Divided in Morals, United in Selfishness

an Experiment*

By MICHAEL KURSCHILGEN

People use moral wiggle room to behave selfish. But does a narrow wiggle room necessarily produce better social outcomes? When people disagree on normative goals, economic theories of self-image predict that narrowing the moral wiggle room will make choices not only less selfish but also even more heterogeneous, and thus potentially more conflictive. I test this conjecture experimentally in a modified dictator game with and without moral introspection. My results confirm the theoretical predictions, revealing a large disparity of moral valuations of efficiency vs. equality. (C91, D03, D63, A13, Z13)

Morality, Selfishness, Wiggle Room, Identity, Self-Image, Introspection, Other-Regarding Preferences, Experiment

Michael Kurschilgen
Technische Universität München
Arcisstraße 21
80333 München
Phone: +49-(0)89-289-25394
E-mail: m.kurschilgen@tum.de

---

* I am grateful to George Akerlof, Marco Casari, Eric Dickson, Christoph Engel, Werner Güth, Shachar Kariv, Oliver Kirchkamp, Isabel Marcin, Moti Michaeli, Paul Schempp, Shyam Sunder, Roberto Weber, and seminar participants at Bologna, Bonn, Jena, Lausanne and Tilburg, for valuable impulses and comments on earlier versions of the paper, as well as Nicolas Meier and Carlos Kurschilgen for programming assistance. Parts of the paper were written at the Max Planck Institute for Research on Collective Goods Bonn, at the European University Institute, and at the University of Bologna. I thank the Max Planck Society for financial support.

I. Introduction

To understand human behavior in economic interactions, we need to understand the concept of morality, and in particular how morality interacts with selfishness. Recently, there have been some important advances in this respect. Theories of self-image or identity[2] (Akerlof & Kranton, 2000; Bénabou & Tirole, 2011; Falk & Tirole, 2016) rationalize human behavior as the result of trading off selfish, material utility against the utility derived from complying with a certain normative ideal, e.g. a moral standard. Accordingly, a person's choice of action depends on the cost of deviating from one's ideal, moral self, relative to potential material benefits. But moral concerns are malleable. Experimental evidence shows that people's concern for others decreases noticeably as soon as there is moral wiggle room (Dana et al., 2007), i.e. the possibility to act selfishly while maintaining a positive self-image due to readily available moral excuses; for instance by delegating responsibility to other people (Bartling & Fischbacher, 2012; Hamman et al., 2010) or to market forces (Bartling et al., 2015; Falk & Szech, 2013), as well as by distorting beliefs about others' likely behavior (Di Tella et al., 2015).

It thus seems straightforward to assume that narrowing people's moral wiggle room should produce better social outcomes. Just as a large wiggle room produces more selfish behavior, a narrow wiggle room should generate more moral behavior, so the logic. But what exactly do we mean by *moral*? Whilst extant research has focused on the dichotomy of selfish vs. moral behavior, essentially equating moral with unselfish, little attention has been

[2] Please note the critically different meanings of the term identity in the sense of self-image, as in the present paper and the articles cited, in contrast to the notion of group identity in the sense of togetherness, as for instance in Chen and Li (2009).

devoted to the fact that in many situations people have rather diverse, even conflicting conceptions of what is moral.

A typical example are allocation problems. It is a robust empirical regularity that some people sacrifice personal earnings for the sake of higher social efficiency, whereas others are willing to pay for higher equality (Engelmann & Strobel, 2004; Fisman et al., 2007). I show that, according to identity utility (Akerlof & Kranton, 2000), heterogeneity of revealed choices may be a reflection of underlying moral ideals that are even more heterogeneous. As a consequence, narrowing the moral wiggle room (i.e. increasing the relative cost of deviating from one's moral ideal) should make choices not only less selfish but also more heterogeneous. Reducing selfishness by emphasizing morality would thus come along with the side-effect of more potential conflict.

I test this prediction by means of a laboratory experiment. In particular, I elicit participants' other-regarding preferences (Charness & Rabin, 2002) with the help of a modified dictator game (Iriberri & Rey-Biel, 2011) and use the idea of introspection (Krupka & Weber, 2009; Smith, 1790) to induce an increase of the relative cost of deviating from one's moral ideal. My results show that, indeed, choices become both less selfish and more heterogeneous, reflecting even more heterogeneous moral valuations of efficiency vs. equality. The results are robust to providing subjects with information about other people's choices.

The next section describes the theoretical framework and derives testable predictions. Section three presents the experimental design and section four reports the experimental results. The paper ends with a discussion of potential implications.

II. Theoretical Framework

I conceptualize individual allocation decisions following the framework of Charness and Rabin (2002):

$$U_i(a_i) = \begin{cases} (1-\rho)\pi_i(a_i) + \rho\pi_j(a_i) & \text{if } \pi_i \geq \pi_j \\ (1-\sigma)\pi_i(a_i) + \sigma\pi_j(a_i) & \text{if } \pi_i \leq \pi_j \end{cases} \quad (1)$$

An individual i 's utility $U_i(\bullet)$ from taking a certain action a_i depends on the action's payoff consequences for herself and for another player j. Every person is characterized by her concern for others when she is richer (ρ) and when she is poorer (σ). Selfish types, who are only interested in their own material payoff, have $\rho = \sigma = 0$. Efficiency orientation is described by $\rho > 0$ and $\sigma > 0$ whereas inequality aversion is captured by $\rho > 0$ and $\sigma < 0$.[3] See Figure I for an illustration. Other types are theoretically possible but empirically irrelevant.

Fehr and Schmidt (1999) propose a functionally equivalent model but place more restrictions on the range of plausible parameter values thus not allowing for pure efficiency orientation, which empirically has been shown to be highly relevant. The same argument applies to Bolton and Ockenfels (2000), who have a slightly different conceptualization of equality concerns that focuses on relative payoff differences instead of absolute differences. Andreoni and Miller (2002) suggest a more general functional form for other-regarding preferences. Their CES-function not only allows for the payoffs of i an j to be perfect substitutes, as both Fehr and Schmidt (1999) and Charness and Rabin (2002) assume, but also Leontief, Cobb-Douglas and many others. However, the increase in generality comes along with an additional free parameter (i.e. the elasticity of substitution), which makes the empirical

---

[3] Other common labels for those types are „social welfare oriented“ on the one hand as well as „difference averse“ and „inequity averse“ on the other.

elicitation disproportionately more involved without adding important substance for answering this paper's research question.

Empirically, other-regarding preferences have been found to differ particularly along two dimensions (Charness & Rabin, 2002; Fisman et al., 2007): (a) the extent to which a person deviates from material selfishness, i.e. the relative price someone is willing to pay to attain a certain distributional goal, and (b) the nature of the deviation, i.e. in which type of situation someone is willing to give up money. This second empirical regularity makes it an interesting test-bed for identity utility.

