TSE WP working paper series

12-347

Research Group: Behavioral and Experimental Economics

November 2012

“Noblesse Oblige? Moral Identity and Prosocial Behavior in the Face of Selfishness”

Roberta Dessí and Benoît Monin

Toulouse School of Economics (TSE) - Manufacture des Tabacs
Aile Jean-Jacques Laffont - 21, allée de Brienne - 31015 TOULOUSE Cedex 6
Tél : +33 (0) 5 61 12 85 89 - Fax : +33 (0) 5 61 12 86 37 - **www.tse-fr.eu** - contact@tse-fr.eu

Toulouse School of Economics

Noblesse Oblige? Moral Identity and Prosocial Behavior in the Face of Selfishness.[*]

Roberta Dessí[†]and Benoît Monin[‡]
Toulouse School of Economics and Stanford University.

November 2, 2012

Abstract

What makes individuals conform or diverge after observing prosocial or selfish behavior by others? We study experimentally how social comparison (observing a peer's behavior) interacts with identity motives for cooperation. Participants play two games. We increase the strength of the identity motive by inducing subjects in a treatment condition to infer their identity from behavior in the first game. Cooperators who observe a peer defect donate 28% more to their unknown partner in the second game in the treatment than in the control group. Our results are consistent with the predictions of Bénabou and Tirole (2011), and show that the "sucker-to-saint effect" identified by Jordan and Monin (2008) can have important behavioral consequences.

JEL classification: A1, A12, D1, D3, D64, Z1

Keywords: prosocial behavior, cooperation, moral identity, self-image, social comparison, conformity, rationality.

[*] We would like to thank Roland Bénabou, Florian Englmaier, Ernst Fehr, Luigi Guiso, Yinghua He, John Helliwell, David Levine, Massimo Marinacci, Klaus Schmidt, Jean Tirole, Joel Van der Weele, seminar participants in Munich and Zurich, and participants in the CIFAR June 2012 Meeting for many valuable comments and suggestions. We thank Samuele Centorrino and Olivier Darmouni for excellent research assistance, and the ANR for financial support. The paper was partly written while Roberta Dessí was visiting EIEF, whose hospitality is gratefully acknowledged.

[†] IDEI, Toulouse School of Economics, Manufacture des Tabacs, Aile Jean-Jacques Laffont, 21 Allée de Brienne, 31000 Toulouse, France (roberta.dessi@tse-fr.eu).

[‡] Stanford Graduate School of Business, Knight Management Center, 655 Knight Way, Stanford, CA 94305-7298 (monin@stanford.edu)

1. Introduction

There is substantial evidence that individuals tend to behave more prosocially after observing others behave prosocially - and more selfishly after seeing others acting selfishly: for example, this is evident in studies of charitable contributions (e.g. Frey and Meier (2004), Croson and Shang (2008))[1]. Yet people do not always follow others' example, and sometimes they even seem to react against it.[2] What makes individuals conform or diverge when it comes to prosocial behavior, and to what extent? Understanding this seems crucial to designing incentives and organizational structures, and implementing a broad range of public policies.[3]

One possible answer is that individuals choose actions to build and/or affirm their identity (Akerlof and Kranton (2000), Bénabou and Tirole (2011)). This identity motive then interacts with social comparison, leading sometimes to conformity and sometimes to divergence.[4] To investigate this interaction empirically, we need to be able to vary social comparison (the observed behavior of others), but also the strength of the identity motive. We do this by designing an experiment in which subjects play two games, where they can act prosocially or selfishly. In each game, acting prosocially is socially efficient: it increases others' payoffs more than it decreases own payoff. Before the second game, subjects observe how another subject behaved in the first game: this is the source of social comparison. To vary the strength of the identity motive, we have a treatment condition in which subjects, after the first game but before learning they will be playing a second game, are induced to infer their identity from their choices in the first game. In the control condition this induced identity inference step is absent. Individuals may still, to some extent, infer their identity from their previous actions, but on average they are less likely to think carefully about the identity implications of behavior than in the treatment condition, where they have to answer a sequence

[1] See also, among others, Alpizar et al. (2008), Andreoni et al. (1998), Heldt (2005), Martin and Randal (2008), Rothstein (2000), Shang and Croson (2009).

[2] For instance, in a study of charitable contributions, Cueva and Dessí (2012) find that the average contribution of "followers" is *higher* after observing two leaders who donate zero, relative to one leader who donates zero and one who donates a positive amount.

[3] For example, organizations routinely make decisions that affect the visibility and salience of different individuals' actions. Governments and government-funded agencies aiming to improve health and environmental quality, or fight crime, carefully select information about individuals' and groups' behavior for their communication and advertising campaigns, and showcase stories to inspire people to emulate certain behaviors and reject others.

[4] For models of either conformity *or* divergence see, for example, Akerlof (1982), Andreoni (1989), Bernheim (1994), and Sugden (1984).

of questions that focus attention specifically on this issue.

We can therefore explore how behavior in the second game varies with own behavior in the first game, with the observed behavior of a peer, and with the strength of the identity motive (control versus treatment condition). We test three main hypotheses. First, we would expect to find that individuals who care more about the well-being of others are more willing to behave prosocially. Thus, to the extent that decisions in the first game reflect such individual differences in "social" preferences, they should be correlated with decisions in the second game: participants who are more generous in the first game should also be more generous in the second. Second, we would expect behavior in the second game to be influenced by the observed choice made by the peer in the first game. In particular, the well-documented tendency for people to behave more (less) prosocially after observing that others have behaved more (less) prosocially, noted above, suggests a second prediction: individuals should be more generous in the second game after observing a generous peer than after observing a selfish peer.

