108 |
| # Subjects (treated) | 121 | 121 | 121 | 121 |
| *Fit statistics* | | | | |
| $R^2$ | 0.006 | 0.057 | 0.057 | 0.057 |

*Note:* Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered as reported in the table.

Table A.7: Linear regression of Selfish on VOI under Market Frame - within subjects (M1-M2). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise).*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VOI | -0.1866*** | -0.1866*** | -0.1866*** | -0.1866*** |
| | (0.0147) | (0.0139) | (0.0265) | (0.0350) |
| z | – | 1.0420*** | 1.0420*** | 1.0420*** |
| | – | (0.0442) | (0.0913) | (0.1156) |
| *Fixed-effects* | | | | |
| Subject | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Subject | Subject and Payoff |
| Observations | 4,480 | 4,480 | 4,480 | 4,480 |
| Mean Sell (baseline) | 0.584 | 0.584 | 0.584 | 0.584 |
| # Subjects | 112 | 112 | 112 | 112 |
| # Subjects (baseline) | 112 | 112 | 112 | 112 |
| # Subjects (treated) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| $R^2$ | 0.035 | 0.141 | 0.367 | 0.367 |
| Within $R^2$ | — | — | 0.183 | 0.183 |

*Note:* Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered as reported in the table.

Table A.8: Linear regression of Selfish on VOI under Market Frame - between subjects (M1-B1). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise).*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VOI | -0.1839*** | -0.1879*** | -0.1879*** | -0.1879*** |
| | (0.0147) | (0.0140) | (0.0219) | (0.0402) |
| z | – | 0.8602*** | 0.8602*** | 0.8602*** |
| | – | (0.0445) | (0.0591) | (0.0744) |
| *Socio-dem. controls* | | | | |
| | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Payoff | Subject and Payoff |
| Observations | 4,480 | 4,480 | 4,480 | 4,480 |
| Mean Sell (baseline) | 0.584 | 0.584 | 0.584 | 0.584 |
| # Subjects | 224 | 224 | 224 | 224 |
| # Subjects (baseline) | 112 | 112 | 112 | 112 |
| # Subjects (treated) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| $R^2$ | 0.034 | 0.13 | 0.13 | 0.13 |

*Note:* Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered as reported in the table.

Table A.9: Logistic regression of Selfish on VOI. Dependent variable: *Selfish (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | Neutral – Within (N1–N2) | Neutral – Bet. (N1–A1) | Market – Within (M1–M2) | Market – Bet. (M1–B1) |
|---|---|---|---|---|
| VOIYes | -0.1636*** | -0.0708** | -0.2002*** | -0.1860*** |
| | (0.0231) | (0.0318) | (0.0216) | (0.032) |
| *Socio-dem. controls* | | | | |
| | — | Yes | — | Yes |
| *Fixed-effects* | | | | |
| Subject | Yes | — | Yes | — |
| Payoff config. | Yes | Yes | Yes | Yes |
| *Clustering* | | | | |
| | Subject and Payoff | Subject and Payoff | Subject and Payoff | Subject and Payoff |
| Observations | 4,040 | 4,580 | 4,200 | 4,480 |
| Mean Sell (baseline) | 0.443 | 0.443 | 0.584 | 0.584 |
| # Subjects | 108 | 229 | 112 | 224 |
| # Subjects (baseline) | 108 | 108 | 112 | 112 |
| # Subjects (treated) | 108 | 121 | 112 | 112 |
| *Fit statistics* | | | | |
| Log Likelihood | -1961.61 | -2949.16 | -2026.78 | -2766.23 |
| Pseudo $R^2$ | 0.04 | 0.008 | 0.058 | 0.049 |
| BIC | 4,928.01 | 6,117.48 | 5,096.41 | 5,751.06 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.10: Logistic regression of Selfish on VOI under Neutral Frame - within subjects (N1-N2). Dependent variable: *Selfish (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VOI | -0.1528*** | -0.1528*** | -0.1636*** | -0.1636*** |
| | (0.01448) | (0.0140) | (0.0231) | (0.0231) |
| z | – | 0.7198*** | 0.7802*** | 0.7802*** |
| | – | (0.0400) | (0.0576) | (0.0576) |
| *Fixed-effects* | | | | |
| Subject | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Subject | Subject and Payoff |
| Observations | 4,320 | 4,320 | 4,040 | 4,040 |
| Mean Sell (baseline) | 0.443 | 0.443 | 0.443 | 0.443 |
| # Subjects | 108 | 108 | 108 | 108 |
| # Subjects (baseline) | 108 | 108 | 108 | 108 |
| # Subjects (treated) | 108 | 108 | 108 | 108 |
| *Fit statistics* | | | | |
| Log Likelihood | -2,784.34 | -2,647.76 | -1,980.62 | -1,980.62 |
| Pseudo $R^2$ | 0.019 | 0.067 | 0.116 | 0.116 |
| BIC | 5,585.43 | 5,320.62 | 4,816.54 | 4,816.54 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.11: Logistic regression of Selfish on VOI under Neutral Frame - between subjects (N1-A1). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VOI | -0.0745*** | -0.0729*** | -0.0729*** | -0.0729* |
| | (0.0145) | (0.0142) | (0.0142) | (0.0402) |
| z | – | 0.6511*** | 0.6511*** | 0.6511*** |
| | – | (0.0414) | (0.0409) | (0.0528) |
| *Socio-dem. controls* | | | | |
| | — | Yes | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Payoff | Subject and Payoff |
| Observations | 4,580 | 4,580 | 4,580 | 4,580 |
| Mean Sell (baseline) | 0.443 | 0.443 | 0.443 | 0.443 |
| # Subjects | 229 | 229 | 229 | 229 |
| # Subjects (baseline) | 108 | 108 | 108 | 108 |
| # Subjects (treated) | 121 | 121 | 121 | 121 |
| *Fit statistics* | | | | |
| Log Likelihood | -3,076.01 | -2,958.41 | -2,958.41 | -2,958.41 |
| Pseudo $R^2$ | 0.004 | 0.042 | 0.042 | 0.042 |
| BIC | 6,168.88 | 5,984.25 | 5,984.25 | 5,984.25 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Socio-demographic controls are: nationality, gender, couple status, prior experience in experiments, and attendance to a private school.

Table A.12: Logistic regression of Selfish on VOI under Market Frame - within subjects (M1-M2). Dependent variable: *Selfish (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VOI | -0.1866*** | -0.1866*** | -0.2002*** | -0.2002*** |
| | (0.0147) | (0.0138) | (0.0216) | (0.0216) |
| z | – | 1.054*** | 1.105*** | 1.105*** |
| | – | (0.0410) | (0.0600) | (0.0600) |
| *Fixed-effects* | | | | |
| Subject | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Subject | Subject and Payoff |
| Observations | 4,480 | 4,480 | 4,200 | 4,200 |
| Mean Sell (baseline) | 0.584 | 0.584 | 0.584 | 0.584 |
| # Subjects | 112 | 112 | 112 | 112 |
| # Subjects (baseline) | 112 | 112 | 112 | 112 |
| # Subjects (treated) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| Log Likelihood | -3,026.1 | -2,762.9 | -2,054.31 | -2,054.31 |
| Pseudo $R^2$ | 0.025 | 0.11 | 0.181 | 0.181 |
| BIC | 6,069.01 | 5,551.03 | 5,001.31 | 5,001.31 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.13: Logistic regression of Selfish on VOI under Market Frame - between subjects (M1-B1). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VOI | -0.1839*** | -0.1878*** | -0.1878*** | -0.1878*** |
| | (0.0147) | (0.0140) | (0.0209) | (0.0397) |
| z | – | 0.8643*** | 0.8643*** | 0.8643*** |
| | – | (0.042) | (0.0604) | (0.0736) |
| *Socio-dem. controls* | | | | |
| | — | Yes | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Payoff | Subject and Payoff |
| Observations | 4,480 | 4,480 | 4,480 | 4,480 |
| Mean Sell (baseline) | 0.584 | 0.584 | 0.584 | 0.584 |
| # Subjects | 224 | 224 | 224 | 224 |
| # Subjects (baseline) | 112 | 112 | 112 | 112 |
| # Subjects (treated) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| Log Likelihood | -3,028.55 | -2,794.37 | -2,794.37 | -2,794.37 |
| Pseudo $R^2$ | 0.025 | 0.1 | 0.1 | 0.1 |
| BIC | 6,073.92 | 5,656.01 | 5,656.01 | 5,656.01 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.14: Linear regression of Selfish on Market Frame under VOI condition – within subjects (A1–A2). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise).*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Market | 0.0674*** | 0.0674*** | 0.0674** | 0.0674** |
| | (0.0141) | (0.0136) | (0.0265) | (0.0291) |
| z | – | 0.7743*** | 0.7743*** | 0.7743*** |
| | – | (0.0435) | (0.0730) | (0.0810) |
| *Fixed-effects* | | | | |
| Subject | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Subject | Subject and Payoff |
| Observations | 4,840 | 4,840 | 4,840 | 4,840 |
| Mean Selfish (Neutral Frame) | 0.369 | 0.369 | 0.369 | 0.369 |
| # Subjects | 121 | 121 | 121 | 121 |
| # Subjects (Neutral Frame) | 121 | 121 | 121 | 121 |
| # Subjects (Market Frame) | 121 | 121 | 121 | 121 |
| *Fit statistics* | | | | |
| $R^2$ | 0.005 | 0.066 | 0.343 | 0.343 |
| Within $R^2$ | — | — | 0.091 | 0.091 |

*Note:* Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered as reported in the table.