Identity utility (Akerlof & Kranton, 2000, 2002, 2005) describes human behavior as the result of trading off material utility against self-esteem. A straightforward formalization reads:

$$U_i(a_i) = \pi_i(a_i) - \gamma_i N_i(d_i) \tag{2}$$

with $d_i = |\tilde{a}_i - a_i|$ representing a given action's deviation from that individual's normative ideal, $N(0) = 0$ and $N'(d_i) > 0$. According to this model, a person's utility increases in her material payoffs $\pi_i(a_i)$ but decreases as her action a_i deviates from her normative ideal $\tilde{a}_i$. The relative importance of normative compliance is determined by $\gamma_i \geq 0$. The smaller (larger) γ_i the closer a person's utility maximizing choice a_i^{**} will be to her selfish optimum a_i^* (to her normative ideal $\tilde{a}_i$). Similar formalizations have also been advanced by Levitt and List (2007), Krupka and Weber (2013), as well as Kimbrough and Vostroknutov (2016). Bénabou and Tirole (2011) propose an alternative conceptualization of identity utility as a game of self-signaling, in which γ, the "strength of the self-esteem motive, increases with the salience of the normative context".

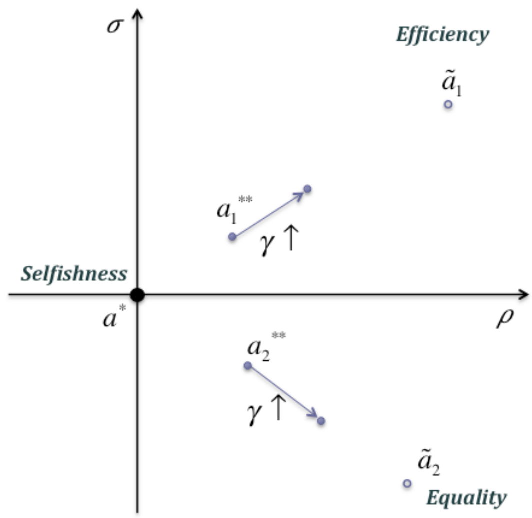

FIGURE I: BEHAVIORAL PREDICTIONS

In the context of other-regarding preferences, identity utility has some interesting implications, illustrated in Figure I. If in a situation with a unique selfish optimum, i.e. $\forall i\left(a_i^* = a^*\right)$, two players' revealed choices (a_1^{**} and a_2^{**}) differ, not only in their distance from the selfish optimum but also in their orientation, this implies that their underlying normative ideals ($\tilde{a}_1$ and $\tilde{a}_2$) will differ even more than their revealed choices. As a consequence, increasing γ should make choices (1) move away from selfish optimization, (2) move closer to people's respective normative ideals, and thus (3) become even more heterogeneous along the dimension of normative dissent. Empirically, it has been shown that virtually all participants display $\rho \geq 0$. In contrast, people have been found to differ substantially along the σ-dimension (Andreoni & Miller, 2002; Charness & Rabin, 2002; Iriberri & Rey-Biel, 2011), revealing differing relative inclinations toward efficiency (large σ) and equality (small σ). The subsequent experiment tests these three predictions.

III. Experimental Design

A. Paradigm

Participants play a modified dictator game (MDG) similar to the one of Iriberri and Rey-Biel (2011). There are two roles: one player is decider i, the other is recipient j. Both players know their respective roles (i.e. no role uncertainty) at the time the game is played. The recipient is passive. The decider is confronted on her computer screen, successively, with four decision panels, each panel consisting of nine decision tasks (see Table I), i.e. a total of 36 tasks. In every task, the decider may choose between an option A and an option B. Option A is always profit maximizing. By choosing option B a decider can either create ($\pi_j^B > \pi_j^A$) additional income for the recipient (panels 1 and 2) or destroy ($\pi_j^B < \pi_j^A$) parts of it (panels 3 and 4). In panels 1 and 3 the decider is richer than the recipient ($\pi_i > \pi_j$) whereas in panels 2 and 4 she is poorer ($\pi_i < \pi_j$), independent of the specific option chosen.

The MDG serves as an instrument to elicit participants' other-regarding preferences in the two-dimensional parameter space of the Charness and Rabin (2002) utility function (see Appendix II for details). The design of the MDG aims for subjects to make deliberate, well-thought choices in the spirit of the Holt and Laury (2002) test for risk attitudes. For that purpose, I deviate from the MDG of Iriberri and Rey-Biel (2011) in two respects: First, I let deciders choose between two options (Option A: selfish, Option B: destroy *or* create) instead of three (Option A: selfish, Option B: create, Option C: destroy). Second, instead of presenting the tasks randomly, I classify them into four panels and sort them within every panel by the relative price of creating/destroying.

TABLE I: DECISION PANELS AND TASKS IN THE MDG

	(1) Ahead – Create				(2) Behind – Create			
	Option A		Option B		Option A		Option B	
Task	π_i^A	π_j^A	π_i^B	π_j^B	π_i^A	π_j^A	π_i^B	π_j^B
1	170	70	160	82	110	120	100	132
2	170	70	160	84	110	120	100	134
3	170	70	160	88	110	120	100	138
4	170	70	160	94	110	120	100	144
5	170	70	160	102	110	120	100	152
6	170	70	160	112	110	120	100	162
7	170	70	160	124	110	120	100	174
8	170	70	160	138	110	120	100	188
9	170	70	160	154	110	120	100	204
	(3) Ahead – Destroy				**(4) Behind – Destroy**			
	Option A		Option B		Option A		Option B	
Task	π_i^A	π_j^A	π_i^B	π_j^B	π_i^A	π_j^A	π_i^B	π_j^B
1	140	130	130	118	90	180	80	168
2	140	130	130	116	90	180	80	166
3	140	130	130	112	90	180	80	162
4	140	130	130	106	90	180	80	156
5	140	130	130	98	90	180	80	148
6	140	130	130	88	90	180	80	138
7	140	130	130	76	90	180	80	126
8	140	130	130	62	90	180	80	112
9	140	130	130	46	90	180	80	96

Note: Player *i* is the decider and player *j* the recipient in the MDG. Each of the 4 decision panels consists of 9 tasks in which the decider chooses between an option A and an option B.

Deciders' choice behavior in the MDG allows categorizing them in a two-dimensional Cartesian type space (see Appendix for details). Players who always choose the profit maximizing Option A are categorized as selfish, which corresponds to the origin of the graph. Every time a decider chooses option B over option A she pays a price of 10 tokens. The further East (West) off the origin a dot is, the more money a decider is willing to give up in order to create (destroy) recipient's income when she is *richer* than the recipient. The further North (South) off the origin a dot is, the more money a decider gives up to create (destroy) recipient's income when she is *poorer* than the recipient. Hence, the North-East (South-East) extreme of the graph represents the maximum amount of efficiency (equality) a decider can choose.

The point x in Figure II, for instance, corresponds to a decider who being richer pays 80 tokens to create income for the recipient whereas being poorer pays 20 tokes to destroy income of the recipient. In Cartesian space, point x thus deviates 80+20=100 tokens from Max Profits, 10+110=120 tokens from Max Efficiency, and 10+70=80 tokens from Max Equality.