Our third and key hypothesis is that behavior in the second game will differ between the control and the treatment condition. In the latter, after observing a peer's decision in the first game but before learning that they will play again a similar game, participants are asked to rate themselves, and the peer, on two dimensions: a seven-point generosity scale, going from "very selfish" to "very generous", and a seven-point rationality scale, going from "very rational" to "very irrational". Those who behaved prosocially (henceforth called "givers" for ease of exposition) can think of themselves as having a generous disposition. Those who behaved selfishly (henceforth called "keepers") can think of themselves as rational. Since the questions in the treatment condition have no impact on monetary payoffs, standard economic models would not predict any treatment effect on behavior in the second game. Bénabou and Tirole (2011), on the other hand, would predict that, by making participants think more carefully about the identity implications of their actions, the treatment condition would increase what they call "identity investments" - i.e., behavior aimed at increasing confidence in one's identity. If indeed this is the case, we might expect givers to be more generous in the treatment than in the control condition (because of investments in the "generous" identity), and keepers to be less generous (because of investments in the "rational" identity).

We find support for all three hypotheses. Givers are more generous than keepers in the second game, in both the control and the treatment condition, and in the face of congruent or divergent behavior by the observed peer. This is consistent with an important role for *social preferences.* The role of *social comparison* is also important: participants are more generous in the second game after observing a generous peer than after observing a selfish peer. This is true for keepers as well as givers, in both the control and the treatment condition.

Our third main finding is that in the treatment condition, givers are more generous and keepers more selfish, on average, than in the control condition. This is consistent with an important role for *identity investments.* However, the only statistically significant treatment effect is the greater generosity by givers who have observed divergent behavior by the peer (i.e., a keeper).

To better understand this result, we go on to study subjects' responses to the ratings questions in the treatment condition. This reveals our fourth main result: when we examine givers who have observed a keeper (i.e. the subjects for whom we obtain a significant treatment effect), we find that those who have rated themselves as more *rational* tend to behave more prosocially in the second game. Those who have rated themselves as both more rational *and* more generous give even more. Thus the significance of our treatment effect is not driven simply by givers investing in a "generous" identity, as conjectured earlier: investments are greater for a combined (and more appealing) identity as both generous and rational. This interpretation is also consistent with our last finding, namely that givers and keepers tend to agree in rating givers as more generous than keepers, but givers think they are more rational than keepers while keepers think they are more rational than givers. In other words, generosity has a clear, consensual meaning for our participants, while the concept of rationality lends itself to different interpretations: keepers can view themselves as more "rational", as we predicted, but interestingly, givers can also view themselves as more "rational", perhaps by thinking in terms of collective rationality. This has the advantage of yielding a very positive self-image for givers; on the other hand, maintaining that positive self-image requires consistency in subsequent behavior. Indeed, it is precisely the givers who emphasize the rationality as well as generosity of prosocial behavior who go on to behave more prosocially in the second game: *noblesse oblige*?

The remainder of the paper is organized as follows. We complete this section with a review of the related literature in economics and psychology. Section 2 describes our experimental design and procedures. Our results are presented in section 3. We discuss our findings and provide some conclusions in section 4.

1.1. Literature Review

Our paper is related to several strands of the literature in economics and psychology that insist on the importance of identity processes[5]. Akerlof and Kranton (2000) were the first to highlight the importance of identity for economics, arguing that individuals experience a loss of utility when they fail to conform to the prescriptions of a chosen or ascribed identity. Bénabou and Tirole (2011) have developed a model of identity investments in which individuals, who have imperfect self-knowledge, infer their identity ("true nature") in part from observation of their past behavior. In psychology, this approach can be traced back to Bem (1967), who argued that individuals tend to infer their dispositions from their behavior. Our experimental methodology essentially manipulates the extent to which subjects engage in such inference, by focusing their attention on the identity implications of past behavior in the treatment condition (and not in the control condition). This enables us to test the theoretical prediction from Bénabou and Tirole that identity investments should increase when identity concerns are made more salient.

By asking individuals to reflect on their identity between two iterations of the same game, we rely both on the fact that people's identity is shaped by past behavior, and on the fact that future behavior is shaped by past identity. In other words, identity can serve as a bridge between past and future behavior, and, as we show, the consistency of past and future behavior can be increased by the crystallization of an identity, or the fact that the implications of past behavior for one's identity are drawn. Thus our paper is also related to recent work in economics showing that individuals value consistency in a variety of contexts (Falk and Zimmermann (2011)). Our study takes this line of research forward by exploring the link between consistency and identity.

What identities or personality traits do individuals infer from their own choices in an experimental game? Psychologists investigating this question have shown that individuals who tend to act self-interestedly in such games see cooperators as naïve or weak, whereas individuals who tend to act cooperatively see defectors as unfair or mean (the "might-over-morality" phenomenon, see Liebrand, Jansen,

[5] Our main focus is on moral identity. There is now a substantial body of evidence in economics showing that many people care about their moral identity: they do not like to perceive themselves, and to be perceived by others, as unfair, selfish, opportunistic or dishonest. See, among others, Andreoni and Bernheim (2009), Ariely, Bracha and Meier (2009), Broberg, Ellingsen and Johannesson (2007), Cueva and Dessí (2012), Dana, Weber and Kuang (2007), Dana, Cain and Dawes (2006), Van der Weele (2012).

Rijken, & Suhre, 1986). Thus individuals draw different conclusions based on their own value orientation, and both groups are able to conclude that they possess a self-enhancing identity: defectors think they are strong and smart, while cooperators think they are kind and sympathetic. Our results are similar in one respect - both givers and keepers in our experiment are able to conclude that they possess a self-enhancing identity - but also different in an interesting way: rather than full separation of identities, with givers viewing themselves as generous and keepers as rational, we observe separation along the generosity/selfishness identity dimension, but competing claims to a "rational" identity.[6]

Our findings are therefore related to recent work by Butler, Giuliano and Guiso (2012), who find that participants in trust game experiments have different notions of what constitutes cheating in a trust exchange, correlated with values instilled by their parents. While we do not investigate the link with parental values, we find that participants in our experiment hold quite different notions of what constitutes rational behavior, and these notions have a significant impact on behavior.