Table A.15: Linear regression of Selfish on Market Frame under VOI condition – within subjects (B1–B2). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise).*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Market | 0.0728*** | 0.0728*** | 0.0728*** | 0.0728** |
| | (0.0143) | (0.0136) | (0.0245) | (0.0295) |
| z | – | 0.9955*** | 0.9955*** | 0.9955*** |
| | – | (0.0433) | (0.0708) | (0.0777) |
| *Fixed-effects* | | | | |
| Subject | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Subject | Subject and Payoff |
| Observations | 4,480 | 4,480 | 4,480 | 4,480 |
| Mean Selfish (Neutral Frame) | 0.328 | 0.328 | 0.328 | 0.328 |
| # Subjects | 112 | 112 | 112 | 112 |
| # Subjects (Neutral Frame) | 112 | 112 | 112 | 112 |
| # Subjects (Market Frame) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| $R^2$ | 0.006 | 0.111 | 0.316 | 0.316 |
| Within $R^2$ | — | — | 0.139 | 0.139 |

*Note:* Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered as reported in the table.

Table A.16: Linear regression of Selfish on Market Frame under non-VOI condition – between subjects (N1–M1). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise).*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Market | 0.1413*** | 0.1425*** | 0.1425*** | 0.1425*** |
| | (0.0149) | (0.0147) | (0.0137) | (0.0402) |
| z | – | 0.8027*** | 0.8027*** | 0.8027*** |
| | – | (0.0459) | (0.0854) | (0.0990) |
| *Socio-dem. controls* | | | | |
| | — | Yes | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Payoff | Subject and Payoff |
| Observations | 4,400 | 4,400 | 4,400 | 4,400 |
| Mean Selfish (Neutral Frame) | 0.443 | 0.443 | 0.443 | 0.443 |
| # Subjects | 220 | 220 | 220 | 220 |
| # Subjects (Neutral Frame) | 108 | 108 | 108 | 108 |
| # Subjects (Market Frame) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| $R^2$ | 0.02 | 0.091 | 0.091 | 0.091 |

*Note:* Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered as reported in the table.

Table A.17: Linear regression of Selfish on Market Frame – between subjects (A1–B1). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise)*

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Market | 0.0319** | 0.0488*** | 0.0488** | 0.0488 |
| | (0.0143) | (0.0139) | (0.0183) | (0.0387) |
| z | – | 0.7354*** | 0.7354*** | 0.7354*** |
| | – | (0.0438) | (0.0411) | (0.0530) |
| *Socio-dem. controls* | | | | |
| | — | Yes | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Payoff | Subject and Payoff |
| Observations | 4,660 | 4,660 | 4,660 | 4,660 |
| Mean Selfish (Neutral Frame) | 0.369 | 0.369 | 0.369 | 0.369 |
| # Subjects | 233 | 233 | 233 | 233 |
| # Subjects (Neutral Frame) | 121 | 121 | 121 | 121 |
| # Subjects (Market Frame) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| $R^2$ | 0.001 | 0.075 | 0.075 | 0.075 |

*Note:* Standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors are clustered as reported in the table.

Table A.18: Logistic regression of Selfish on Market. Dependent variable: *Selfish (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | VOI – Within (A1–A2) | VOI – Within (B1–B2) | non-VOI – Bet. (N1–M1) | VOI – Bet. (A1–B1) |
|---|---|---|---|---|
| FrameMarket | 0.07487*** | 0.07371*** | 0.1423*** | 0.04586 |
| | (0.02351) | (0.0204) | (0.03079) | (0.02808) |
| *Socio-dem. controls* | | | | |
| | — | — | Yes | Yes |
| *Fixed-effects* | | | | |
| Subject | Yes | Yes | — | — |
| Payoff config. | Yes | Yes | Yes | Yes |
| *Clustering* | | | | |
| | Subject and Payoff | Subject and Payoff | Subject and Payoff | Subject and Payoff |
| Observations | 4,440 | 4,280 | 4,400 | 4,660 |
| Mean Sell (Neutral Frame) | 0.369 | 0.328 | 0.443 | 0.369 |
| # Subjects | 121 | 112 | 220 | 233 |
| # Subjects (Neutral Frame) | 121 | 112 | 108 | 121 |
| # Subjects (Market Frame) | 121 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| Log Likelihood | -2,226.26 | -2,039.57 | -2,811.28 | -2909.34 |
| Pseudo $R^2$ | 0.008 | 0.01 | 0.023 | 0.016 |
| BIC | 5,552.71 | 5,141.07 | 5,840.68 | 6,038.3 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.19: Logistic regression of Selfish on Market Frame under VOI condition – within subjects (A1–A2). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Market | 0.06736*** | 0.06736*** | 0.07487*** | 0.07487*** |
| | (0.0141) | (0.0137) | (0.0236) | (0.0236) |
| z | – | 0.739*** | 0.8061*** | 0.8061*** |
| | – | (0.0393) | (0.0529) | (0.0529) |
| *Fixed-effects* | | | | |
| Subject | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Subject | Subject and Payoff |
| Observations | 4,840 | 4,840 | 4,440 | 4,440 |
| Mean Sell (Neutral Frame) | 0.369 | 0.369 | 0.369 | 0.369 |
| # Subjects | 121 | 121 | 121 | 121 |
| # Subjects (Neutral Frame) | 121 | 121 | 121 | 121 |
| # Subjects (Market Frame) | 121 | 121 | 121 | 121 |
| *Fit statistics* | | | | |
| Log Likelihood | -3,250.36 | -3,101.78 | -2,238.93 | -2,238.93 |
| Pseudo $R^2$ | 0.004 | 0.049 | 0.092 | 0.092 |
| BIC | 6,517.69 | 6,229.01 | 5,426.87 | 5,426.87 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.20: Logistic regression of Selfish on Market Frame under VOI condition – within subjects (B1–B2). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Market | 0.0728*** | 0.0728*** | 0.0737*** | 0.0737*** |
| | (0.0143) | (0.0136) | (0.0204) | (0.0204) |
| z | – | 0.9122*** | 0.9652*** | 0.9652*** |
| | – | (0.0363) | (0.0459) | (0.0459) |
| *Fixed-effects* | | | | |
| Subject | — | — | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Subject | Subject and Payoff |
| Observations | 4,480 | 4,480 | 4,280 | 4,280 |
| Mean Sell (Neutral Frame) | 0.328 | 0.328 | 0.328 | 0.328 |
| # Subjects | 112 | 112 | 112 | 112 |
| # Subjects (Neutral Frame) | 112 | 112 | 112 | 112 |
| # Subjects (Market Frame) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| Log Likelihood | -2,924.8 | -2,689.4 | -2,086.36 | -2,086.36 |
| Pseudo $R^2$ | 0.004 | 0.084 | 0.134 | 0.134 |
| BIC | 5,866.42 | 5,404.03 | 5,084.15 | 5,084.15 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.21: Logistic regression of Selfish on Market Frame under non-VOI condition – between subjects (N1–M1). Dependent variable: *Selfish decision (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Market | 0.1413*** | 0.1425*** | 0.1425*** | 0.1425*** |
| | (0.0149) | (0.0147) | (0.0137) | (0.0403) |
| z | – | 0.8174*** | 0.8174*** | 0.8174*** |
| | – | (0.0444) | (0.0752) | (0.0896) |
| *Socio-dem. controls* | | | | |
| | — | Yes | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Payoff | Subject and Payoff |
| Observations | 4,400 | 4,400 | 4,400 | 4,400 |
| Mean Sell (Neutral Frame) | 0.443 | 0.443 | 0.443 | 0.443 |
| # Subjects | 220 | 220 | 220 | 220 |
| # Subjects (Neutral Frame) | 108 | 108 | 108 | 108 |
| # Subjects (Market Frame) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| Log Likelihood | -3,003.76 | -2,837.55 | -2,837.55 | -2,837.55 |
| Pseudo $R^2$ | 0.014 | 0.069 | 0.069 | 0.069 |
| BIC | 6,024.3 | 5,742.22 | 5,742.22 | 5,742.22 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.22: Logistic regression of Selfish on Market Frame under VOI condition – between subjects (A1–B1). Dependent variable: *Selfish (1 if subject chose the selfish action, 0 otherwise)*. Average Marginal Effects.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Market | 0.0319** | 0.0486*** | 0.0486*** | 0.0486 |
| | (0.0143) | (0.0139) | (0.0181) | (0.0386) |
| z | – | 0.6965*** | 0.6965*** | 0.6965*** |
| | – | (0.0395) | (0.0387) | (0.0479) |
| *Socio-dem. controls* | | | | |
| | — | Yes | Yes | Yes |
| *Clustering* | | | | |
| | None | None | Payoff | Subject and Payoff |
| Observations | 4,660 | 4,660 | 4,660 | 4,660 |
| Mean Sell (Neutral Frame) | 0.369 | 0.369 | 0.369 | 0.369 |
| # Subjects | 233 | 233 | 233 | 233 |
| # Subjects (Neutral Frame) | 121 | 121 | 121 | 121 |
| # Subjects (Market Frame) | 112 | 112 | 112 | 112 |
| *Fit statistics* | | | | |
| Log Likelihood | -3,100.8 | -2,927.49 | -2,927.49 | -2,927.49 |
| Pseudo $R^2$ | 0.001 | 0.057 | 0.057 | 0.057 |
| BIC | 6,218.5 | 5,922.55 | 5,922.55 | 5,922.55 |