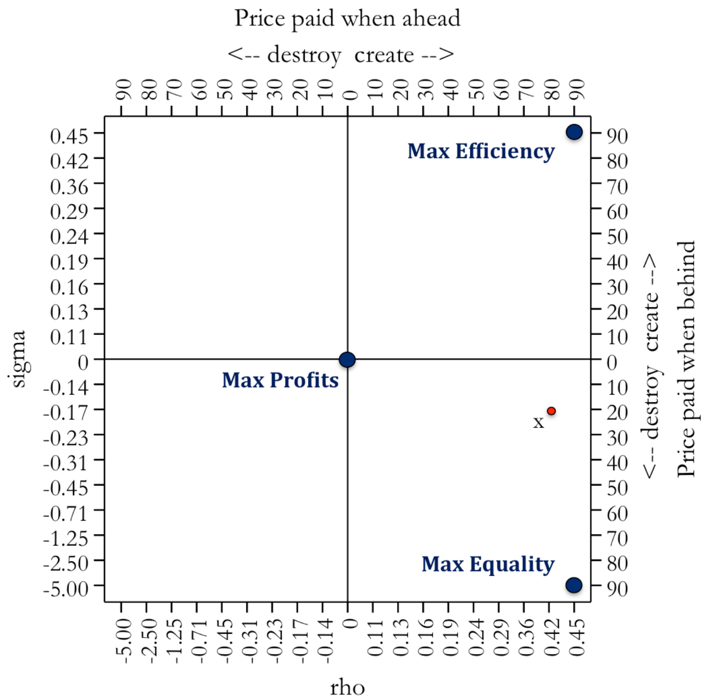

FIGURE II: CARTESIAN TYPE SPACE OF THE MDG

In order to test the three predictions stated in the previous section, I will analyze treatment differences in (1) the distance from Max Profits, (2) the distance from Max Efficiency and Max Equality, and, as measure of heterogeneity, (3) the standard deviation of choices along the σ-dimension.

B. Treatments

The BASE treatment is the MDG described above. I run two sets of treatment comparisons: BASE vs. INTROSPECT, and iBASE vs. iINTROSPECT (see Figure III).

In the INTROSPECT treatment, after reading the instructions but before assigning the roles of decider and recipient (i.e. behind the veil of ignorance), subjects are asked for their moral judgments. Specifically, they have to privately state for each of the 36 tasks they will subsequently be seeing in the MDG: “Which of the two options (A or B) do you find morally right?” The instructions on the computer screen make it clear that the answers to this question are not payoff-relevant and will not be revealed to other participants. Subsequently, subjects are assigned their roles and play the payoff-relevant MDG. When playing the MDG, deciders are reminded on their screens of their own, previously stated, moral judgments.

The idea of the INTROSPECT treatment is to reduce deciders' moral wiggle room by strengthening awareness of their own moral ideal and thus its relative weight vis-à-vis material payoffs (Engel & Kurschilgen, 2015), without imposing any specific normative content, i.e. to increase γ_i without altering $\tilde{a}_i$. This idea can in fact be traced back to Adam Smith's concept of introspection, who called for strengthening one’s moral self by becoming “the impartial spectator of one’s own character and conduct” (Smith, 1790). A similar approach is used by Krupka and Weber (2009), who have subjects deliberate about what others possibly said one should do (“injunctive focus”), as well as Gächter and Riedl (2005), who ask negotiators before a bargaining game to judge the situation from “the vantage point of a neutral arbitrator”. It contrasts for instance with the more intrusive approach of Dal Bó and Dal Bó (2014), who provide participants in a public good game with messages that define moral behavior.

Bicchieri and Xiao (2009) manipulate subjects' normative expectations in a dictator game by telling them what a majority of previous participants *thought should be done* as well as what the majority of participants *did* in previous sessions. My second set of treatments (iBASE and iINTROSPECT) relates to their work in the sense that I also introduce social information albeit with a different purpose. Whereas Bicchieri and Xiao (2009) use *selective* information in order to *change* baseline behavior, I use *representative* information in order to *reinforce* baseline behavior.

In the treatment iBASE subjects are informed about majority behavior in BASE. Specifically, they learn that the experiment has been run before with more than 100 deciders. On the decision screens of the MDG they are shown, for each of the 36 decision tasks, which of the two options (A or B) was chosen by the majority of previous deciders and how large the respective majority was.[4]

In the iINTROSPECT treatment players receive the same information as subjects in iBASE (i.e. they are informed about majority behavior in BASE). Apart from the information, iINTROSPECT is identical to INTROSPECT.

The idea of the information-treatments is to provide subjects with a strong anchor on the type of behavior that subjects typically display in the baseline. Note that subjects in iINTROSPECT receive the social information already before making their moral judgments and that the information reveals the predominance of selfish behavior. Consequently, this should be a more difficult environment for introspection to affect choices and thus a harder test for the theoretical predictions. Figure III provides an overview of the experimental setup.

---

[4] The information is taken from the real choice behavior of deciders in the BASE treatment, see Appendix I.

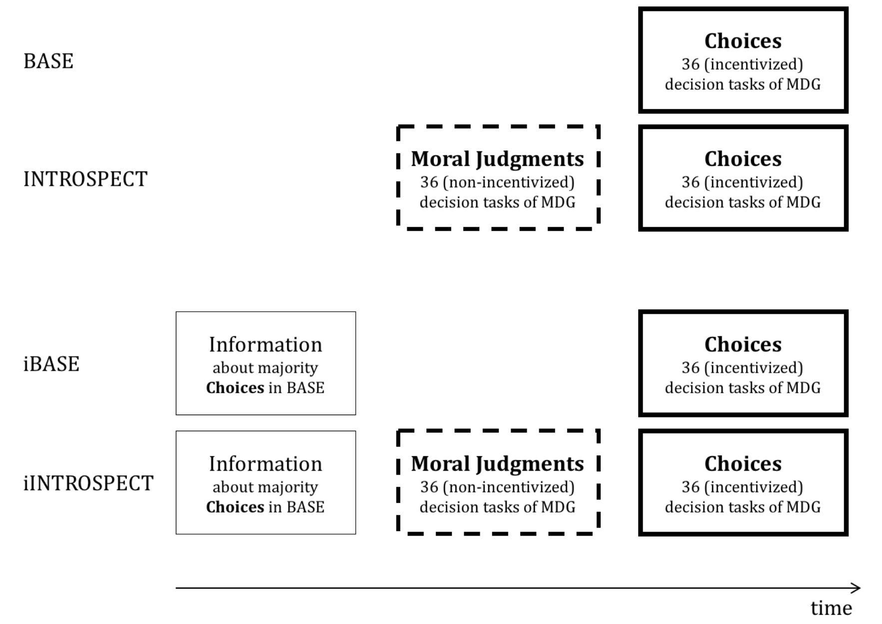

FIGURE III: EXPERIMENTAL SETUP

C. Procedures

The experiment was conducted at the BonnEconLab, Germany. Subjects were recruited via email from a pool of more than 5000 people, using the software ORSEE (Greiner, 2004). In a between-subject design 304 participants (152 deciders) took part in the BASE treatment, 144 (72) in iBASE, 96 (48) in INTROSPECT, in 96 (48) for iINTROSPECT.[5] Participants were mainly University of Bonn undergraduates from a variety of disciplines, 58% were female. Participants were seated in visually completely isolated cubicles. Experimental instructions (see Appendix I) were identical in all treatments. They were handed out in paper to the participants and read aloud by the

[5] The disproportionately large number of participants in BASE is due to the need for collecting a large number of choices in order to provide sufficiently representative information to subjects in iBASE and iINTROSPECT.

experimenter. Participants were then asked to turn their attention to the computer screens in front of them.

