

# Eliciting respect for social norms\*

Erik O. Kimbrough<sup>†</sup>

Alexander Vostroknutov<sup>‡</sup>

July 4, 2016

## Abstract

Recent models of prosociality suggest that cooperation in laboratory games may be better understood as resulting from concern for social norms than from prosocial preferences over outcomes. Underlying this interpretation is the idea that people exhibit heterogeneous respect for shared norms. We elicit individual norm-following propensity by asking subjects to choose from two actions (A or B), where action B is costly. The instructions state “the rule is to do B,” and willingness to incur the cost of B reveals respect for norms. Behavior in this task is correlated with norm-consistent behavior in dictator games, providing support for our interpretation.

*JEL classifications: C91, D03*

*Keywords: experimental economics, norms, prosocial behavior, social preferences*

---

\*Thanks to Mahsa Akbari and Nidhi Dahiya for excellent research assistance and to the Social Sciences and Humanities Research Council of Canada Insight Grants (435-2015-0798) and Insight Development Grants (430-2012-0146) programs for providing research funds. We received helpful comments from Cary Deck, Vernon Smith, and Bart Wilson. Thanks also to the Visiting Scholar’s Program in the Walton College of Business at the University of Arkansas. The experiments were programmed using zTree (Fischbacher, 2007). All remaining errors are our own.

<sup>†</sup>Corresponding Author: Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada, e-mail: ekimbrou@sfu.ca

<sup>‡</sup>Center for Mind/Brain Sciences, University of Trento, Via delle Regole 101, 38123 Mattarello (TN), Italy. e-mail: a.vostroknutov@unitn.it

# 1 Introduction

Extensive experimental and empirical evidence suggests that many individuals readily incur costs in order to cooperate, donate, reciprocate and punish others, often in opposition to the axiom of self-interested payoff maximization. Earlier attempts to account for such observations relied on the idea that subjects had ‘social preferences’ defined over payoff distributions, rather than own payoffs (see e.g. Fehr and Schmidt, 1999), but evidence suggests that such accounts are incomplete, as prosocial and other-regarding behavior (e.g. giving in dictator games, rejection in ultimatum games) is known to be sensitive to payoff-irrelevant manipulations such as who is watching and how decision rights are allocated (Hoffman *et al.*, 1994, 1996; Cherry *et al.*, 2002). Moreover, recent evidence suggests that when individuals can conceal their selfishness, they regularly take advantage of this “moral wiggle room” and behave more selfishly than they otherwise would (Dana *et al.*, 2007; Andreoni and Bernheim, 2009; Lazear *et al.*, 2012).

This sensitivity of prosocial behavior to context cannot be rationalized via simple models in which utilities are defined only over payoffs, but one theoretical approach that better fits the available evidence suggests that prosocial behavior is driven by a desire to adhere to social norms (López-Pérez, 2008; Kessler and Leider, 2012; Krupka and Weber, 2013; Cappelen *et al.*, 2007; Kimbrough and Vostroknutov, 2016). In such models, individuals evaluate the appropriateness of own behavior by comparing it to exogenously defined and commonly known injunctive social norms, that is, to a set of shared beliefs about the appropriateness of possible actions available in the setting.<sup>1</sup> Each possible action has both direct payoff consequences and a normative valence, and the simplest such models assume these enter additively in the utility function. Social norms are assumed to vary with context, and when norms and self-interest conflict, individuals face tradeoffs between following the norm and maximizing their payoffs. Crucially, individuals may differ in the degree to which they suffer from norm violations (or gain from norm adherence), and this can be captured in a single parameter that assigns a weight to the normative component of utility. In such models, heterogeneous behavior within a given game can be explained via heterogeneity in these weights; while, heterogeneous behavior across games, or across contexts, can be explained by the fact that norms vary with the setting.