*Note:* Bootstrapped standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.23: Logistic regression of Selfish on VOI, Market, and their interaction. Dependent variable: *Selfish decision* (1 if subject chose the selfish action, 0 otherwise). Average Marginal Effects.

| | (1) | (2) |
|---|---|---|
| VOI | -0.1425*** | -0.1425*** |
| | (0.0144) | (0.0144) |
| Market Frame | 0.0954*** | 0.0992*** |
| | (0.0142) | (0.0145) |
| VOI × Market Frame | -0.0235 | -0.0216 |
| | (0.1200) | (0.1219) |
| *Socio-dem. controls* | No | Yes |
| *Fixed-effects* | Payoff config. | Payoff config. |
| Observations | 18,120 | 18,120 |
| *Fit statistics* | | |
| Log Likelihood | -11,199.94 | -11,146.01 |
| Pseudo $R^2$ | 0.085 | 0.089 |
| BIC | 22,625.38 | 22,566.54 |

*Note:* Bootstrapped standard errors in parentheses. *** $p < 0.01$. Socio-demographic controls are: nationality, gender, couple status, prior experience in experiments, and attendance to a private school.

Table A.24: Estimated preferences for the representative agent in Neutral and Market frames — Full sample

| | Neutral | Market | $H_0$: Neutral = Market[+] |
|---|---|---|---|
| $\beta$: Aheadness aversion | 0.230*** | 0.106*** | 0.003 |
| | (0.034) | (0.023) | |
| $\kappa$: Degree of morality | 0.223*** | 0.196*** | 0.738 |
| | (0.071) | (0.035) | |
| $\sigma$: Choice sensitivity | 0.024*** | 0.034*** | 0.331 |
| | (0.007) | (0.007) | |
| LR test ($H_0$: $\beta = \kappa = 0$) | $p < 0.001$ | $p < 0.001$ | |
| Pseudo-$R^2$ | 0.185 | 0.191 | |
| BIC (Full model) | 5,383.59 | 5,677.71 | |
| BIC (Null: $\beta = \kappa = 0$) | 6,407.34 | 6,784.48 | |
| Number of observations | 4,320 | 4,480 | |
| Number of subjects | 108 | 112 | |
| Log likelihood (Full) | −2,665.80 | −2,845.86 | |

Notes: Clustered SE's (subject level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[+] $p$-value of two-sided $z$-test for equality of coefficients across frames.

Table A.25: Estimated preferences for the representative agent in Neutral and Market frames — Core 2 sample

| | Neutral | Market | $H_0$: Neutral = Market[+] |
|---|---|---|---|
| $\beta$: Aheadness aversion | 0.147*** | 0.089*** | 0.159 |
| | (0.035) | (0.022) | |
| $\kappa$: Degree of morality | 0.441*** | 0.255*** | 0.120 |
| | (0.112) | (0.041) | |
| $\sigma$: Choice sensitivity | 0.025*** | 0.045*** | 0.254 |
| | (0.011) | (0.013) | |
| LR test ($H_0$: $\beta = \kappa = 0$) | $p < 0.001$ | $p < 0.001$ | |
| Pseudo-$R^2$ | 0.213 | 0.227 | |
| BIC (Full model) | 3,211.69 | 3,681.09 | |
| BIC (Null: $\beta = \kappa = 0$) | 3,873.50 | 4,287.22 | |
| Number of observations | 2,720 | 3,160 | |
| Number of subjects | 68 | 79 | |
| Log likelihood (Full) | −1,576.85 | −1,846.44 | |

Notes: Clustered SE's (subject level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[+] $p$-value of two-sided $z$-test for equality of coefficients across frames.

Table A.26: Two-type Finite Mixture Model estimates — Full sample

| | Neutral | | Market | |
|---|---|---|---|---|
| | Type 1 | Type 2 | Type 1 | Type 2 |
| $\beta$: Aheadness aversion | 0.373*** | -0.057 | 0.188*** | -0.104 |
| | (0.086) | (0.159) | (0.012) | (0.072) |
| $\kappa$: Degree of morality | 0.125** | 0.309*** | 0.229*** | 0.136** |
| | (0.064) | (0.054) | (0.020) | (0.040) |
| $\sigma$: Choice sensitivity | 0.043*** | 0.026*** | 0.055*** | 0.037*** |
| | (0.009) | (0.003) | (0.011) | (0.011) |
| Share | 0.553*** | 0.447*** | 0.643*** | 0.357*** |
| | (0.134) | (0.134) | (0.026) | (0.026) |
| Number of Observations | 4,320 | | 4,480 | |
| Number of subjects | 108 | | 112 | |
| Log likelihood | -2,297.761 | | -2,403.156 | |

Notes: Clustered SE's (subject level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.27: Two-type Finite Mixture Model estimates, Core 2 sample

| | Neutral | | Market | |
|---|---|---|---|---|
| | Type 1 | Type 2 | Type 1 | Type 2 |
| $\beta$: Aheadness aversion | 0.287*** | -0.145 | 0.180*** | -0.305 |
| | *(0.024)* | *(0.194)* | *(0.012)* | *(0.283)* |
| $\kappa$: Degree of morality | 0.344*** | 0.595*** | 0.176*** | 0.597** |
| | *(0.085)* | *(0.146)* | *(0.020)* | *(0.297)* |
| $\sigma$: Choice sensitivity | 0.047*** | 0.023*** | 0.073*** | 0.029* |
| | *(0.018)* | *(0.006)* | *(0.012)* | *(0.015)* |
| Share | 0.522*** | 0.478*** | 0.623*** | 0.377*** |
| | *(0.084)* | *(0.084)* | *(0.028)* | *(0.028)* |
| Number of Observations | 2,720 | | 3,160 | |
| Number of subjects | 68 | | 79 | |
| Log likelihood | -1,384.981 | | -1,617.195 | |

Notes: Clustered SE's (subject level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