At the end of experiment, the computer randomly picked one decision task per panel for payoff. The corresponding token amounts from those four decision tasks were added and converted into Euros (100 tokens = € 1). Participants earned on average € 6 (ca. US$ 8) for approximately 20 minutes of lab-time, which corresponds to about twice the typical student's hourly wage. Immediately after the MDG, subjects answered a non-incentivized questionnaire, covering socio-demographics (age, gender, number of siblings), self-stated risk[6] and trust[7] attitudes as commonly elicited in the German Socioeconomic Panel (SOEP), and the Big-Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism, openness) according to Rammstedt and John (2007). The experiment was computerized in ztree (Fischbacher, 2007).

IV. Results

I first report treatment differences (BASE vs. INTROSPECT and iBASE vs. iINTROSPECT) of the incentivised choice behavior of experimental deciders and subsequently analyze the relationship of individual choices and self-stated moral judgments within INTROSPECT and iINTROSPECT.

---

[6] The risk question read: "Are you, generally speaking, a person willing to take risks or do you rather try to avoid risks?" (0-not at all willing to take risks, ... , 10-very willing to take risks)

[7] The trust questions read: "Please rate the following three statements on a scale from 1 to 4 (1-fully agree, 2-rather agree, 3-rather disagree, 4-fully disagree): (A) Generally, people can be trusted. (B) Nowadays you cannot trust anybody. (C) When dealing with strangers it's better to be careful before trusting them." The composite trust measure is (5-A)+B+C and ranges from 3 (low trust) to 12 (high trust).

A. Choices

Figure IV depicts deciders' choices in BASE and INTROSPECT.[8] In both treatments, the modal behavior is pure selfishness, as illustrated by the large black dot in the origin of both graphs. In BASE, 46 percent of deciders never pick Option B, in INTROSPECT the share of purely selfish deciders decreases to 31 percent. On average, deciders in BASE only deviate 37 tokens from the selfish maximum, nearly doubling the distance to 71 tokens in INTROSPECT (Mann-Whitney ranksum test, two-sided, N=159, p=0.005). The largest part of the change away from selfishness goes into the direction of efficiency, reducing the average distance to Max Efficiency from 152 tokens in BASE to 121 in INTROSPECT (MW, two-sided, N=159, p=0.010). The distance to Max Equality is also reduced from 173 to 163 tokens but fails to reach significance (MW, two-sided, N=159, p=0.160). Heterogeneity along the σ-dimension increases, as predicted, from 32 tokens in BASE to 43 in INTROSPECT (Levene's F-test for equality of variances, two-sided, N=159, p=0.002).

---

[8] I report the results of all deciders whose choice behavior in the MDG was consistent with GARP and thus directly revealed their preferences. The percentage of consistent deciders is virtually identical (Probit regression, p=0.73) in BASE (79 percent) and INTROSPECT (81 percent). Including inconsistent deciders would require additional assumptions about the interpretation of deciders' "errors". For a discussion of GARP-consistency see Appendix II. For a similar procedure see for instance Sutter et al. (2013).

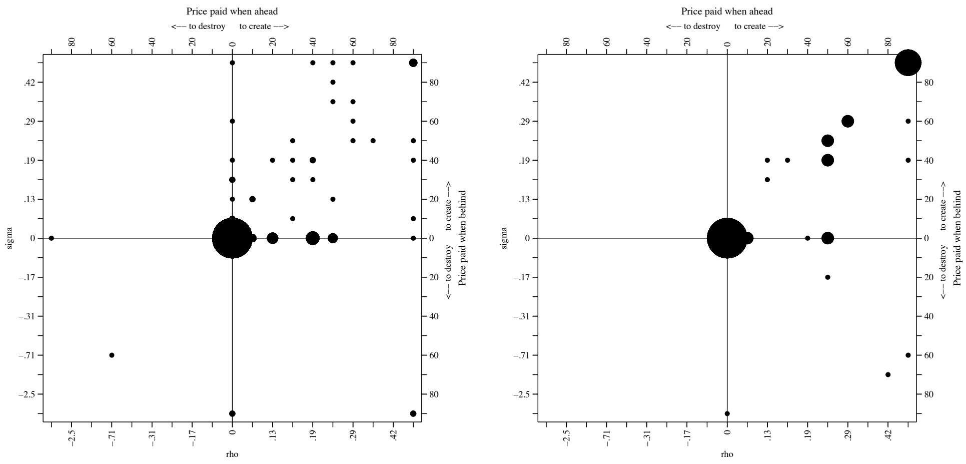

FIGURE IV: CHOICES

Note: The left (right) graph depicts individual choices in BASE (INTROSPECT).

For informed players[9] the effect of introspection is very similar, see Figure V. The deviation from Max Profits increases from 32 tokens in iBASE to 60 in iINTROSPECT (MW, two-sided, N=107, p=0.006). The deviation from Max Efficiency decreases from 150 to 127 (MW, two-sided, N=107, p=0.029). Choices approach Max Equality from 182 to 163 (MW, two-sided, N=107, p=0.007). σ-heterogeneity increases significantly from 28 tokens to 36 (Levene test, two-sided, N=107, p=0.027).[10]

---

[9] Also for the informed players, the share of consistent deciders does not vary significantly (Probit regression, p=0.472) across treatments: 88 in iBASE and 92 in iINTROSPECT.

[10] In an additional control treatment (INTROSPECTi), with 96 subjects (48 deciders), I provide deciders with the social information *after* they made their moral judgments (and *before* they make the incentivized choice). There are no significant differences between iINTROSPECT and INTROSPECTi.

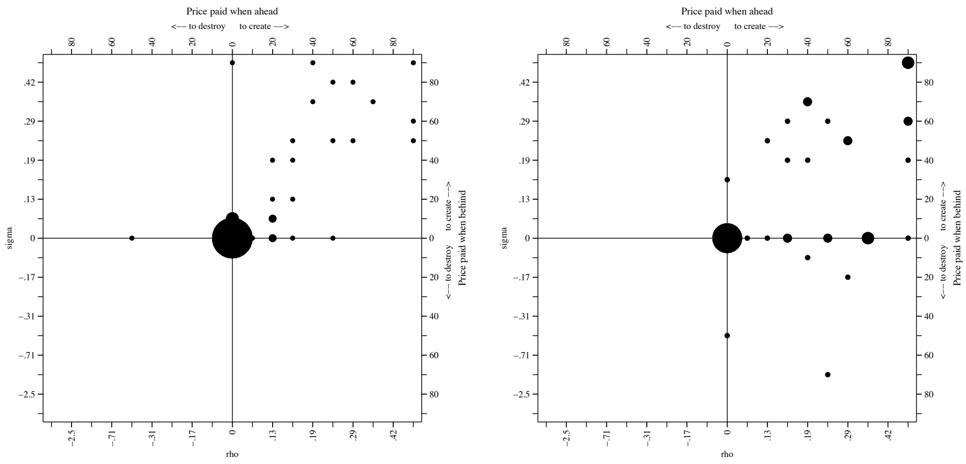

FIGURE V: CHOICES OF INFORMED DECIDERS

Note: The left (right) graph depicts individual choices in iBASE (iINTROSPECT).