Finally, our work is related to recent research suggesting that individuals may moralize their behavior to compensate for feeling deficient after witnessing others acting more self-interestedly. Jordan and Monin (2008) demonstrated in two psychological studies that individuals boosted their moral self-image to justify not acting in self-interested ways. In one study, participants witnessed a peer (really a confederate) refuse to help out the experimenter on an optional task. Participants rated themselves as more moral (and the other as less moral) if they had first been induced to help out than if they had not. They also rated themselves as more moral if they helped and witnessed the peer refuse than if they helped without observing any peer. Jordan and Monin propose that after helping the experimenter, individuals witnessing the peer acting selfishly felt like suckers, and were able to reduce this threat to their self-image by deciding instead that they were better persons – and the refusing peer a worse person. That this tendency resulted from self-threat was demonstrated by a second study where a "self-affirmation" manipulation (Steele, 1988), in which individuals were induced to dwell on valued aspects of their own personality, eliminated this "sucker-to-saint" effect. These studies focused on self-report outcomes, and left open the question of behavioral

[6] Our results also differ in another important way: Liebrand et al. (1986) found that defectors exposed to cooperators continued to defect, whereas cooperators exposed to defectors started acting more individualistically. We do not find an analogous "race to the bottom" among our experimental subjects.

consequences. Our results show that ascribing past behavior to a more rational and more generous nature can have a significant impact on future behavior, and hence on economic outcomes.

2. Experimental Design and Procedures

Participants in our experiment were students at the University of Toulouse, and all the experimental sessions were carried out in the Experimental Economics Laboratory of the Toulouse School of Economics, using the software z-Tree (Fischbacher (2007)). The details of the experimental instructions are available in the Appendix.

Subjects were recruited by visiting the first or last five minutes of lectures given to students in Economics, Business and Finance, and Law. We informed students that they could, if they wished, volunteer to participate in experiments on decision-making by registering on the Laboratory's recruitment website. We told them that sessions could take up to 90 minutes, inclusive of individual confidential payments at the end of the experiment. Payments would depend on their decisions and those of other participants.

Participants were randomly allocated to control or treatment sessions. At the beginning of each session, each subject picked one out of a set of number identifiers, which determined the computer that was allocated to him for the duration of the experiment. At the end of the experiment, all participants left the laboratory and waited outside; they were then called back individually to return the number identifier and receive their earnings.

Throughout the experiment, earnings were referred to in terms of experimental currency units or "points". Participants were told at the beginning of the session that each point was worth 50 (Euro)cents.

In total, 256 subjects participated in the experiment, 120 in the control condition and 136 in the treatment. In the control group, 46% of participants were female, 58% were first-year undergraduates, and the average age was 20. In the treatment group, the proportions were: 58% female, 48% first-year undergraduates. The average age was again 20.

2.1. Treatment Condition

Subjects in this condition were allocated to groups of four, randomly and anonymously. Within each group, they were referred to neutrally as participants A, B,

C and D. They all received an initial endowment of 12 units of experimental currency or "points", worth a total of 6 euros. They were told that the experiment would proceed as follows. Participants A and B would decide, simultaneously and without knowing the other's decision, whether to keep all their endowment or to send 4 points (2 euros) to the other. Participants C and D faced the same decision. All amounts sent would be trebled by the experimenter; thus each subject could receive either zero or 12 points (6 euros), depending on his partner's decision. Payoffs were therefore interdependent within each pair and independent between pairs in any group.

After the allocation decisions were made, each participant was reminded of his choice (gave zero or gave four) and at the same time learned the choice made by one of the players in the other pair in his group (e.g. A learned whether C had given zero or four), without learning his own partner's decision (which was only revealed at the end of the experiment) and hence his payoff. In what follows, we refer to the other player whose decision is observed as "the peer".

At this point, and without knowing that they would then face another decision, subjects were asked six questions. In the first two, they were asked to rate their peer, and themselves, on a seven-point scale going from 1 = very selfish to 7 = very generous. The following two questions asked for the same ratings (peer and self) on a scale going from 1 = very rational to 7 = very irrational. The last two questions asked subjects to imagine an experiment where participants were given the same endowment (12 points), and could freely allocate it between themselves and their partner, who was facing the same choice (knowing that the amount sent to the partner would be trebled by the experimenter). In question 5, they were asked to estimate the minimum amount that a generous person would send. In question 6, they were asked to estimate the maximum amount that a rational person would send.

After answering these questions, subjects were told that they would now participate in a second, similar experiment, which would be the last. They would then receive their combined earnings from the two experiments. In this second experiment, they would be re-matched, randomly and anonymously, so as to play in another group of four, where none of the other players had belonged to their group in the first experiment. Moreover, no participant in the new groups would know how the other members of his group had played during the first experiment. Subjects would receive a new endowment of 12 points for this second experiment, and play the same allocation game as in the first, with one difference: they would be free to allocate their endowment as they wished (to the nearest unit) between themselves and their partner. Amounts sent would be trebled by the experimenter, as before.

After all the allocation decisions were made, subjects learned their payoffs from each of the two experiments, and hence their total earnings.

2.2. Control Condition

In this condition, there was only one difference relative to the treatment: subjects were asked to answer the six questions described above only at the very end (i.e. after the two experiments). The questions were therefore very slightly modified so as to remind subjects of the first experiment and obtain their ratings accordingly. Similarly the questions about maximum and minimum amounts were phrased with reference to the second experiment rather than an imagined experiment. In this condition answering the questions could not have any impact on one's choice in the second experiment, because the questions came after the choice.