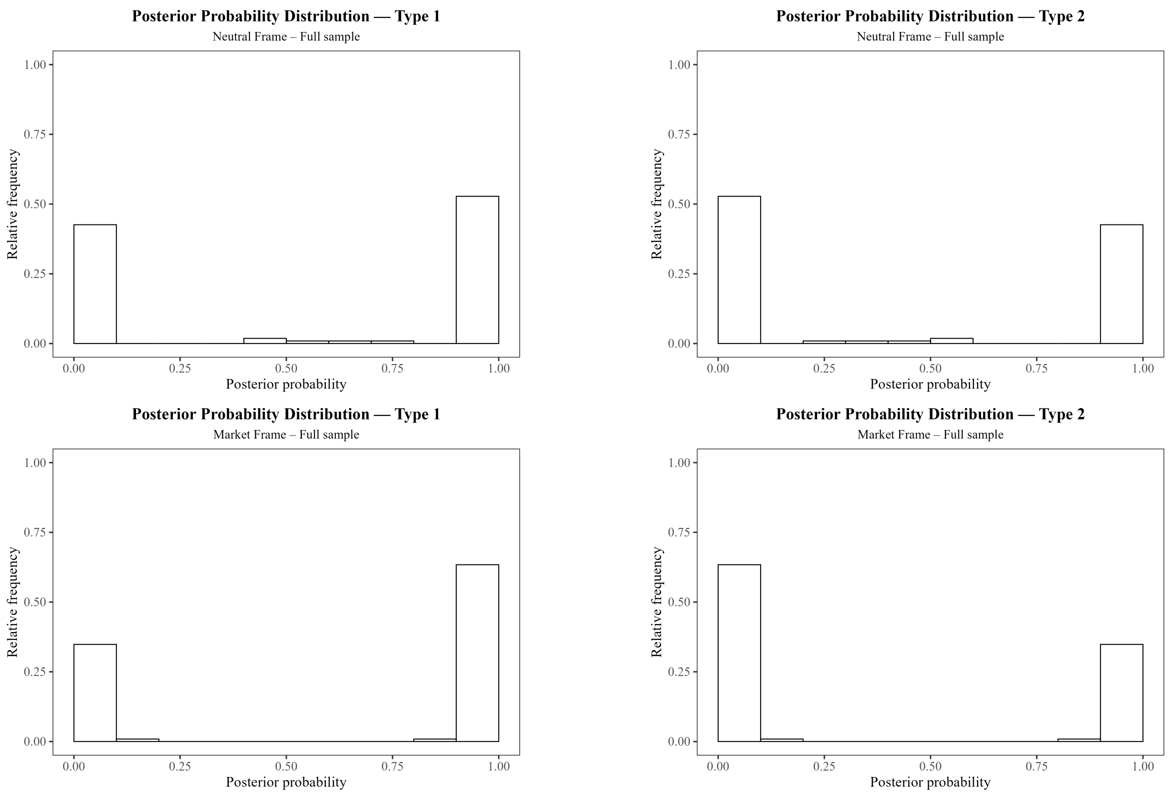
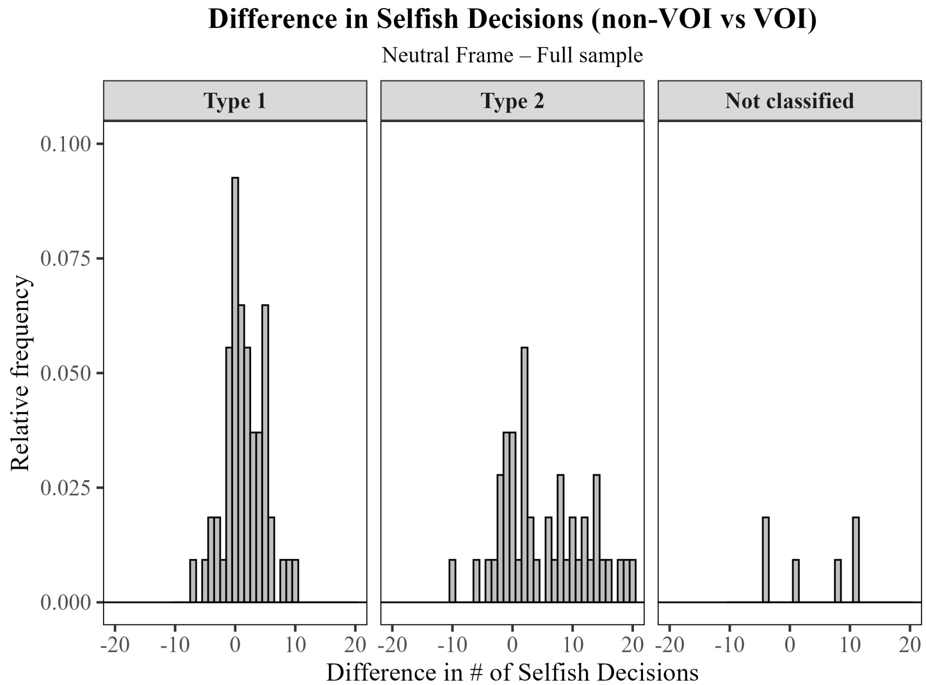
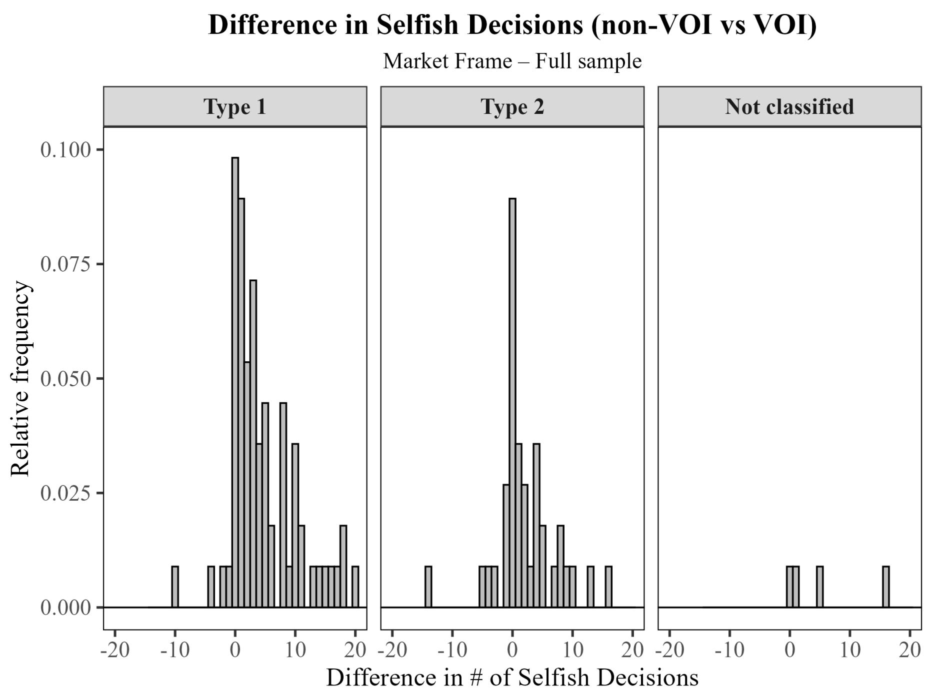



Figure A.4: Distribution of posterior probabilities of individual type-membership in Neutral (first row) and Market frames (second row). Full sample estimates.