All results are confirmed by the parametric analysis of Table II. Narrowing the moral wiggle room through introspection makes choices less selfish and more heterogeneous as some deciders move towards more efficiency whereas others turn towards more equality. Interestingly, the regression results reveal that the general increase in heterogeneity through introspection coincides largely with an increase of gender differences. Column 2b shows that whilst introspection makes men substantially more efficiency-oriented (by 46.27 tokens; p=0.001), this effect is significantly smaller for women (32.82 tokens less; p=0.044) and in sum insignificantly different from zero (13.45 tokens; p=0.156). In sharp contrast, column 3b indicates that introspection makes women significantly more equality-oriented (by 25.06 tokens; p=0.003) whereas the effect for men is virtually zero (-0.466 tokens; p=0.885).

These findings add an interesting new element to the economic literature on gender differences in allocation problems. Non-incentivized survey studies suggest that women care more about equality and tend to be more favorable to redistribution than men (Alesina & La Ferrara, 2005; Corneo & Grüner, 2002; Fong, 2001; Ravallion & Lokshin, 2000). This has been corroborated by several lab experiments (Andreoni & Vesterlund, 2001; Eckel & Grossman, 1998; Krawczyk, 2010; Schildberg-Hörisch, 2010) whilst others have found no difference (Bolton et al., 1998). My results suggest that the observed gender gap may increase with the particular setting's emphasis on the moral dimension of the question.

TABLE II: PARAMETRIC ANALYSIS

	Distance from Max **Selfishness**		Distance from Max **Efficiency**		Distance from Max **Equality**	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
INFO	-3.846	-3.732	-2.051	-2.169	3.927	3.996
	(6.525)	(6.519)	(6.656)	(6.641)	(3.856)	(3.811)
INTROSPECT	29.66	46.41	-28.42	-45.78	-11.01	-0.859
	(6.656)	(10.76)	(6.816)	(11.18)	(4.142)	(5.930)
female	-15.37	-3.217	17.01	4.423	-1.018	6.348
	(6.825)	(7.735)	(6.897)	(7.958)	(4.284)	(5.331)
INTROSPECT*female		-30.28		31.37		-18.35
		(13.61)		(14.00)		(8.350)
age	1.018	1.051	-0.753	-0.788	-0.406	-0.386
	(0.765)	(0.741)	(0.735)	(0.716)	(0.533)	(0.544)
siblings	2.773	2.956	-3.801	-3.991	-0.831	-0.719
	(3.641)	(3.573)	(3.707)	(3.607)	(2.174)	(2.129)
risk	0.962	1.244	0.593	0.301	-1.341	-1.169
	(1.416)	(1.414)	(1.495)	(1.516)	(0.972)	(0.976)
trust	3.596	4.095	-2.805	-3.322	-0.795	-0.492
	(2.118)	(2.089)	(2.156)	(2.128)	(1.157)	(1.168)
extraversion	-5.907	-6.012	4.743	4.852	0.414	0.351
	(1.832)	(1.807)	(1.865)	(1.866)	(1.253)	(1.284)
agreeableness	1.876	2.087	-3.367	-3.586	1.120	1.248
	(2.091)	(2.077)	(2.234)	(2.187)	(1.354)	(1.350)
conscientiousness	-0.843	-0.625	0.570	0.343	0.687	0.819
	(1.914)	(1.911)	(1.916)	(1.908)	(1.028)	(1.028)
neuroticism	-0.704	-0.675	1.466	1.437	-0.998	-0.981
	(1.620)	(1.644)	(1.617)	(1.639)	(1.221)	(1.205)
openness	2.834	2.877	-1.243	-1.287	-0.619	-0.593
	(1.717)	(1.695)	(1.734)	(1.728)	(1.207)	(1.198)
Constant	3.282	-11.80	168.8	184.4	194.8	185.7
	(34.46)	(33.95)	(33.52)	(33.37)	(22.08)	(20.95)
Observations	311	311	311	311	311	311
R-squared	0.143	0.159	0.123	0.140	0.043	0.059

Note: Robust standard errors in parentheses.

B. Moral Judgments

In the preceding section, I have related choices to two stylized normative ideals: maximum attainable efficiency and equality. Figure VI shows what deciders, behind the veil of ignorance, actually state as "morally right behavior". I pool the moral judgments from INTROSPECT and iINTROSPECT since there is no significant difference between informed and uninformed deciders, neither with respect to choices, nor to moral judgments.[11]

According to their self-stated moral judgments, 47 percent self-declare strictly efficiency-oriented (i.e. upper-right quadrant), 18 percent of deciders can be categorized as strictly inequality-averse as they state that is morally right to increase the recipient's income when one is richer but to reduce it when one is poorer (i.e. the lower-right quadrant), and another 18 percent are "effi-equi" types, who believe it is morally right to give when being richer but who would neither pay for destroying nor for creating income for other people when being poorer (i.e. exactly on the horizontal line that separates strict efficiency from strict equality orientation). 14 percent state it is morally right to be selfish (i.e. the origin) whereas 47 percent of deciders state either maximum efficiency, maximum equality or some weighted average thereof as their normative ideal (i.e. the right boundary).

Note the extent of dissimilarity of moral judgments among those players who are furthest away from material selfishness. Players depicted in the upper-right corner believe it is "morally right" to prefer allocation (100, 204) to (110, 120), i.e. to incur 94 additional units of inequality in order to create 74 units of additional wealth. In sharp contrast, players depicted in the lower-right corner believe it is "morally right" to prefer allocation (80, 96) to

[11] For a similar procedure, see for instance Hamman et al. (2010).

(90, 180), i.e. to destroy 94 units of wealth in order to reduce inequality by 74 units.

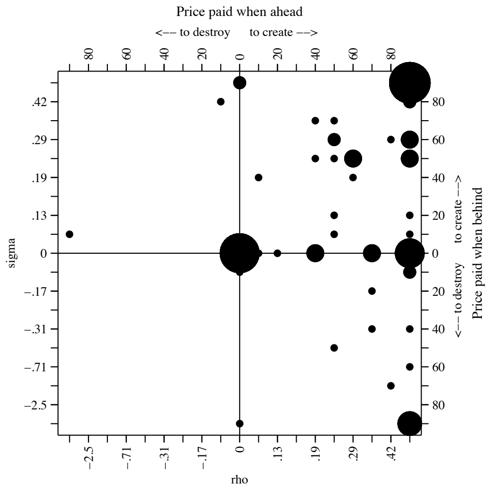

FIGURE VI: MORAL JUDGMENTS

How do these moral judgments translate into actual choice behavior, once the veil of ignorance is lifted? Identity utility would predict them to become more selfish (i.e. reduce the distance to the origin) but to not change their general normative orientation (i.e. the angle). The data thoroughly supports that prediction. Faced with the real incentivized choice, the 47 percent self-declared efficiency-oriented players become significantly more selfish; the mean distance form Max Profits decreases from 137 to 96 tokens (Wilcoxon signrank test, two-sided, N=39, p<0.001) just as the distance from Max Efficiency increases from 43 to 84 tokens (WSR, two-sided, N=39, p<0.001). But these players do not revise their general normative orientation as their distance from Max Equality shrinks only negligibly (and insignificantly) from 177 to 174 tokens (WSR, two-sided, N=39, p=0.241).