A key implication of these models is that observed behavior in social settings (i.e. where norms of conduct apply) cannot offer direct evidence on preferences since behavior deviating from payoff maximization is a joint product of both norms and the weight assigned to them. One major concern with the norm-based approach is that we currently lack a solid theory of how and why norms vary with context, which introduces a potentially dubious source of “researcher

---

<sup>1</sup>Other- and self-signalling models have also been proposed as a possible unifying account of prosocial behavior in games. In these models agents wish to maintain their reputations and behave prosocially in order to do so (e.g. Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009). As argued in Kimbrough and Vostroknutov (2016) such approaches are complementary to a norms-based approach since the ability to know which actions will help or harm a reputation is predicated on shared normative beliefs (i.e. injunctive norms).

degrees of freedom” into the framework. To alleviate this it is important to be able to pin down the norm in a given setting *ex ante*. To this end, Krupka and Weber (2013) introduced a method of directly eliciting shared beliefs about social norms using incentivized coordination games and showed that context-driven changes in norms can account for observed changes in behavior in a set of dictator game experiments reported in List (2007). Adapting their methods appropriately, it is possible to empirically identify social norms in any game, and thus while we may not be able to say something about the sources of norms in general, we can pin the norm down in any particular application and thereby learn something about the sources of social behavior.<sup>2</sup>

A second issue is that these models assume a single, portable weight placed on the normative component of utility, which is also observationally confounded with norms in social settings.<sup>3</sup> One solution is to attempt to measure (a proxy for) this parameter directly using an experimental task. Kimbrough and Vostroknutov (2016) introduced such an individual decision task, called the rule-following task, designed to measure how much weight individuals place on social norms. In their task norm- (or rule-) following is pitted directly against material interest in a setting in which social preference models and payoff-maximizing models make the same predictions, namely that all individuals will break the rule. Using this task and separately eliciting norms using variants of the task due to Krupka and Weber (2013), the authors found that rule-following behavior was correlated with norm-consistent behavior in public goods, trust, ultimatum and dictator games; in a followup paper, they also found that this task was correlated with norm consistent behavior in dynamic common pool resource games (Kimbrough and Vostroknutov, 2015).

While this approach provides solid evidence in favor of the norms-based account of prosociality, the rule-following task employed contextual cues (traffic lights) which may have introduced noise into the measurement of subjects’ concern for social norms. Here we present a new, abstract task addressing these issues which can be widely used to measure the propensity to adhere to social norms. We have conducted a number of experiments that included this task in three countries (Canada, Netherlands, USA), and for a subset of the rule-following data collected in each country, we also subsequently collected data in a variant of the standard dictator game. The distribution of behavior in the rule-following task across these samples is quite similar, and the correlation between rule-following and norm-consistent behavior in the dictator game is large in magnitude and highly significant overall, though we find some differences between countries. Thus the task offers a simple and direct method of eliciting individual-level respect for social norms.

---

<sup>2</sup>See the examples in Kimbrough and Vostroknutov (2015, 2016). Recent work also suggests that their method is robust to asking subjects to identify the norm from the perspective of both second- and third-party roles in games as well to order-effects in within-subject designs (Erkut *et al.*, 2015; D’Adda *et al.*, 2015).

<sup>3</sup>With repeated observations of the same subject, making strong assumptions about the form of the utility function, it is possible to estimate this parameter from data. However, to our knowledge existing work has only attempted to estimate the population mean concern for norms.

## 2 The Task

### 2.1 The Rule Following Task

In the rule following task participants drag-and-drop balls one-by-one into two buckets: yellow or blue. For each ball in the yellow bucket they receive 10 cents and for each ball in the blue bucket they receive 5 cents (see Figure 1). The current earnings from the two buckets are shown above them. The total earnings in this task is the sum of earnings from the buckets. The position of the buckets on the screen is randomized across individuals.

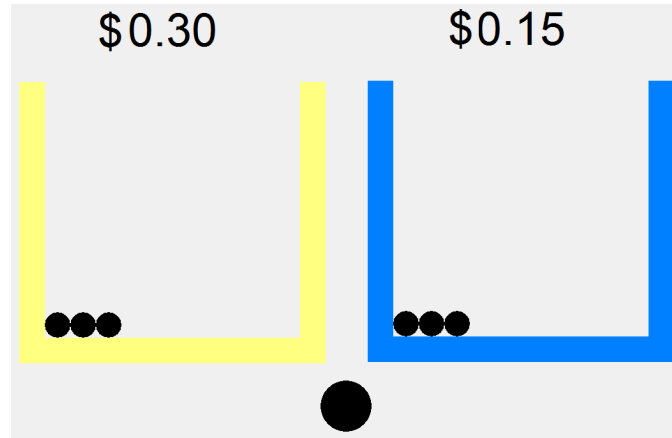


Figure 1: Presentation of the rule following task.