(a) Neutral frame




(b) Market frame

Figure A.5: Difference in selfish actions non-VOI versus VOI. Full sample. Subjects classified as Type 1 if $\tau_1 > 0.95$ or Type 2 if $\tau_1 < 0.05$

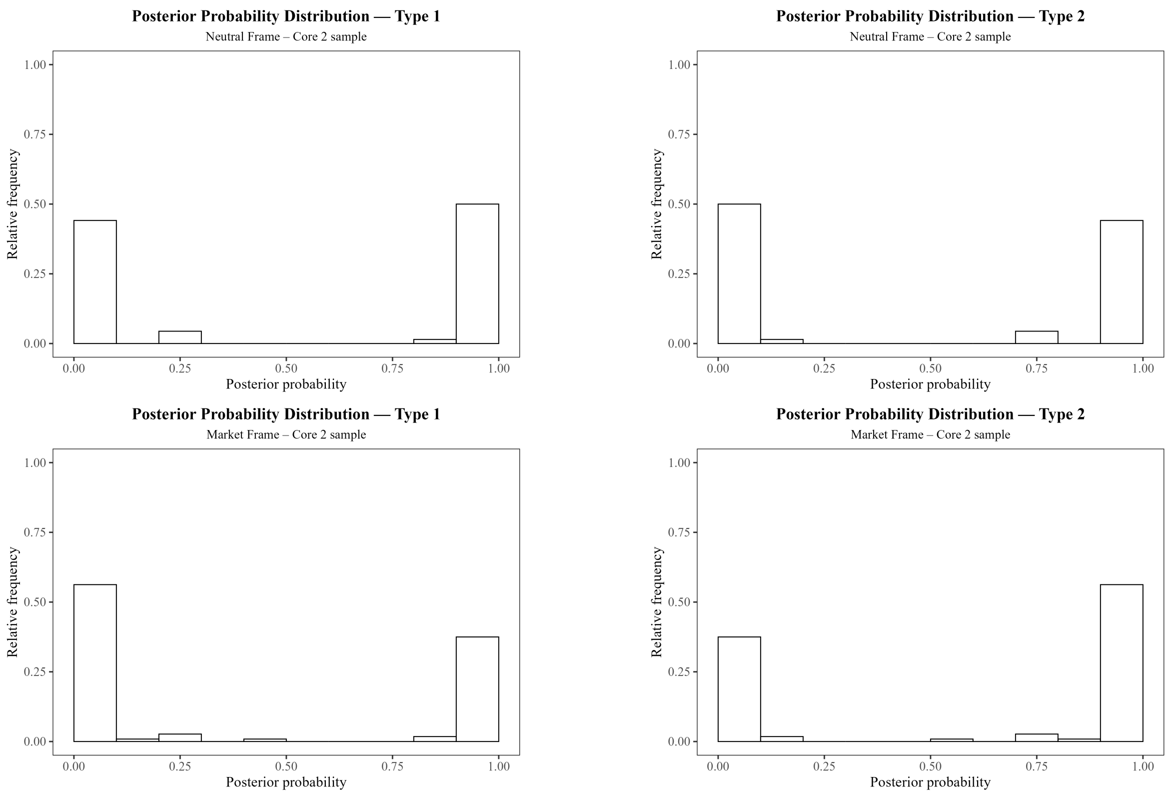
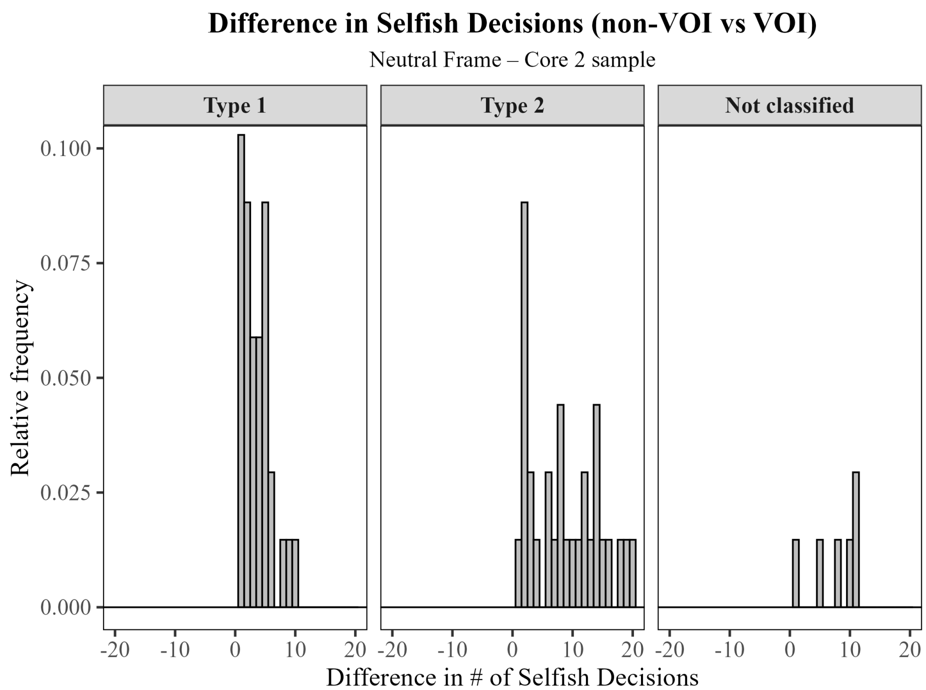
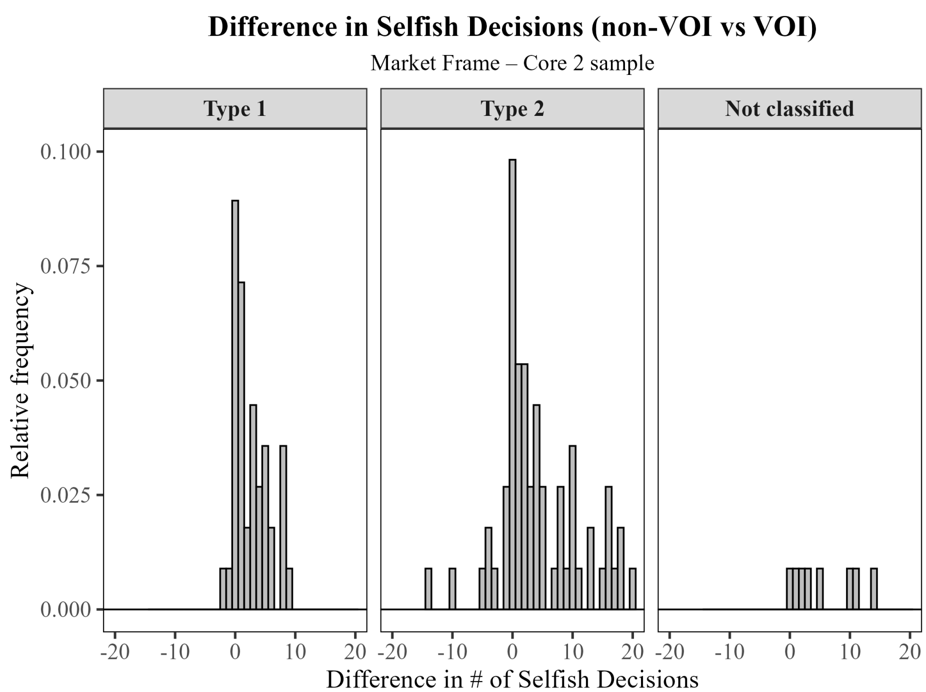



Figure A.6: Distribution of posterior probabilities of individual type-membership in Neutral (first row) and Market frames (second row). Core 2 sample estimates.




(a) Neutral frame




(b) Market frame

Figure A.7: Difference in selfish actions non-VOI versus VOI. Core 2 sample. Subjects classified as Type 1 if $\tau_1 > 0.95$ or Type 2 if $\tau_1 < 0.05$

## Appendix A2 Cross-validation procedure

To evaluate the model's predictive performance in the Core 1 sample, we implemented a three-fold cross-validation using the log-loss as the scoring metric. The partitioning was done at the *subject* level rather than at the individual decision level, so that all observations from a given subject were kept together in the same fold. This ensures that parameter estimation and validation are performed on independent sets of subjects, avoiding cross-contamination of information across folds.

For each fold $k = 1, 2, 3$, the model was estimated on the training set $\mathscr{D}_{-k}$ (the union of the other two folds) and evaluated on the hold-out test set $\mathscr{D}_k$. For each observation $i \in \mathscr{D}_k$, the model generates a predicted probability $\hat{p}_i$ of choosing the selfish option. The log-loss for that fold is computed as:

$$\text{LogLoss}_k = -\frac{1}{|\mathscr{D}_k|} \sum_{i \in \mathscr{D}_k} \left[ y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i) \right], \tag{28}$$

where $y_i = 1$ if the selfish option was chosen and 0 otherwise. The mean log-loss across folds,

$$\overline{\text{LogLoss}} = \frac{1}{3} \sum_{k=1}^{3} \text{LogLoss}_k, \tag{29}$$

provides a measure of the model's expected predictive error on new data. Lower values indicate better predictive performance. For comparison, we computed the same metric both for the full model and for the null (restricted) model containing only the choice sensitivity parameter $\sigma$, and report both below. The full model achieved a consistently lower log-loss across all folds, indicating meaningful out-of-sample predictive power.