The same applies to the 18 percent self-declared equality-oriented players. Also their choices are significantly more selfish than their moral judgments; the distance from Max Profits decreases from 129 to 66 tokens (WSR, two-sided, N=15, p=0.002) just as the distance from Max Equality rises from 51 to 122 tokens (WSR, two-sided, N=15, p=0.002). But they do not become more or less efficiency oriented; their distance to Max Efficiency stays virtually unchanged at 165 tokens (WSR, two-sided, N=15, p=0.871).

The 18 percent effi-equi types display a similar pattern, their choices being significantly more selfish than their moral judgments (WSR, two-sided, N=15, p=0.015) but without becoming more efficiency nor equality oriented. Finally, of the 14 percent self-declared selfish deciders, all but one confirmed their moral judgment with an identical subsequent choice.

Comparing Figure VI with Figure IV and Figure V it is striking how the rather extended type space of moral judgments translates into a much narrower type space of actual choices. This supports the fundamental assumption of identity utility that people have, independent of their individual conception of a moral ideal, selfishness as a common denominator. As players trade off compliance with their individual moral ideal against selfish profit maximization, their actual incentivized choices become much more homogeneous than their moral judgments behind the veil of ignorance. This is evidenced by the significant decrease of σ-heterogeneity from 52 tokens in moral judgments to 39 in choices (Levene test, two-sided, N=83, p=0.0088).

V. Discussion

This paper has analyzed how people react to narrowing moral wiggle room in a situation in which competing normative goals are plausible. In particular, I have looked at decision behavior in allocation problems. Allocation problems are of obvious importance to many economic applications, ranging from wage negotiations, to questions of redistribution and taxation, to issues of political representation. Moreover, allocation problems have the interesting property to feature two plausible normative goals: efficiency and equality. Previous research has established that people disagree on which of the two should be prioritized and by how much. Relying on the concept of identity utility, which views material selfishness as a common denominator of human behavior, I predict that in such a setting narrowing people's moral wiggle room would lead to choices that are not only less selfish but also more heterogeneous, as people move closer to their respective normative "extremes".

My experimental findings yield strong support to those predictions. In particular, I show that incentivized choices become less selfish and more heterogeneous after subjects do moral introspection, and that those choices reflect even less selfish and even more heterogeneous moral judgments of efficiency versus equality. Strikingly, narrowing the moral wiggle room through introspection exacerbates gender differences as men value efficiency stronger whilst women put more weight on equality. All results are robust to providing participants with information about typical choice behavior.

These findings both validate the concept of identity utility and shed some interesting new light on the concept of other-regarding preferences. Whilst it might be convenient to interpret them as outcome preferences, one should be aware that they apparently reflect a tradeoff between material selfishness and other desirable allocative goals, whose relative weight may crucially depend on the "salience of the normative context" (Bénabou & Tirole, 2011).

A significant part of social science, most notably the vast literature on social dilemmas, is concerned with finding ways to restrain people's selfishness and encourage their moral responsibility. Recent research has suggested that narrowing moral wiggle room might be an avenue of improving social outcomes. My findings point at a potentially important caveat. When people differ with respect to their particular moral goals, as participants do already in the simple, non-strategic, two-dimensional setting of this paper, appealing to morality might actually increase the difficulty of finding consensual solutions to social problems. There might sometimes be a social cost to a moral frame. Further research should test this conjecture in situations of strategic interaction, for instance in problems of coordination with conflicting interests. When people disagree on normative goals, they might actually find some common ground in human selfishness. In that sense, future work should attempt to identify situations in which recurring to material selfishness may, ironically, actually serve a social purpose.

Whilst my results underpin that an exogenous variation of a situation's moral implications can have important effects on people's behavior, endogenously managing the moral framing of an issue is a rather usual variable in the action space of political players. Future research should study how those players use moral framing strategically for their particular purposes, how people react to it, and which types of framing equilibria are likely to emerge in political competition.

Appendix 1: Experimental Instructions

A. Paper Instructions for all treatments

General Information

Welcome to our experiment!

If you read the following explanations carefully, you will be able to earn a substantial sum of money, depending on the decisions you make. It is therefore crucial that you read these explanations carefully.

During the experiment there shall be absolutely no communication between participants. Any violation of this rule means you will be excluded from the experiment and from any payments. If you have any questions, please raise your hand. We will then come over to you.

During the experiment we will not calculate in euro, but instead in tokens. Your total income is therefore initially calculated in tokens. The total number of tokens you accumulate in the course of the experiment will be transferred into Euro at the end, at a rate of

100 tokens = 1 Euro.

At the end you will receive from us the **cash** sum, in euro, based on the number of tokens you have earned.

The Experiment

In the experiment, there are two roles: **decider** and **recipient**.

At the beginning of the experiment you will be randomly allotted one of the two roles. One half of the participants will be deciders, the other half will be recipients. During the entire experiment, you will remain in the same role.

On your computer screen you will be shown **4 tables**, one after the other. Every table consists of **9 decision tasks**.

A decision task could for example read as follows:

	Option A	Option B	
Decider (You)	12	10	
Recipient	5	7	Your decision (A or B):

In every decision task the decider has to choose between **Option A** and **Option B**. The two options define how many **tokens** the decider gets and how many the recipient gets.

In this example the decider gets 12 tokens and the recipient 5 tokens if the decider chooses Option A. If the decider chooses Option B, the decider gets 10 tokens and the recipient 7 tokens.

In every decision task the computer will **randomly** match every decider with a different recipient. Thus the decider-recipient pairs change in every decision task.

The decider will never know the identity of the recipient.

The recipient will never know the identity of the decider.

At the end of every table please press the "OK" button on the lower right hand side of your screen. Only after pressing "OK" your decisions are saved and become effective. You will then be shown the next table.

Payoffs

At the end of experiment the computer will **randomly** pick one decision task out of every table. The computer thus picks in total **4 decision tasks**, one from every table. The corresponding token amounts from those 4 decision tasks will be added and changed into Euros.

If you are **decider**, your payoffs only depend on your own choices and on the random draw at the end of the experiment.

If you are **recipient**, your payoffs only depend on the choices of the corresponding decider and the random draw at the end of the experiment.

B. Additional screen in INTROSPECT treatment

Before the computer randomly determines who will be Decider and who will be Recipient, we would like to know your opinion.