3. Results

We first describe the data, then go on to investigate our main hypotheses.

3.1. Descriptive Statistics

We begin by summarizing the data in Table 3.1. Roughly half of the participants in both conditions chose to give a third of their endowment to their partner in the first game, while the other half kept it all for themselves. Specifically, the proportion of "givers" (who sent 4 units to their partner) was 51% in the control and 45% in the treatment condition. As a consequence, the average amount sent, shown in Table 3.1, was 2.033 in the control and 1.794 in the treatment. The difference was not statistically significant.[7]

In both conditions, but crucially at different times, subjects were asked six questions, described in detail in the Appendix. Four of these questions involved rating the peer and the self on two dimensions, generosity and rationality. Average ratings on the generosity dimension were very similar in the two conditions, with peers being rated slightly more generous in the treatment than in the control

[7] Throughout the paper, p-values for pairwise comparisons refer to Mann-Whitney U-tests unless otherwise stated.

Table 3.1: Descriptive Statistics

	Control	Treatment	Difference
Variable	Mean (SD)	Mean (SD)	p-value
Amount sent	2.033	1.794	0.340
in first stage	(2.008)	(1.997)	
Amount sent	2.125	2.162	0.679
in second stage	(2.522)	(2.842)	
Minimum sent	4.167	4.140	0.751
if generous	(2.059)	(2.486)	
Maximum sent	4.133	3.743	0.279
if rational	(3.462)	(3.455)	
Self-rating	3.925	3.949	0.889
(generosity)	(1.788)	(1.421)	
Peer-rating	3.867	4.022	0.684
(generosity)	(1.815)	(1.594)	
Self-rating	2.717	2.478	0.310
(rationality)	(1.562)	(1.333)	
Peer-rating	3.633	3.243	0.103
(rationality)	(1.883)	(1.745)	

Generosity rating: 1= very selfish; 7 = very generous

Rationality rating: 1 = very rational; 7 = very irrational group. Average ratings on the rationality dimension were quite similar, with both peers and the self being rated a little less rational in the treatment group.

Average amounts were very similar in the two conditions when answering the question about the minimum that a generous person would give (a little over four units). When deciding the maximum that a rational person would give, subjects' answers were on average somewhat lower in the treatment, but the difference is not significant.

In the second game, while almost all possible allocations were observed, the average amount sent was 2.125 units in the control and 2.162 in the treatment condition (i.e. a little over the average amount sent in the first game).

3.2. Main Results

We begin by analyzing behavior in the second game, then examine participants' answers to the questions about generosity and rationality. In our analysis, we distinguish between "givers" (who chose the option to send four units to their partner in the first game) and "keepers" (who chose to keep all their endowment in the first game).

3.2.1. Behavior in Game 2

Table 3.2 investigates our first two hypotheses. On average, givers sent 3.164 units, while keepers sent 1.216 units. The difference is highly significant, and supports our first hypothesis concerning the role of social preferences: *individuals who gave more in the first game also gave more in the second game.* The table then shows the mean amount sent in the second game as a function of the observed peer's behavior in the first game: we find that subjects sent *significantly more* in the second game *after observing a giver* (2.525 units) than after observing a keeper (1.799 units). This result supports our second hypothesis concerning the role of social comparison: *individuals tend to behave more prosocially (cooperatively) after observing more prosocial behavior by peers.*

Our third and key hypothesis is that *in the treatment condition, givers send more and keepers send less, on average, than in the control condition.* The means presented in Table 3.3 are consistent with the hypothesis. However, the p-values from the Mann-Whitney U-tests reported in the last line of the table show that the treatment effect is only statistically significant for one group: givers who observe a keeper. Moreover, the difference in means for this group implies a larger effect.

Table 3.2: Behavior in Game 2: Hypotheses 1 & 2

	Mean amount sent		p-value
Hypothesis 1	Givers	Keepers	
	3.164	1.216	0.000
Hypothesis 2	Peer: giver	Peer: keeper	
	2.525	1.799	0.042

To better understand this result, we go on to examine subjects' responses to the ratings questions.

Table 3.3: Mean amounts sent in Game 2: Hypothesis 3

	Givers		Keepers	
Peer	Giver	Keeper	Giver	Keeper
Means				
Control	3.346	2.486	1.629	1.000
Treatment	3.750	3.172	1.517	0.826
p-values				
	0.770	0.020	0.203	0.281

3.2.2. Generosity and Rationality

Tables 3.4 to 3.6 summarize the answers to our questions about generosity and rationality. Table 3.4 shows the average ratings for the self and the peer on the generosity scale, going from 1 = very selfish to 7 = very generous.

For the treatment condition, we see that givers who have a keeper as a peer have an average *self*-rating of 4.759, while their average rating for the *peer* is 2.414. Thus on average *givers rate themselves as substantially more generous than keepers*, and the difference is significant ($p = 0.000$). Keepers who have a giver as a peer share this view: their average *self*-rating is 2.724, while their average rating for the *peer* is 5.483 ($p = 0.000$).