Table A.28: Three-fold cross-validation results (log-loss)

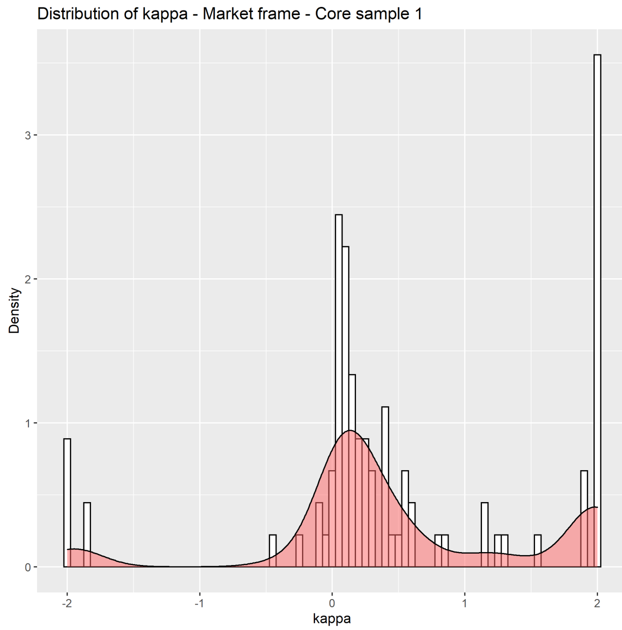
| | **Fold 1** | **Fold 2** | **Fold 3** | **Mean** |
|---|---|---|---|---|
| **Neutral frame** | | | | |
| Full model | 0.616 | 0.595 | 0.662 | 0.624 |
| Null model | 0.693 | 0.693 | 0.693 | 0.693 |
| Δ (Null – Full) | 0.0768 | 0.0980 | 0.0314 | 0.0687 |
| **Market frame** | | | | |
| Full model | 0.665 | 0.611 | 0.604 | 0.627 |
| Null model | 0.693 | 0.693 | 0.693 | 0.693 |
| Δ (Null – Full) | 0.0279 | 0.0819 | 0.0892 | 0.0663 |

*Note:* Folds were defined at the subject level to ensure independence between training and test samples. Lower log-loss values indicate better predictive performance. Δ measures the improvement of the full model relative to the restricted (null) specification.

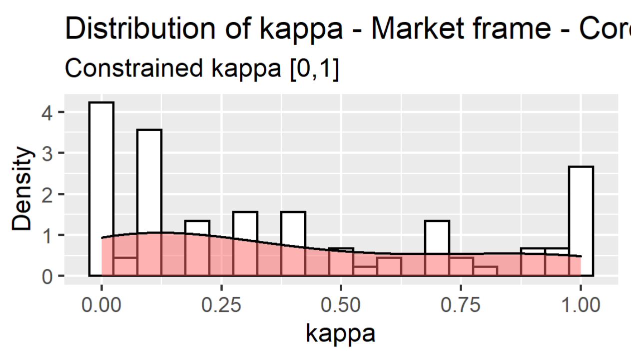
# Appendix A3 Preference parameter estimation at the individual level

Figures A.8 and A.9 show the distributions of the estimates of $\kappa$ at the individual level, for the Market and the Neutral frames, respectively.

These figures show that for a large number of subjects the estimates are negative and/or exceed 2 in absolute value (this is especially true in the Neutral frame). This suggests that our experimental design is not ideal to structurally estimate individual preference parameters. To see why, consider Figure A.10, which shows the threshold values for $\beta$ and $\kappa$ (recall (6) and (7)) for three of the payoffs used in the experiment: those labeled 1, 4, and 20 in Table 1. As indicated by the numbers in the blue balls, the solid lines correspond to payoff 1, the dashed ones to payoff 4, and the dotted ones to payoff 20. We discuss the challenge of identifying $(\beta_i, \kappa_i)$ for several different behavioral scenarios (a discussion which clearly generalizes to subjects taking decisions in the 20 payoff configurations of the experiment). For this discussion, recall that, for any given payoff, a subject $i$: (a) does not sell under either non-VOI or VOI if $(\beta_i, \kappa_i)$ is to the right of the vertical line; (b) sells under both non-VOI or



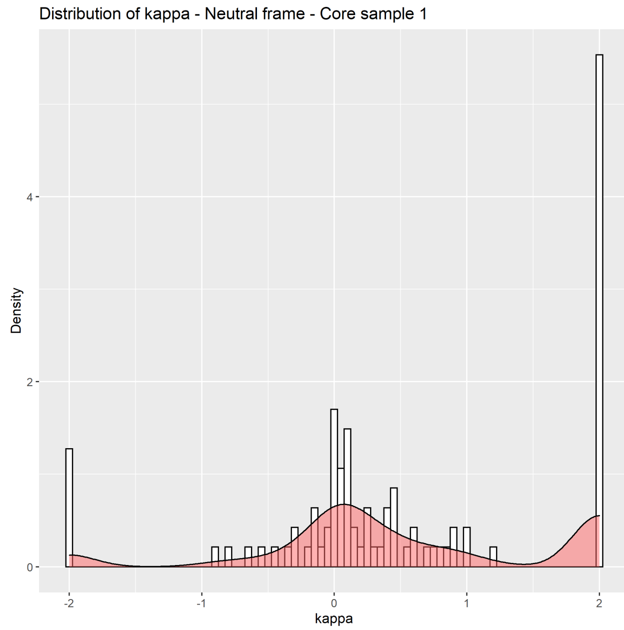

(a)



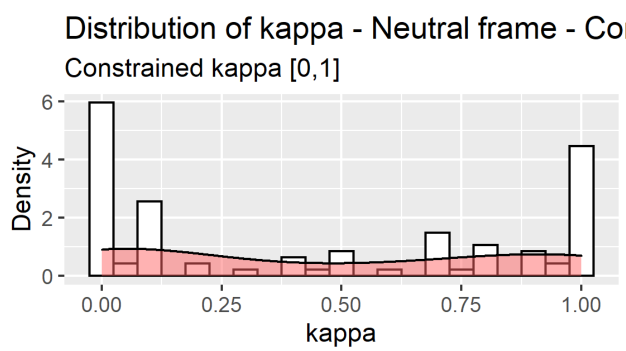

(b)

Figure A.8: Distributions of individual estimates of $\kappa$ for the 100 subjects for whom the decision was different under VOI and non-VOI for at least one of the 20 payoffs, in the Market frame (treatment M). In panel (a), no restriction was imposed on the estimates. In panel (b) the estimates were restricted to the interval $[0,1]$. In panel (a) the estimates whose value is above 2 in absolute value are lumped together at -2 and 2.

VOI if $(\beta_i, \kappa_i)$ is below the downward-sloping curve; and (c) makes a switch (i.e., sells under non-VOI but not under VOI) if $(\beta_i, \kappa_i)$ is between the vertical and the downward-sloping curve (recall Figure 3).




(a)




(b)

Figure A.9: Distributions of individual estimates of $\kappa$ for the 98 subjects for whom the decision was different under VOI and non-VOI for at least one of the 20 payoffs, in the neutral frame (treatment N). In panel (a), no restriction was imposed on the estimates. In panel (b) the estimates were restricted to the interval $[0,1]$. In panel (a) the estimates whose value is above 2 in absolute value are lumped together at -2 and 2.

The first issue is the lack of bounds on $\beta$ and/or $\kappa$ inherent in individual estimations. Consider a subject who switches under payoff 1 and never sells under payoffs 4 and 20. Such behavior is consistent with any $(\beta,\kappa)$ in zone I (where the red ball labeled I appears).

While the value of $\beta$ is thus bounded below by 0.33 and above by 0.6, the value of $\kappa$ is bounded below only. In a similar vein, a subject who switches under payoff 20 and sells under payoffs 1 and 4 is consistent with any $(\beta, \kappa)$ in zone II: for this subject, the value of $\beta$ is bounded above only, while the value of $\kappa$ is bounded below by the downward-sloping dashed and dotted lines. As a third example, a subject who switches under the three payoffs is consistent with any $(\beta, \kappa)$ in zone III. Again, identification of this subject's $(\beta, \kappa)$ is clearly impossible.

The second issue arises for subjects whose behavior is compatible with disjoint sets of values of $(\beta, \kappa)$. For example, consider a subject who switches under payoff 1 and 20, but not under payoff 4. Such a subject could have a $(\beta, \kappa)$ in zone I (consistent with the switch under payoff 1), in which case (s)he would have made a large mistake by switching under payoff 20; or (s)he could have a $(\beta, \kappa)$ in zone II (consistent with the switch under payoff 20), in which case (s)he would have made a large mistake by switching under payoff 1.