We would like to know from you:

Which of the two Options (A or B) do you find morally right?

The answers to these questions will be kept anonymous. No other participant will get to know them at any time. Your answers to these questions are not relevant for your payoffs.

C. Additional screen in iBASE treatment

This Experiment has been run before with more than 100 Deciders.

In the column on the right hand side of your screen you can see how the Deciders in those previous Experiments decided. Specifically, you will be shown which percentage of Deciders chose Option A or Option B in the corresponding Choice Task.

D. Information about choices of previous players

Table A.I displays the information given to players in the iBASE and iINTROSPECT treatments. The information represents the percentage of deciders choosing the majority option in the BASE treatment.

TABLE A.I: SOCIAL INFORMATION

Panel Task	Ahead – Create	Ahead – Destroy	Behind – Create	Behind – Destroy
1	89% chose A	95% chose A	91% chose A	95% chose A
2	89% chose A	95% chose A	92% chose A	91% chose A
3	87% chose A	97% chose A	89% chose A	95% chose A
4	83% chose A	98% chose A	88% chose A	90% chose A
5	76% chose A	99% chose A	84% chose A	90% chose A
6	68% chose A	97% chose A	76% chose A	89% chose A
7	64% chose A	97% chose A	74% chose A	91% chose A
8	58% chose A	97% chose A	68% chose A	90% chose A
9	53% chose A	97% chose A	67% chose A	89% chose A

Appendix 2: Logic of the Modified Dictator Game (MDG)

Player *i* prefers allocation B to allocation A iff $U_i^B \geq U_i^A$. Assuming Charness and Rabin (2002) preferences and $\pi_i \geq \pi_j$ this implies:

$$\pi_i^B - \pi_i^A \geq \rho\left(\left(\pi_i^B - \pi_i^A\right) - \left(\pi_j^B - \pi_j^A\right)\right) \tag{A1}$$

which, for convenience, I rewrite as:

$$\Delta_i \geq \rho\left(\Delta_i - \Delta_j\right) \tag{A2}$$

The same argument applies to $\pi_i \leq \pi_j$ by simply replacing ρ with σ. Assuming $\Delta_i < 0$ (i.e. allocation B is less profitable to player *i* than allocation A) a person's choice reveals her ρ and σ parameters as depicted in Table A.II.

TABLE A.II: PARAMETER SPACE OF THE MDG

	$\pi_i \geq \pi_j$	$\pi_i \leq \pi_j$
$\Delta_i < \Delta_j$	$\rho \geq \frac{\Delta_i}{\Delta_i - \Delta_j} > 0$	$\sigma \geq \frac{\Delta_i}{\Delta_i - \Delta_j} > 0$
$\Delta_i > \Delta_j$	$\rho \leq \frac{\Delta_i}{\Delta_i - \Delta_j} < 0$	$\sigma \leq \frac{\Delta_i}{\Delta_i - \Delta_j} < 0$

Table A.III illustrates how the experimental MDG devotes one decision panel to each of these four situations. In the two panels on the left, the decider's payoff is always higher than the recipient's ($\pi_i \geq \pi_j$) whereas in the two right-hand panels it is the other way around ($\pi_i \leq \pi_j$). In the two upper panels the decider can create income for the recipient ($\Delta_i < 0 < \Delta_j$) whereas in the two lower panels she can reduce the recipient's income ($0 > \Delta_i > \Delta_j$).

TABLE A.III: LOGIC OF DECISION TASKS IN THE MDG

	(1) Ahead – Create					(2) Behind – Create				
	$\pi_i^A = 170$ $\pi_j^A = 70$					$\pi_i^A = 110$ $\pi_j^A = 120$				
Task	π_i^B	π_j^B	Δ_i	Δ_j	$\rho \geq$	π_i^B	π_j^B	Δ_i	Δ_j	$\sigma \geq$
1	160	82	*-10*	*12*	*0.45*	100	132	*-10*	*12*	*0.45*
2	160	84	*-10*	*14*	*0.42*	100	134	*-10*	*14*	*0.42*
3	160	88	*-10*	*18*	*0.36*	100	138	*-10*	*18*	*0.36*
4	160	94	*-10*	*24*	*0.29*	100	144	*-10*	*24*	*0.29*
5	160	102	*-10*	*32*	*0.24*	100	152	*-10*	*32*	*0.24*
6	160	112	*-10*	*42*	*0.19*	100	162	*-10*	*42*	*0.19*
7	160	124	*-10*	*54*	*0.16*	100	174	*-10*	*54*	*0.16*
8	160	138	*-10*	*68*	*0.13*	100	188	*-10*	*68*	*0.13*
9	160	154	*-10*	*84*	*0.11*	100	204	*-10*	*84*	*0.11*
	(3) Ahead – Destroy					**(4) Behind – Destroy**				
	$\pi_i^A = 140$ $\pi_j^A = 130$					$\pi_i^A = 90$ $\pi_j^A = 180$				
Task	π_i^B	π_j^B	Δ_i	Δ_j	$\rho \leq$	π_i^B	π_j^B	Δ_i	Δ_j	$\sigma \leq$
1	130	118	*-10*	*-12*	*-5.00*	80	168	*-10*	*-12*	*-5.00*
2	130	116	*-10*	*-14*	*-2.50*	80	166	*-10*	*-14*	*-2.50*
3	130	112	*-10*	*-18*	*-1.25*	80	162	*-10*	*-18*	*-1.25*
4	130	106	*-10*	*-24*	*-0.71*	80	156	*-10*	*-24*	*-0.71*
5	130	98	*-10*	*-32*	*-0.45*	80	148	*-10*	*-32*	*-0.45*
6	130	88	*-10*	*-42*	*-0.31*	80	138	*-10*	*-42*	*-0.31*
7	130	76	*-10*	*-54*	*-0.23*	80	126	*-10*	*-54*	*-0.23*
8	130	62	*-10*	*-68*	*-0.17*	80	112	*-10*	*-68*	*-0.17*
9	130	46	*-10*	*-84*	*-0.14*	80	96	*-10*	*-84*	*-0.14*

Note: To ensure that stakes are comparable across panels, every panel has approximately (i.e. constrained on only using integers) the same average pie size $\overline{P} = \frac{1}{18}\sum_{t=1}^{9} \pi_{i,t}^A + \pi_{j,t}^A + \pi_{i,t}^B + \pi_{j,t}^B$. Ahead-Create has 254 tokens, Ahead-Destroy 246, Behind-Create 244, and Behind-Destroy 246.

In each panel, there are nine decision tasks. In each task the decider has to choose between option A and option B, specifying two different payoff allocations for the decider and the corresponding recipient. Option A is the same for every task within a given panel. Option B creates or destroys income of the recipient at a cost of 10 tokens. Take for example task 1 of the Ahead-Create panel. If the decider chooses option A she receives 170 tokens and the recipient 70 tokens and if she chooses option B she gets 160 and the recipient 82.