Table 3.5 presents average ratings for the self and the peer on the (ir)rationality scale, going from 1 = very rational to 7 = very irrational. For the treatment condition, we find that the average *self*-rating for keepers who have a giver as a peer

Table 3.4: Generosity: mean ratings

Self	Peer: Keeper	Peer: Giver	Peer: Keeper	Peer: Giver
	Control		Treatment	
Self-rating				
Givers	4.857	5.038	4.759	5.219
Keepers	3.000	2.800	3.326	2.724
Peer-rating				
Givers	2.086	5.077	2.414	5.188
Keepers	2.917	5.400	3.304	5.483
1 = very selfish		7 = very generous		

Table 3.5: Rationality: mean ratings

Self	Peer: Keeper	Peer: Giver	Peer: Keeper	Peer: Giver
	Control		Treatment	
Self-rating				
Givers	3.000	2.885	3.103	2.813
Keepers	2.292	2.600	1.913	2.379
Peer-rating				
Givers	4.171	2.769	4.000	3.344
Keepers	2.208	4.714	2.000	4.345
1 = very rational		7 = very irrational		

is 2.379, while their average rating for the *peer* is 4.345, implying that keepers indeed view givers as much more irrational than themselves ($p = 0.000$). Interestingly though, givers do *not* share this view: givers who have a keeper as a peer tend to rate keepers as more irrational than themselves (average self-rating: 3.103; average peer-rating: 4.000), although the difference is not statistically significant ($p = 0.104$).

Overall, the evidence from self ratings and peer ratings shows that, as predicted, givers view themselves as significantly more generous than keepers, while keepers view themselves as significantly more rational than givers. In addition, we see that givers and keepers *agree* that givers are more generous, but they *disagree* on who is more rational: keepers think that keeping all the money is more rational, while givers seem to find that there are good rational reasons for sending some of the money. This difference also emerges in the answers to the question asking what is the maximum amount that a rational person would send to the partner, summarized in Table 3.6. In the treatment condition, the mean answer for givers is substantially higher than for keepers (5.966 versus 2.370; $p = 0.000$; 5.125 versus 2.172; $p = 0.000$).

Table 3.6: Mean thresholds: Maximum amount a rational person would send, and Minimum amount a generous person would send

Self	Peer: Keeper	Peer: Giver	Peer: Keeper	Peer: Giver
	Control		Treatment	
Max if rational				
Givers	5.143	4.538	5.966	5.125
Keepers	3.958	2.943	2.370	2.172
Min if generous				
Givers	4.143	4.846	4.207	5.000
Keepers	4.292	3.600	3.630	3.931

These results suggest that keepers, as we conjectured, rationalized their behavior in the first game as rationally selfish. Givers, on the other hand, were able to rationalize their behavior as both more generous *and* at least as rational as keepers' behavior, perhaps by thinking in terms of collective rationality. This yielded a more appealing identity, which may explain why identity effects were then stronger for these subjects in the second game.

For participants in the treatment condition, we further explored this interpretation by examining the relationship between the amount sent in the second game, and the answers given to the ratings questions. Tables 3.7 $-$ 3.10 present the results of Tobit regressions for the four different groups in the treatment condition. The dependent variable is the amount sent in the second game. The variable "generosity" measures the subject's self-rating on the generosity scale. We constructed the "rationality" variable by recoding our rationality scale from $1 =$ very irrational to $7 =$ very rational. Thus "rationality" measures the subject's self-rating on this recoded scale. We normalized both variables to ease interpretation of coefficients[8]. The variable "genrational" is the product of these normalized ratings and captures the interaction between "generosity" and "rationality". The

[8] Let X be the raw self-rating score: we use the normalized score $X' = (X - \hat{\mu})/\hat{\sigma}$, where $\hat{\mu}$ is

answers to the last two questions are coded as "mingen" (the minimum amount that a generous person would send) and "maxrat" (the maximum amount that a rational person would send).

Table 3.7: Tobit Regression for Amount Sent in Second Game: Givers Who Observed a Keeper

Variable	Coeff.	Std. Err.	t	p-value
Generosity	-0.630	0.386	-1.63	0.116
Rationality	0.785**	0.376	2.09	0.048
Genrational	0.702*	0.354	1.99	0.059
Mingen	-0.233	0.197	-1.19	0.248
Maxrat	-0.042	0.103	-041	0.685

*p $<$ 0.10, **p $<$ 0.05, ***p $<$ 0.01
Obs: 29

Table 3.7 shows the results for the group we are most interested in: givers who observe a keeper. For this group, the coefficient for generosity (estimated at the mean for rationality) is not significant, but the coefficient for rationality (estimated at the mean for generosity) is positive and significant, with individuals who rated themselves as more rational donating more in the second game. Moreover, the interaction term is positive, and significant at the 10% level ($p = 0.059$): the greatest donations in the second game thus came from people who had rated themselves *both* rational and generous in the first game (where all givers had objectively given the same amount), and the lowest donations came from individuals who rated themselves as generous but irrational.

Table 3.8 shows that for givers who observed another giver, two effects were significant. Individuals gave more in the second game after stating a higher threshold for the maximum amount that a rational person would give. They gave less, on the other hand, after rating themselves as more rational. This contrasts with the results for givers who observed a keeper: they gave more after rating themselves as more rational. Interestingly, the difference is consistent with the implications of Bénabou and Tirole (2011), to the extent that givers who rated themselves as more rational were more confident (making identity investments unnecessary) after observing another giver than after observing a keeper.

the sample mean of X and $\hat{\sigma}$ is the sample standard deviation of X. With this normalization, coefficients measure the expected change in the amount sent associated with a change in the self-rating score of one standard deviation (at the mean of the other self-rating).