By contrast, estimation of $(\beta, \kappa)$ at the aggregate level is possible. To see why, consider again Figure A.10. Suppose that 20% of the subjects switched under all the payoffs (zone III), 25% switched under payoffs 1 and 4 and did not sell under payoff 20 (zone IV), 35% switched under payoff 1 and did not sell under payoffs 4 and 20 (zone I), while the remaining 20% switched only under payoff 20 and sold under payoffs 1 and 4 (zone II). If these decisions are interpreted as emanating from one single individual who sometimes makes mistakes, the estimated $(\beta, \kappa)$ of this hypothetical individual would likely be in zone V, or in zone IV close to zone V. The between-subject heterogeneity in behavior would thus help put boundaries on the estimates of $\beta$ and $\kappa$.

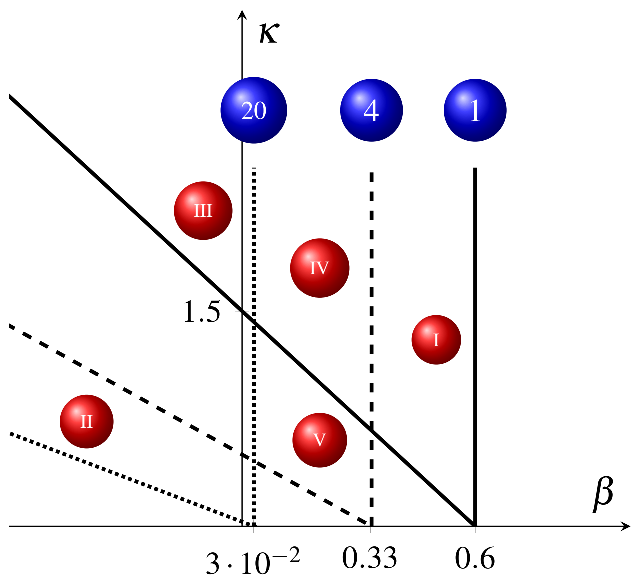



Figure A.10: The numbers in the blue balls indicate the payoff to which the V-shaped lines correspond (see Table 1). The numbers in the red balls indicate zones with different switching behaviors between non-VOI and VOI.

# Appendix A4 Experimental instructions

**Screen 1 [common to all the treatments]:**

Welcome to this experiment!

Please read the following instructions carefully and from now on, do not communicate with any of the other participants. If you have any questions during the experiment, please raise your hand and wait until we come to you to answer your questions in private.

All the participants here today will be asked to take decisions. These decisions will generate points. At the end of the experiment points will be converted into money. Every point is worth 0.025 euros. Each participant also receives 4 euros for attending until the end and answering all questions.

Your decisions during the experiment are anonymous. They will not be linked to your name in any way. Other participants can never trace your decisions back to you. Moreover, the amount of money you receive at the end will be handed over to you in an opaque envelope, and no other participant will see what is inside the envelope.

During the experiment, your cell phone should be switched off and out of reach. And remember not to talk with the other participants. We would need to exclude you from the experiment (and the payment) if you breach these rules.

The experiment consists of two parts, followed by a short questionnaire. At the beginning of each part, you will receive new instructions. Your decisions made in one part will never affect outcomes in the other part, so you can treat both parts as independent.

## A4.1 Remaining screens for the Market treatment

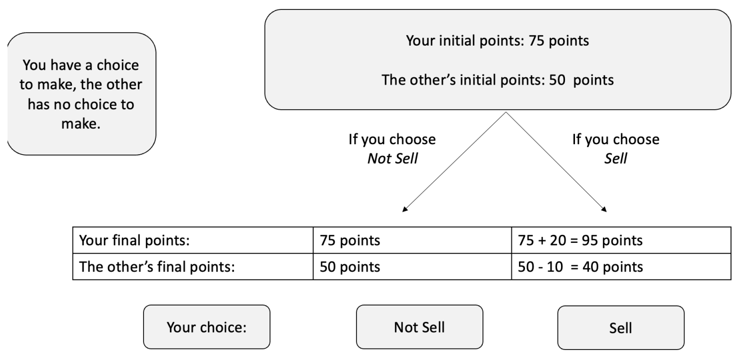
**Screen 2:**

**Part I**

In this part you will be asked to make choices in 20 different decision situations, which represent an interaction between a Seller and a Buyer.

In each decision situation you will be paired with one of the other participants here today, each time with a different participant.

In each decision situation you are the Seller and the person you are paired with is the Buyer. You will be asked to choose one of two options, Sell and Not Sell, while the Buyer has no choice to make. Think of this as representing a situation in which the Buyer is willing to pay the price for a good that you own. Here is an example of a decision screen:




The number of points that you get if you choose Not Sell can be interpreted as the value you attach to owning the good. The number of points that you get if you choose Sell can be interpreted as the amount of money you have if you Sell the good.

The number of points that the Buyer gets if you choose Not Sell can be interpreted as the amount of money he/she has initially. The number of points that the Buyer gets if you choose Sell can be interpreted as the value he/she attaches to owning the good.

As you can see on the decision screen above, the Buyer would be better off if you chose Not Sell, while you are better off if you Sell. Think of this as representing a situation in which the good that you sell has a defect which makes the Buyer enjoy owning the good less than you do. At the end of the experiment one of these 20 decision situations will be randomly drawn. Your decision in this situation (and only this situation) will have an effect on the number of points you and the person you were paired with will get. This other participant will make no decision that affects your number of points, however.

Remember that:

- all your decisions are anonymous, and no participant will ever learn with whom he/she was paired in any decision situation;
- all the decision situations are equally important, in the sense that they are all equally likely to count towards the amount of money you and the person you were paired with will receive at the end;
- each point will be converted to 0.025 euros at the end of the experiment.

**Screen 3 [comprehension quiz]:**

1. Among the decision situations I will be presented with in this part:
   - ○ all will be taken into account to calculate my final number of points.
   - ○ one randomly drawn decision will be taken into account to calculate my final number of points.
   - ○ two randomly drawn decisions will be taken into account to calculate my final number of points.

2. In each decision situation:
   - ○ I will be paired with the same participant.
   - ○ I will be paired with a new, randomly determined participant.

3. At the end of the experiment:
   - ○ I will learn with whom I was paired with.
   - ○ I will not learn with whom I was paired with.

4. In this experiment, each point is worth:
   - ○ 0.02 euros.
   - ○ 0.025 euros.
   - ○ 0.03 euros.

**Screens 4-33:** //the first 20 decision situations//

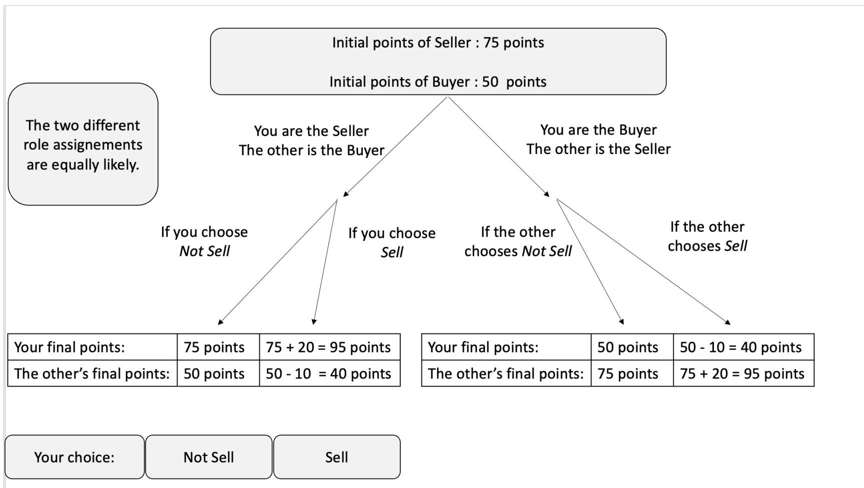
**Screen 34:**

**Part II**

In this part you will be asked to make choices in 20 different decision situations.

In each decision situation you will be paired with one of the other participants here today, each time with a different participant.