The instructions explicitly state that “*the rule is to put the balls into the blue bucket*” (see Appendix). Participants have a fixed number balls to allocate (say 100); thus, their earnings can vary from \$5, if they follow the rule to the letter, to \$10, if they break the rule with each ball.<sup>4</sup> Across a variety of experiments we have elicited rule-following task behavior from 600 subjects (44 NDL, 354 USA, 202 CAN). This was always the first task subjects performed.

### 2.2 The Dictator Game

In some experiments reported here, a Dictator game was played after the rule following task. A visually and structurally similar two buckets design was used (Figure 2). Subjects chose whether to put the balls in their “own” bucket or the bucket of “another player.” The difference between the Dutch version and the US/Canadian version of the Dictator game was that in the Dutch version the buckets were originally empty, and the subjects had to make a choice to which bucket to put each of the 100 balls. In the US version, the buckets were filled with an equal number of balls and subjects had to decide *for each ball* whether to keep it in the bucket where it began or

---

<sup>4</sup>When subjects ask for clarification about the statement that “*the rule is...*”, we always answered that “this is the rule of the experiment,” and when they ask whether anything will happen to them if they don’t follow the rule, we told them that all information they need to make their decision is in the instructions.

to move it to (“give to” or “take from”) another bucket. All subjects made choices as if they were dictators, but subjects were randomly paired and half of the decisions were implemented. Instructions are available in the Appendix. In total we have 180 dictator observations (44 NDL, 70 USA, 66 CAN); this task was always performed immediately after the rule-following task, so that we can test the model .

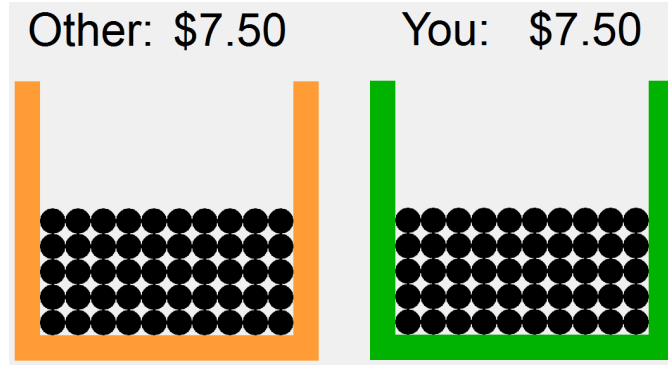


Figure 2: Presentation of the dictator game (USA and Canada).

### 3 How to Use the Rule Following Task

Following Kimbrough and Vostroknutov (2016), we use the cost incurred following the rule as a measure of an individual’s propensity to follow norms in other environments. To give an intuition, consider a norm-dependent utility of subject  $i$  in the rule-following task. Suppose that  $i$  gets linear consumption utility from money and incurs costs from not following the norm (put all the balls in the blue bucket). The utility can be written as follows:

$$U_i(x) = x + \phi_i g(x).$$

Here  $\phi_i \geq 0$  is the propensity of  $i$  to follow norms and  $g : \mathbb{R}_+ \rightarrow [-1, 1]$  is a function that assigns a normative *social appropriateness* (or inappropriateness) to each *action*. In case of the rule following task  $g(x)$  is decreasing in  $x$  (balls allocated to the higher return yellow bucket). Suppose that subject  $i$  maximizes her utility; then we would have the following equation  $\phi_i = -1/g'(x^*)$ , where  $x^*$  is the optimal choice (observed by us in the experiment). Given appropriate assumptions on the shape of  $g(x)$  (e.g., strictly concave) we obtain a positive monotonic relationship between unobserved  $\phi_i$  and observed  $x^*$ . In this sense we consider the observations in the rule following task as a proxy for  $\phi_i$ .

Consider a Dictator who keeps  $x$  dollars in the Dictator game and has norm-dependent utility

$$U_i(x) = u(x) + \phi_i g(x).$$

Again,  $g(x)$  reflects the *social appropriateness* of action  $x$ . The social appropriateness of each action can be directly measured using the norm-elicitation task introduced in Krupka and Weber (2013), which uses a coordination game to measure beliefs about norms. Figure 3 illustrates.