Table 3.8: Tobit Regression for Amount Sent in Second Game: Givers Who Observed a Giver

Variable	Coeff.	Std. Err.	t	p-value
Generosity	-1.071	0.733	-1.46	0.156
Rationality	-1.819**	0.851	-2.14	0.042
Genrational	1.316	0.904	1.46	0.157
Mingen	0.167	0.228	0.73	0.471
Maxrat	0.774***	0.228	3.39	0.002

*p < 0.10, **p < 0.05, ***p < 0.01
Obs: 32

Table 3.9: Tobit Regression for Amount Sent in Second Game: Keepers Who Observed a Keeper

Variable	Coeff.	Std. Err.	t	p-value
Generosity	0.269	0.565	0.48	0.637
Rationality	-0.164	0.490	-0.34	0.739
Genrational	0.484	0.562	0.86	0.394
Mingen	-0.274	0.281	-0.97	0.337
Maxrat	1.083***	0.238	4.56	0.000

*p < 0.10, **p < 0.05, ***p < 0.01
Obs: 46

The two groups of keepers in the treatment condition (Tables 3.9 and 3.10) also exhibited a strong positive relationship between their declared maximum amount a rational person would send, and the amount they actually sent in the second game. This was the only significant explanatory variable for both groups.[9]

[9] To explore the potential role of collinearity between explanatory variables, all the Tobit regressions in Tables 3.7 to 3.10 were estimated again excluding the Mingen and Maxrat variables. We found the same pattern of signs and significance for the estimated coefficients, with one exception: the coefficient for Rationality was no longer significant in the regression for givers who observed a giver, corresponding to Table 3.8.

Table 3.10: Tobit Regression for Amount Sent in Second Game: Keepers Who Observed a Giver

Variable	Coeff.	Std. Err.	t	p-value
Generosity	0.294	0.612	0.48	0.635
Rationality	0.762	0.576	1.32	0.198
Genrational	0.113	0.643	0.17	0.863
Mingen	0.336	0.276	1.22	0.235
Maxrat	1.407***	0.340	4.14	0.000

*p < 0.10, **p < 0.05, ***p < 0.01
Obs: 29

4. Conclusions

We have examined how inducing individuals to infer their identity from their past behavior affects their subsequent decisions to cooperate. By asking them to rate themselves, and a peer, on a "generosity" scale and on a "rationality" scale, we focused the attention of subjects in the treatment condition on the identity implications of behavior. We found that subsequent choices were consistent with the implications of Bénabou and Tirole (2011).

Our main result is that individuals who behave generously and then observe a peer who has behaved selfishly tend to behave *more generously* after being induced to estimate how generous and rational they and their peer were being, relative to a control condition where subjects are not explicitly induced to make any inference about their identity. The effect is economically important: average donations for this group are 28% higher in the treatment than in the control condition. To trace the causes of this effect, we examined subjects' responses to the ratings questions in the treatment condition. We found that givers who observe a keeper, when asked to evaluate the peer and themselves, tend to rationalize the difference in previous behavior by rating themselves as much more generous, and no less rational - hence "superior" to their peer. However, this self-image construal has an effect on preferences, since the psychological benefits from a positive self-image as a rationally generous person would be undermined by subsequent behavior that contradicted this perception. Hence preferences evolve, endogenously, towards greater generosity. This is reflected in higher amounts being sent to partners in the second game relative to the control condition, in which subjects are not required to evaluate themselves and their peer, and therefore do not engage in the same self-image construal.

This interpretation is consistent with the finding that, within the group of givers who observe a keeper, the ones who go on to donate most are precisely those who have rated themselves as both rational and generous. Those who had rated themselves as generous but irrational, on the other hand, donate least: these subjects essentially felt like "suckers" (in the Jordan and Monin (2008) terminology), and may have tried to make up for it by acting more "rationally" in the second game.

The psychology of the givers who behaved more generously in the second game, having rated themselves as rational and generous, deserves further study. Two very different processes could be at work, with very different welfare implications. One possibility is that the opportunity to justify behavior in the first game as both generous and rational has a liberating effect, enabling the individual who genuinely cares about collective welfare and efficiency to be more generous in the second game without fear of coming across as a sucker or an idiot, or feeling undue pressure to fall in line with a selfish peer. In this view, the individual is made better off by being able to dig her heels and stick to her guns. The second possibility is that individuals were not strongly motivated by a concern for collective welfare and efficiency, but gave in the first game partly because of image concerns (and optimistic expectations about their peers' behavior): in this case, observing divergent behavior by a peer could have a liberating effect - but less so after individuals are led to rationalize their initial behavior relative to the peer. In this view, individuals who play the rational generosity card to feel better after the first game are then stuck with having to be consistent in the second game, for fear of coming across as hypocritical otherwise. They are thus forced to be generous, rather than liberated to be: *noblesse oblige*. This implies that they might have been better off if they had not been induced to infer their identity from their initial behavior. We leave it to future research to distinguish between these two possibilities.

5. References

Akerlof, G.A. (1982) "Labor contracts as partial gift exchange", *Quarterly Journal of Economics*, 97(4), 543-569.

Akerlof, G.A. and R. E. Kranton (2000) "Economics and identity", *Quarterly Journal of Economics*, 115(3), 715-753.

Alpizar, F., Carlsson, F., and O. Johannsson-Stenman (2008) "Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica", *Journal of Public Economics*, 92(5-6), 1047-1060.

Andreoni, J. (1989) "Giving with impure altruism: applications to charity and Ricardian equivalence", *Journal of Political Economy*, 97(6), 1447-1458.

Andreoni, J., Erard, B., and J. Feinstein (1998) "Tax compliance", *Journal of Economic Literature*, 36(2), 818-860.

Andreoni, J. and D. Bernheim (2009) "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects", *Econometrica,* 77(5), 11-28.

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

Bem, D.J. (1967) "Self-perception: An alternative interpretation of cognitive dissonance phenomena", *Psychological Review*, 74, 183-200.

Bénabou, R. and J. Tirole (2011) "Identity, Morals and Taboos: Beliefs as Assets", *Quarterly Journal of Economics*, 126(2), 805-855.

Bernheim, D. (1994) "A theory of conformity", *Journal of Political Economy*, 102(5), 841-877.

Broberg, T., Ellingsen, T. and M. Johannesson (2007) "Is Generosity Involuntary?", *Economic Letters*, 94, 32-37.