In each panel the relative price of creating/destroying decreases with every task. In task 1, the decider has to give up 10 tokens to create/destroy 12 tokens whereas in task 9 for the same cost the decider creates/destroys 84 tokens. Consequently, choosing option B in task 1 and option A in task 2 of the same panel would violate the General Axiom of Revealed Preferences (GARP). In the MDG, a GARP-consistent decider should have at most one switch from Option A to option B per panel, and no switch from B to A. In addition, consistency requires players not to both create and destroy when they are ahead (or behind). If these consistency requirements are met, the ρ and σ of a given decider are defined by the point at which she switches from option A to option B.

For example, a player who chooses option A in the first 3 tasks of the Ahead-Create panel and option B in the remaining 6 tasks, would have $0.36 > \rho \geq 0.29$. The same player might then for instance choose always option A in the Ahead-Destroy panel and in the Behind-Create panel but then switch to option B in task 7 of the Behind-Destroy panel. This would yield $-0.31 > \sigma \leq -0.23$. The type classification is straightforward: Selfish players will never choose option B since this is costly. Efficiency oriented players will create income for the recipient both when ahead and when behind as long as the relative price of creating is low enough. Equality oriented players will also create when ahead but destroy when behind. Competitive types will destroy recipients' income no matter whether they are ahead or behind.

References

Akerlof, George A., & Kranton, Rachel E. (2000). Economics and Identity. *The Quarterly Journal of Economics, 115*(3), 715-753.

-------. (2002). Identity and Schooling. Some Lessons for the Economics of Education. *Journal of Economic Literature, 40*(4), 1167-1201.

-------. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives, 19*, 9-32.

Alesina, Alberto, & La Ferrara, Eliana. (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics, 89*(5), 897-931.

Andreoni, James, & Miller, John. (2002). Giving According to GARP. An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica, 70*, 737-753.

Andreoni, James, & Vesterlund, Lise. (2001). Which is the Fair Sex? Gender Differences in Altruism. *The Quarterly Journal of Economics, 116*, 293-312.

Bartling, Björn, & Fischbacher, Urs. (2012). Shifting the Blame. On Delegation and Responsibility. *Review of Economic Studies, 79*(1), 67-87.

Bartling, Björn, Weber, Roberto A., & Yao, Lan. (2015). Do Markets Erode Social Responsibility? *The Quarterly Journal of Economics, 130*(1), 219-266.

Bénabou, Roland, & Tirole, Jean. (2011). Identity, Morals, and Taboos: Beliefs as Assets. *The Quarterly Journal of Economics, 126*(2), 805-855.

Bicchieri, Cristina, & Xiao, Erte. (2009). Do the Right Thing: But Only If Others Do So. *Journal of Behavioral Decision Making, 22*, 191-208.

Bolton, Gary E., Katok, Elena, & Zwick, Rami. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory, 27*(2), 269-299.

Bolton, Gary E., & Ockenfels, Axel. (2000). ERC: A Theory of Equity, Reciprocity and Competition. *American Economic Review, 90*, 166-193.

Charness, Gary, & Rabin, Matthew. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics, 117*, 817-869.

Chen, Yan, & Li, Sherry Xin. (2009). Group Identity and Social Preferences. *American Economic Review, 99*, 431-457.

Corneo, Giacomo, & Grüner, Hans Peter. (2002). Individual preferences for political redistribution. *Journal of Public Economics, 83*(1), 83-107.

Dal Bó, Ernesto, & Dal Bó, Pedro. (2014). "Do the right thing:" The effects of moral suasion on cooperation. *Journal of Public Economics, 117*, 28-38.

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

Di Tella, Rafael, Perez-Truglia, Ricardo, Babino, Andres, & Sigman, Mariano. (2015). Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism. *The American Economic Review, 105*(11), 3416-3442.

Eckel, Catherine C., & Grossman, Philip J. (1998). Are Women Less Selfish Than Men?: Evidence from Dictator Experiments. *Economic Journal, 108*, 726-735.

Engel, Christoph, & Kurschilgen, Michael. (2015). The Jurisdiction of the Man within – Introspection, Identity, and Cooperation in a Public Good Experiment. *MPI Collective Goods Preprint, No. 2015/1*.

Engelmann, Dirk, & Strobel, Martin. (2004). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *American Economic Review, 94*, 857-869.

Falk, Armin, & Szech, Nora. (2013). Morals and Markets. *Science, 340*, 707-711.

Falk, Armin, & Tirole, Jean. (2016). Narratives, Imperatives and Moral Reasoning.

Fehr, Ernst, & Schmidt, Klaus M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics, 114*, 817-868.

Fischbacher, Urs. (2007). z-Tree. Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics, 10*, 171-178.

Fisman, Raymond, Kariv, Shachar, & Markovits, Daniel. (2007). Individual Preferences for Giving. *American Economic Review, 97*, 1858-1876.

Fong, Christina M. (2001). Social Preferences, Self-interest, and the Demand for Redistribution. *Journal of Public Economics, 82*(2), 225-246.

Gächter, Simon, & Riedl, Arno. (2005). Moral property rights in bargaining with infeasible claims. *Management Science, 51*(2), 249-263.

Greiner, Ben. (2004). An Online Recruiting System for Economic Experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003* (pp. 79-93). Göttingen.

Hamman, John R, Loewenstein, George, & Weber, Roberto A. (2010). Self-interest through delegation: An additional rationale for the principal-agent relationship. *The American Economic Review*, 1826-1846.

Holt, Charles A., & Laury, Susan K. (2002). Risk Aversion and Incentive Effects. *American Economic Review, 92*, 1644-1655.

Iriberri, Nagore, & Rey-Biel, Pedro. (2011). The Role of Role Uncertainty in Modified Dictator Games. *Experimental Economics, 14*, 160-180.

Kimbrough, Erik O., & Vostroknutov, Alexander. (2016). Norms Make Preferences Social. *Journal of the European Economic Association, 14*(3), 608-638.

Krawczyk, Michał. (2010). A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution. *Journal of Public Economics, 94*(1), 131-141.

Krupka, Erin L., & Weber, Roberto A. (2009). The Focusing and Informational Effects of Norms on Pro-Social Behavior. *Journal of Economic Psychology, 30*, 307-320.

-------. (2013). Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association, 11*(3), 495-524.

Levitt, Steven D., & List, John A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives, 21*, 153-174.

Rammstedt, Beatrice, & John, Oliver P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203-212.

Ravallion, Martin, & Lokshin, Michael. (2000). Who wants to redistribute?: The tunnel effect in 1990s Russia. *Journal of Public Economics, 76*(1), 87-104.

Schildberg-Hörisch, Hannah. (2010). Is the veil of ignorance only a concept about risk? An experiment. *Journal of Public Economics, 94*(11), 1062-1066.

Smith, Adam. (1790). *The Theory of Moral Sentiments, or, An Essay Towards an Analysis of the Principles by which Men Naturally Judge Concerning the Conduct and Character, First of Their Neighbours, and Afterwards of Themselves. To which is Added, a Dissertation on the Origin of Languages* (6 ed.). London: Strahan.

Sutter, Matthias, Kocher, Martin G, Glätzle-Rützler, Daniela, & Trautmann, Stefan T. (2013). Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior. *The American Economic Review, 103*(1), 510-531.