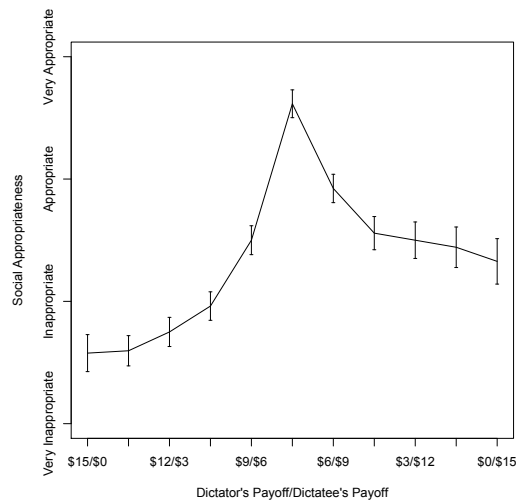


Figure 3: Elicited norms in the dictator game (USA data). Bars indicate  $\pm 1$  SE.

The data was generated by asking subjects ( $N = 52$ ) to assign an appropriateness rating to a set of possible actions by dictators (here we report data from Kimbrough *et al.*, 2014). One of these actions is chosen at random, and subjects are paid \$10 if their appropriateness rating matches the modal appropriateness rating and \$0 otherwise. This captures the notion of social norms as a set of shared beliefs about the appropriateness of actions in a given setting. Very socially inappropriate rating corresponds to function  $g$  equal to  $-1$  and very socially appropriate rating corresponds to  $g$  equal to  $1$ . One can see that giving half in the Dictator game is considered highly appropriate and giving nothing is thought highly inappropriate. The parameter  $\phi_i$  is measured by our rule following task. Thus, utility maximizers who have high parameter  $\phi$  should tend to choose closer to the half/half distribution and utility maximizers with low  $\phi$  should choose closer to keeping the entire pie.

## 4 Results

Now we test whether the above utility specification has merit. Our first result concerns the distributions of rule following choices in three countries. Figure 4 shows the respective distributions in the Netherlands, USA and Canada. All statistical tests are two-tailed.

The distributions show very similar patterns: there are peaks at putting almost all balls in the blue bucket (close to full rule following) and putting almost no balls in the blue bucket (close to full rule breaking). However, the distributions are not the same (KS test: exact  $p = 0.005$  between USA and Canada; exact  $p = 0.141$  between the Netherlands and Canada; exact  $p =$

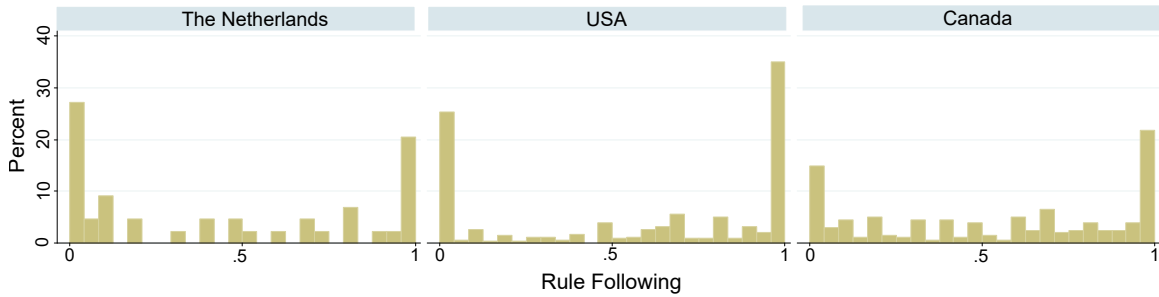


Figure 4: Rule following propensity (percentage of balls in blue bucket) in three countries. Number of observations: 44 in the Netherlands; 354 in USA; 202 in Canada.

0.079 between the Netherlands and USA). Although, all distributions have point masses at full rule following and full rule breaking, the heights of these peaks differ. For example, in the US sample, the percentage of people who choose to put almost all balls into the blue bucket is almost double that of the Dutch and Canadian samples. This observation is supported by a simple OLS regression analysis where (normalized) number of balls in a blue bucket is regressed on USA and Canada dummies (the Netherlands being a baseline). The coefficient on the USA dummy is 0.140 with  $p = 0.029$ , which signifies a 14% increase in rule following in the USA as compared to Canada or the Netherlands. These findings can potentially be explained by the demographic differences in the samples (see below).