Butler, J.V., Giuliano, P., and L. Guiso (2012) "Trust and Cheating", mimeo, EIEF.

Croson, R. and J. Shang (2008) "The impact of downward social information on contribution decisions", *Experimental Economics*, 11(3), 221-233.

Cueva, C. and R. Dessí (2012) "Charitable Giving, Self-Image and Personality", IDEI Working Paper, Toulouse.

Dana, J., Cain, D.M., and R. Dawes (2006) "What You Don't Know Won't Hurt Me: Costly (But Quiet) Exit in Dictator Games", *Organizational Behavior and Human Decision Processes*, 100(2), 193-201.

Dana, J., Weber, R.A. and J.X. Kuang (2007) "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness", *Economic Theory*, 33(1), 67-80.

Falk, A. and F. Zimmermann (2011) "Preferences for consistency", IZA Discussion Paper.

Fischbacher, U. (2007) "z-tree: Zurich toolbox for ready-made economic experiments", *Experimental Economics*, 10(2), 171-178.

Frey, B.S. and S. Meier (2004) "Social Comparisons and Pro-Social Behavior:

Testing "Conditional Cooperation" in a Field Experiment", *American Economic Review*, 94(5), 1717-1722.

Heldt, T. (2005) "Conditional cooperation in the field: Cross-country skiers' behavior in Sweden", Working Paper, Dalarna University College Electronic Archive.

Jordan, A. H. and B. Monin (2008) "From sucker to saint: Moralization in response to self-threat", *Psychological Science*, 19(8), 683-689.

Liebrand W.G.B., Jansen R.W.T.L., Rijken V.M. and C.J.M. Suhre (1986) "Might over Morality: Social Values and the Perception of Other Players in Experimental Games", *Journal of Experimental Social Psychology*, 22, 203-215.

Martin, R. and J. Randal (2008) "How is donation behaviour affected by the donations of others?", *Journal of Economic Behavior and Organization*, 67(1), 228-238.

Rothstein, B. (2000) "Trust, social dilemmas and collective memories", *Journal of Theoretical Politics*, 12(4), 447-501.

Shang, J. and R. Croson (2009) "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods", *The Economic Journal*, 119, 1422-1439.

Steele, C.M. (1988) "The psychology of self-affirmation: sustaining the integrity of the self", in L. Berkowitz (Ed.) *Advances in Experimental and Social Psychology*, Vol.21, 261-302. New York: Academic Press.

Sugden, R. (1984) "Reciprocity: the supply of public goods through voluntary contributions", *The Economic Journal*, 94, 772-787.

Van der Weele, J.J. (2012) "When Ignorance is Innocence: On Information avoidance in Moral Dilemmas", SSRN Working Paper.

6. Appendix

6.1. Experimental Instructions

Welcome; you are going to participate in an economics experiment. Your answers and decisions will have no consequence for your grades or your degree.

This experiment studies decision-making. There are no correct or false answers - you should simply decide according to your preferences.

This experiment will be remunerated. Your remuneration will depend on your decisions and the decisions of the other participants. All amounts will be expressed in units of experimental currency, or "points". One point is worth 50 (Euro)cents.

We ask you to switch off your mobile phones and not to talk to each other during the experiment. If you have a question, raise your hand and we will come to answer.

Are there any questions now? If there are no questions, we can start. You will see some instructions on your screen. Read them carefully. Whenever you are asked a question, take the time you need to answer. If you see the phrase "Waiting for other players" on your screen, it means you have to wait for the others to give their answers before continuing.

6.1.1. General instructions

During this experiment you will sometimes be asked to take decisions that will affect the outcome for you and for other participants. It is important to know that your decisions will remain completely anonymous.

Each person will be assigned to a group of four participants, depending on the ID you chose before the start of the experiment. You will never know who were the other members of your group, and they will never know you were in their group.

Within each group of four participants, there will be a participant "A", a participant "B", a participant "C" and a participant "D". Each participant will learn her role (A, B, C or D) shortly. When we refer to other members of your group, we will always use the letter (A,B, C or D) and never their ID or other information that could allow you to identify them.

If you have a question raise your hand. If there are no questions we can give you the specific instructions.

6.1.2. Specific instructions

Now we give each participant an endowment of 12 points. In each group of four participants, A has to decide how much to keep for herself, and how much to give to B. At the same time, B has to decide how much to keep for herself, and how much to give to A. Each person has to decide before knowing the other's decision. The same applies to participants C and D: C has to decide how much to keep for herself, and how much to give to D. At the same time, D has to decide how much to keep for herself, and how much to give to C.

Important: the amount received will be **three times** the amount given. Example: if A gives 4 points, B will receive 12 points. If B gives 4 points, A will receive 12 points.

Decisions have to respect the following rules. Each participant can choose between two options:

Option 1: He keeps his endowment of 12 points and gives nothing. In this case the other receives nothing from him.

Option 2: He keeps 8 points and gives 4 points. In this case, the other receives 12 points from him.

[NEXT SCREENSHOT]

Instructions to A

[*equivalent instructions were given to B, C and D*]

You are participant A in your group. You now have an endowment of 12 points. You can choose between the following two options:

Option 1: you keep 12 points and give nothing. So B will receive nothing from you.

Option 2: You keep 8 points and give 4 points. So B will receive 12 points from you.

Reminder: B is also choosing between option 1 (giving you zero) and option 2 (giving you 4 points; in this case, you would receive 12 points).

Take the time you need to think before making your decision.

[NEXT SCREENSHOT]

Instructions to A

[*equivalent instructions were given to B, C and D*]

Now every member of your group has decided.