In each decision situation there are two roles: the Seller role and the Buyer role. Either you or the person you are paired with will be assigned to the Seller role, while the other will be assigned to the Buyer role. The two different role assignments are equally likely. The person assigned to the Seller role chooses between two options, Sell and Not Sell, while the other person has no choice to make. However, because the role assignment will only be made at the end of the experiment, both you and the person you were paired with will be asked to state what you would do in Role A in each decision situation. Here is an example of a decision screen:




The number of points that the Seller gets from choosing Not Sell can be interpreted as the value that the Seller attaches to owning the good. The number of points that the Seller gets from choosing Sell can be interpreted as the amount of money the Seller has if he/she Sells the good. The number of points that the Buyer gets if the Seller chooses Not Sell can be interpreted as the amount of money the Buyer has initially. The number of points that the Buyer gets if the Seller chooses Sell can be interpreted as the value the Buyer attaches to owning the good.

As you can see on the decision screen above, the Buyer would be better off if the Seller chose Not Sell, while the Seller is better off if he/she chooses Sell. Think of this as representing a situation in which the good that the Seller sells has a defect which makes the Buyer enjoy owning the good less than the Seller does.

At the end of the experiment one of these 20 decision situations will be randomly drawn. The role assignment for this decision situation will be randomly drawn. Recall that the two roles are equally likely. If you are assigned to the Seller role (this happens with probability

1/2), it is your decision in this situation that will have an effect on the number of points you and the person you were paired with will get. If you are assigned to the Buyer role (this happens with probability 1/2), it is the decision of the person you were paired with in this situation that will have an effect on the number of points you and the person you were paired with will get.

Remember that:

- all your decisions are anonymous, and no participant will ever learn with whom he/she was paired in any decision situation;
- all the decision situations are equally important, in the sense that they are all equally likely to count towards the amount of money you and the person you were paired with will receive at the end;
- each point will be converted to 0.025 euros at the end of the experiment.

**Screen 35 [comprehension quiz]:**

1. Among the decision situations I will be presented with in this part:
o all will be taken into account to calculate my final number of points.
o one randomly drawn decision will be taken into account to calculate my final number of points.
o two randomly drawn decisions will be taken into account to calculate my final number of points.

2. In each decision situation:
o there are two roles and I will be paired with the same participant.
o I will be paired with a new, randomly determined participant.

3. In each decision situation:
o I will be paired with the same participant.
o I will be paired with a new, randomly determined participant.

4. At the end of the experiment:
o I will learn who you were paired with.
o I will not learn who you were paired with.

5. In this experiment, each point is worth:
o 0.02 euros.
o 0.025 euros.
o 0.03 euros.

**Screens 36-65: the 20 decision situations for Part II**

## A4.2 Remaining screens for the Neutral treatment

**Screen 2:**

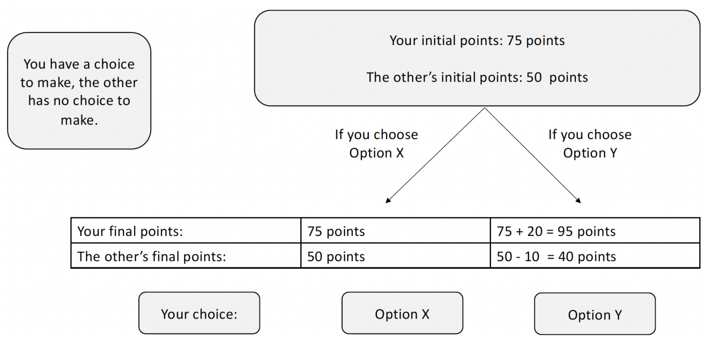
**Part I**

In this part you will be asked to make choices in 20 different decision situations.

In each decision situation you will be paired with one of the other participants here today, each time with a different participant.

In each decision situation you will be asked to choose one of two options, X and Y, while the other participant has no choice to make.

Here is an example of a decision screen:




At the end of the experiment one of these 20 decision situations will be randomly drawn. Your decision in this situation (and only this situation) will have an effect on the number of points you and the person you were paired with will get. This other participant will make no decision that affects your number of points, however.

Remember that:

- all your decisions are anonymous, and no participant will ever learn with whom he/she was paired in any decision situation;
- all the decision situations are equally important, in the sense that they are all equally likely to count towards the amount of money you and the person you were paired with will receive at the end;
- each point will be converted to 0.025 euros at the end of the experiment.

**Screen 3 [comprehension quiz]:**

1. Among the decision situations I will be presented with in this part:
   - o all will be taken into account to calculate my final number of points.
   - o one randomly drawn decision will be taken into account to calculate my final number of points.
   - o two randomly drawn decisions will be taken into account to calculate my final number of points.

2. In each decision situation:
   - o I will be paired with the same participant.
   - o I will be paired with a new, randomly determined participant.

3. At the end of the experiment:
   - o I will learn with whom I was paired with.
   - o I will not learn with whom I was paired with.

4. In this experiment, each point is worth:
   - o 0.02 euros.
   - o 0.025 euros.
   - o 0.03 euros.

**Screens 4-33: the 20 decision situations for Part I**

**Screen 34:**

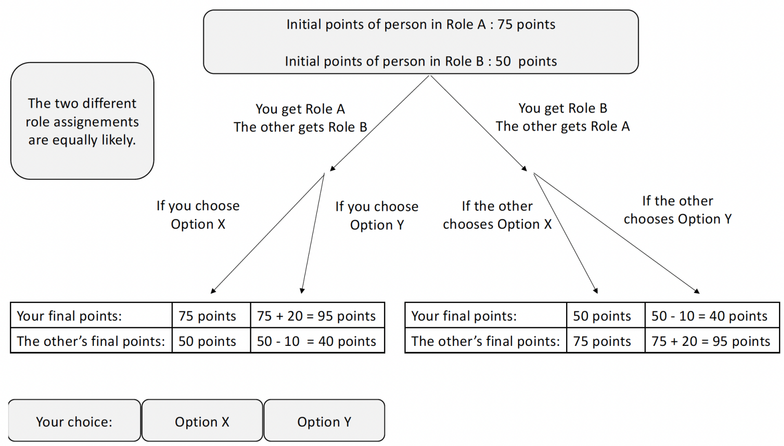
**Part II**

In this part you will be asked to make choices in 20 different decision situations.

In each decision situation you will be paired with one of the other participants here today, each time with a different participant.

In each decision situation there are two roles: Role A and Role B. Either you or the person you are paired with will be assigned to Role A, while the other will be assigned to Role B. The two different role assignments are equally likely. The person assigned to Role A chooses between two options, X and Y, while the other person has no choice to make. However, because the role assignment will only be made at the end of the experiment, both you and the person you were paired with will be asked to state what you would do in Role A in each decision situation.

Here is an example of a decision screen:




At the end of the experiment one of these 20 decision situations will be randomly drawn. The role assignment for this decision situation will be randomly drawn. Recall that the two roles are equally likely. If you are assigned to Role A (this happens with probability 1/2), it is your decision in this situation that will have an effect on the number of points you and the person you were paired with will get. If you are assigned to Role B (this happens with probability 1/2), it is the decision of the person you were paired with in this situation that will have an effect on the number of points you and the person you were paired with will get.

Remember that:

- all your decisions are anonymous, and no participant will ever learn with whom he/she was paired in any decision situation;
- all the decision situations are equally important, in the sense that they are all equally likely to count towards the amount of money you and the person you were paired with will receive at the end;
- each point will be converted to 0.025 euros at the end of the experiment.

**Screen 35 [comprehension quiz]:**

1. Among the decision situations I will be presented with in this part:
- all will be taken into account to calculate my final number of points.
- one randomly drawn decision will be taken into account to calculate my final number of points.
- two randomly drawn decisions will be taken into account to calculate my final number of points.

2. In each decision situation:
- there are two roles and I will be paired with the same participant.
- I will be paired with a new, randomly determined participant.

3. In each decision situation:
- I will be paired with the same participant.
- I will be paired with a new, randomly determined participant.

4. At the end of the experiment:
- I will learn who you were paired with.
- I will not learn who you were paired with.

5. In this experiment, each point is worth:
- 0.02 euros.
- 0.025 euros.
- 0.03 euros.

**Screens 36-65: the 20 decision situations for Part II**

## A4.3 Remaining screens for the two Mixed treatments

Each mixed treatment combines the Market VOI instructions (shown in Part II of subsection "Remaining screens for the Market treatment") and the Neutral VOI instructions (shown in Part II of subsection "Remaining screens for the Neutral treatment").

## A4.4 Example of the two screens shown for one decision