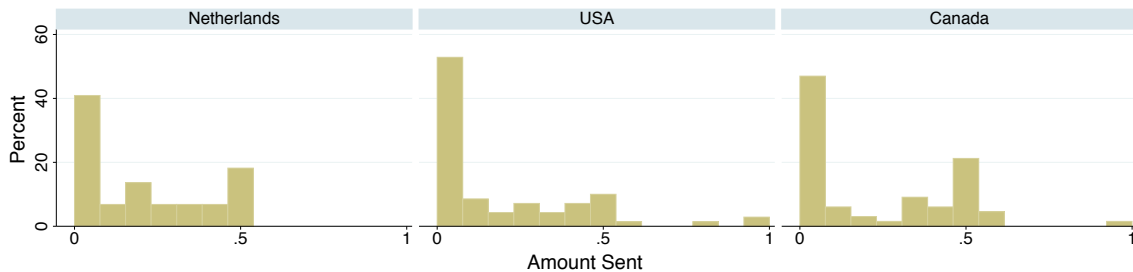


Figure 5: Dictator game giving (number of balls allocated to other) in three countries. Number of observations: 44 in the Netherlands; 70 in USA; 66 in Canada.

Next we report correlations between rule following task and Dictator game choices. Figure 5 displays the distribution of amount sent in the dictator game for each country. In all sessions that also contained a Dictator game, the rule following task was presented first. Pooling data from all three countries, Spearman's rank correlation coefficient between the amount sent to the recipient in the dictator game and the share of balls placed into the rule-following bucket is highly significant and positive ( $\rho = 0.31$ ,  $p < 0.0001$ ,  $N=180$ ). When we analyze the data within-country, we find a strong correlation between the behavior in the rule following task and the dictator game in the Netherlands and USA, but not in Canada. Spearman's correlation coefficient in the Netherlands is  $\rho = 0.42$ , ( $p = 0.006$ ,  $N = 44$ ); in the USA:  $\rho = 0.51$ , ( $p < 0.0001$ ,  $N = 70$ ) and in Canada:  $\rho = 0.06$ , ( $p = 0.62$ ,  $N = 66$ ). We discuss possible reasons for these differences below. OLS regressions of the amount *sent* in the Dictator game and rule following also supports

these findings (Table 1). Wald tests reveal that the coefficient on rule following is significant in the Netherlands ( $\beta = 0.152, p = 0.034$ ) and the USA ( $\beta = 0.248, p < 0.001$ ), but not in Canada ( $\beta = -0.002, p = 0.981$ ). Moreover, in the USA the subjects give slightly less overall.

<b>Amount sent in DG</b>		
	$\beta/se$	$\beta/se$
RF	0.148*** (0.042)	0.152** (0.071)
USA	-0.034 (0.038)	-0.093* (0.048)
USA $\times$ RF		0.095 (0.090)
CAN	0.003 (0.043)	0.097 (0.078)
CAN $\times$ RF		-0.155 (0.115)
Cons	0.867*** (0.033)	0.869*** (0.040)
<i>N</i>	180	180

Table 1: OLS regression of sharing in the dictator game. Errors are robust. \* –  $p < 0.1$ ; \*\* –  $p < 0.05$ ; \*\*\* –  $p < 0.01$ .

Finally, we explore the connection between our behavioral measures and various demographic characteristics. Table 2 shows summary statistics of the demographic data.<sup>5</sup> The samples are relatively similar in age, percent male and years of study, with the Dutch students having slightly fewer years at University on average. Also, though Economics and related fields prevail in all three countries, the US sample has fewer other Social Science students and somewhat more in the Arts and Humanities. The Dutch sample has the lowest share of students studying Natural Sciences. Ethnic composition varies widely across the samples, with the Dutch sample being mostly composed of Europeans, the American sample mostly composed of whites and self-reported “Americans”, and the Canadian sample divided roughly 1/3 to 2/3 between “Canadians” and Asians.