You have chosen:

either Option 1 (give zero to B) or Option 2 (give 4 points to B)

Participant C in your group has chosen:

either Option 1 (give zero to D) or Option 2 (give 4 points to D)

[NEXT SCREENSHOT]

[**treatment condition only**]

Instructions to A

[*equivalent instructions were given to B, C and D*]

You have given ?? points to B. At the same time, participant C has given ?? points to D.

In your opinion, which of the following would best describe participant C?

1. Very selfish
2. Quite selfish
3. A little selfish
4. Neither selfish nor generous.

5. A little generous.
6. Quite generous.
7. Very generous.

[NEXT SCREENSHOT]

[**treatment condition only**]

Instructions to A

[*equivalent instructions were given to B, C and D*]

You have given ?? points to B. At the same time, participant C has given ?? points to D.

In your opinion, which of the following would best describe you?

1. Very selfish
2. Quite selfish
3. A little selfish
4. Neither selfish nor generous.
5. A little generous.
6. Quite generous.
7. Very generous.

[NEXT SCREENSHOT]

[**treatment condition only**]

Instructions to A

[*equivalent instructions were given to B, C and D*]

You have given ?? points to B. At the same time, participant C has given ?? points to D.

In your opinion, which of the following would best describe participant C?

1. Very rational
2. Quite rational
3. A little rational
4. Neither rational nor irrational.
5. A little irrational.
6. Quite irrational.
7. Very irrational.

[NEXT SCREENSHOT]

[**treatment condition only**]

Instructions to A

[*equivalent instructions were given to B, C and D*]

You have given ?? points to B. At the same time, participant C has given ?? points to D.

In your opinion, which of the following would best describe you?

1. Very rational
2. Quite rational
3. A little rational
4. Neither rational nor irrational.
5. A little irrational.
6. Quite irrational.
7. Very irrational.

[NEXT SCREENSHOT]

[**treatment condition only**]

If every participant had an endowment of 12 points and could choose freely how much to keep for herself and how much to give to the other member of her group (who would receive three times the amount given), what would be, in your opinion, the **minimum amount** that a **generous** person would give?

[NEXT SCREENSHOT]

[**treatment condition only**]

If every participant had an endowment of 12 points and could choose freely how much to keep for herself and how much to give to the other designated member of her group (who would receive three times the amount given), what would be, in your opinion, the **maximum amount** that a **rational** person would give?

[NEXT SCREENSHOT]

Now you are going to participate in a similar experiment, but with a different group. In this new group, no participant will have played with you before. You will not know how they played the first time, and they will not know how you played the first time.

The size of each group will be the same: 4 participants.

In this second experiment, which will be the last, each participant will have a new endowment of 12 points. He or she will be able to choose freely how much to keep and how much to give to the other designated member of her group. At the end of this experiment, you will receive the total remuneration for the two experiments.

[NEXT SCREENSHOT]

Instructions to A

[*equivalent instructions were given to B, C and D*]

You are participant A in your new group. You now have a new endowment of 12 points.

You have to choose how much to keep for yourself, and how much to give to participant B of your new group. The amount given will be **multiplied by three**.

Reminder: participant B of your new group is also choosing how much to keep for herself and how much to give you.

Take the time you need to think before making your decision.

[NEXT SCREENSHOT]

[**control condition only**]

Now all the decisions have been taken.

In the first experiment, you kept ?? points of your endowment, and you received ?? points.

In the second experiment, you kept ?? points of your endowment, and you received ?? points.

[NEXT SCREENSHOT]

[**control condition only**]

Instructions to A

[*equivalent instructions were given to B, C and D*]

In the first experiment, you gave ?? points to B. At the same time, participant C gave ?? points to D.

In your opinion, which of the following would best describe participant C in that first experiment?

1. Very selfish
2. Quite selfish
3. A little selfish
4. Neither selfish nor generous.
5. A little generous.
6. Quite generous.
7. Very generous.

[NEXT SCREENSHOT]

[**control condition only**]

Instructions to A

[*equivalent instructions were given to B, C and D*]

In the first experiment, you gave ?? points to B. At the same time, participant C gave ?? points to D.

In your opinion, which of the following would best describe you?

1. Very selfish
2. Quite selfish
3. A little selfish

4. Neither selfish nor generous.
5. A little generous.
6. Quite generous.
7. Very generous.

[NEXT SCREENSHOT]

[**control condition only**]

Instructions to A

[*equivalent instructions were given to B, C and D*]

In the first experiment, you gave ?? points to B. At the same time, participant C gave ?? points to D.

In your opinion, which of the following would best describe participant C in that first experiment?

1. Very rational
2. Quite rational
3. A little rational
4. Neither rational nor irrational.
5. A little irrational.
6. Quite irrational.
7. Very irrational.

[NEXT SCREENSHOT]

[**control condition only**]

Instructions to A

[*equivalent instructions were given to B, C and D*]

In the first experiment, you gave ?? points to B. At the same time, participant C gave ?? points to D.

In your opinion, which of the following would best describe you?

1. Very rational
2. Quite rational
3. A little rational
4. Neither rational nor irrational.
5. A little irrational.
6. Quite irrational.
7. Very irrational.

[NEXT SCREENSHOT]

[**control condition only**]

In the second experiment, every participant had an endowment of 12 points and could choose freely how much to keep for herself and how much to give to the other member of her group (who received three times the amount given). In your opinion, what would be the **minimum amount** that a **generous** person would give?

[NEXT SCREENSHOT]

[**control condition only**]

In the second experiment, every participant had an endowment of 12 points and could choose freely how much to keep for herself and how much to give to the other member of her group (who received three times the amount given). In your opinion, what would be the **maximum amount** that a **rational** person would give?

[NEXT SCREENSHOT]

Thank you for participating in the experiment. Please complete the following questionnaire before leaving the laboratory.