First we would like to know if observed differences in the distributions of rule following propensities stem from these demographic differences. Table A1 in Appendix A shows an OLS regression of the (normalized) number of balls in the blue bucket regressed on field of study, ethnicity, years of study, gender and age. The baseline of the regression is undeclared ethnicity with field of study being economics or related fields (see Table A2 in Appendix A for details). We see that relative to this baseline, subjects who study Art & Humanities tend to follow rules more (dependent variable has range from 0 to 1). Consistent with our observations above, subjects in the USA tend to be more rule following. Finally, consistent with our observations in Kimbrough and Vostroknutov (2016), males are more rule breaking.

<sup>5</sup>See Table A2 in Appendix A for the classification of fields of study and nationalities.



In trying to understand the lack of correlation between rule-following and dictator giving in our Canadian sample, two possible demographic factors stand out. First, there was a relatively larger share of students studying Natural Sciences, and while we have no strong reason to suspect an interaction with field of study, we explore this potential relationship by computing correlation coefficients between rule-following and dictator behavior within each field of study group. The correlation is positive for all but the “other social sciences” students, and is significant for economics/business students ( $\rho = 0.47, p < 0.001, N = 95$ ) and marginally significant, though smaller, for those in the natural sciences ( $\rho = 0.22, p = 0.11, N = 50$ ). This suggests that the lack of correlation in Canada was not driven by students in the natural sciences.

	Full RF Sample			DG Sample		
	NDL	USA	CAN	NDL	USA	CAN
Num. Subjects	44	354	202	44	70	66
Rule Following	44.5%	58.5%	54.6%	44.5%	60.3%	61.2%
Dictator Giving				19.9%	18.8%	22.7%
<b>Demographics</b>						
Males	52.3%	57.6%	54.4%	52.3%	58.6%	57.6%
Years of Study	2.11	2.76	2.85	2.11	2.63	2.70
Age	21.75	21.77	21.98	21.75	21.44	22.08
<b>Field of Study</b>						
Economics/Business	63.6%	58.2%	38.6%	63.6%	58.6%	39.4%
Social Sciences	15.9%	5.9%	18.8%	15.9%	10.0%	4.5%
Arts & Humanities	9.1%	13.0%	8.9%	9.1%	10.0%	9.1%
Natural Sciences	11.4%	22.6%	32.7%	11.4%	21.4%	45.5%
<b>Ethnicity</b>						
“American”		48.6%	2.0%		44.3%	
“Canadian”		0.3%	30.2%			22.3%
White/European	88.6%	24.6%	5.4%	88.6%	27.1%	4.5%
Hispanic		6.0%	2.0%		8.6%	3.0%
Black		7.3%	1.5%		7.1%	
Asian	11.4%	13.0%	57.9%	11.4%	12.9%	70.0%

Table 2: Summary statistics of behavioral and demographic data.

Second, many of the subjects were foreign-born and raised in non-Western cultures, such that cultural differences may influence behavior (e.g. by changing views about the norm, or changing some other aspect of how the decision is perceived). Thus we compute correlation coefficients between rule-following and dictator behavior for each ethnic group, and we find significant correlations among “American”, white, and hispanic subjects ( $\rho = 0.60, 0.41, \text{ and } 0.72$ , respectively, all  $p < 0.05, N = 31, 61, \text{ and } 8$ ), but not among the other groups (all  $p > 0.4$ ). For some of the ethnic background categories, this may merely reflect the small sample size, but there may also be important cultural differences that are correlated with ethnic background. In particular, the sizable Asian sample, which disproportionately comes from the Canadian data exhibits very low correlation between rule-following and dictator giving ( $\rho = 0.05, p = 0.68, N = 60$ ). Future research should test the robustness of these observations and attempt to account

for observed differences between ethnic/cultural backgrounds in the relationship between rule-following and dictator giving.

## 5 Discussion

In this paper we introduce a new method of eliciting individual concern for social norms (in a non-strategic setting), and we provide evidence that prosocial behavior in laboratory experiments is consistent with norm-dependent preferences: a set of preferences under which utility depends on a combination of own payoff and (heterogeneously, according to the parameter  $\phi_i$ ) how closely own action adheres to a normative standard. Intuitively, the role of social norms here is quite close to Adam Smith's notion of the 'impartial spectator', in whose eyes an agent wishes his actions to appear praiseworthy (or at least not blame-worthy); see Smith (1759).<sup>6</sup> As with standard models of preferences, agents exhibit non-satiation in own reward, but when social norms suggest non-payoff maximizing behavior (and when agents care about social norms), they will deviate from own-reward maximizing in the direction of the normatively appropriate action. Thus, such a model can be analyzed using standard tools, with choices in individual decisions and in games resulting from optimization by agents heterogeneously weighing the dictates of self-interest against the normative restraints imposed by sociality.

We argue that thinking about social interaction in terms of norm-dependence has a number of advantages for interpreting behavior relative to existing models. First as Charness and Rabin (2002)<sup>7</sup> and Wilson (2008, 2010) have pointed out, there are always implicit normative assumptions baked into typical social preference models. When agents are assumed to care about what is fair, we must ask how they know what constitutes fair behavior. These judgments are normative and are thus the products of social agreements and cultural factors; thus, "[the] rules that guide fair behavior are not located in an individual's private utility function but instead reside in the connections that the individual has to his cultural environs" (Wilson, 2012, p. 407). The modeler of a particular type of social preference implicitly imports a particular norm into the utility function. For instance, our preferred interpretation of models of inequity aversion is that they implicitly assume that in every decision-making context, the norm is to equalize payoffs. In contrast, norm-dependent utility, in which the normative appropriateness of actions (rather than outcomes) is incorporated into decision-making, makes these implicit normative assumptions explicit in the form of the norm.

A second key contrast between norm-dependent preferences and models in which prefer-

---

<sup>6</sup>"[A person] desires, not only praise, but praise-worthiness; or to be that thing which, though it should be praised by nobody, is, however, the natural and proper object of praise. He dreads, not only blame, but blame-worthiness; or to be that thing which, though it should be blamed by nobody, is, however, the natural and proper object of blame." (Smith, 1759, §. 3.1)

<sup>7</sup>As they put it in the context of reciprocity models, "any reciprocal model must embed assumptions about distributional preferences" (p. 824).

ences are defined over payoff distributions, then, is the former's flexibility, since norms are allowed to vary across contexts. The parameters of an inequity aversion model can allow for heterogeneity in the degree to which individuals care about different types of inequity, but they do not allow for concerns about equity to vanish in other settings (except when individuals are unable to directly influence outcomes). In our view, this helps explain why the assumption of inequity aversion is extremely successful in some contexts and less so in others: equity is often, but not always, considered appropriate.<sup>8</sup> Since norms are assumed to vary with context, a norm-dependent utility framework may thus be able to rationalize many observations that are inconsistent with models where utility is defined over payoff distributions, e.g. sensitivity of behavior to subtle manipulations of the choice set, as in Krupka and Weber (2013), to payoff-irrelevant changes in induced context Kimbrough *et al.* (2014), etc.

Of course, while our task provides a useful measure of individual respect for social norms and our findings suggest that norm-following preferences are an important primitive that shapes human sociality, our work does not provide an explanation of where these norms come from. Unfortunately, economists have not yet developed adequate models to allow us to predict the norm in a given setting, but recent experimental techniques allow us to measure norms in particular settings, perhaps eventually providing data that will encourage the development of a theory of norms. We suggest that future work should continue to explore the robustness of models of norm-dependence and attempt to better understand the origins and evolution of (injunctive) social norms.

## References

- ANDREONI, J. and BERNHEIM, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, **77** (5), 1607–1636.
- BÉNABOU, R. and TIROLE, J. (2006). Incentives and prosocial behavior. *The American Economic Review*, **96** (5), pp. 1652–1678.
- CAPPELEN, A. W., HOLE, A. D., SØRENSEN, E. Ø. and TUNGODDEN, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, **97** (3), 818–827.
- CHARNESS, G. B. and RABIN, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, **117** (3), 817–869.
- CHERRY, T. L., FRYKBLOM, P. and SHOGREN, J. F. (2002). Hardnose the dictator. *American Economic Review*, **92** (4), 1218–1221.
- D'ADDA, G., DROUVELIS, M. and NOSENZO, D. (2015). Norm elicitation in within-subject designs: Testing for order effects, cEDEX Working Paper.

---

<sup>8</sup>An obvious example is in sporting competitions where perceived attempts by third-parties such as referees to equalize outcomes arouse anger.

- DANA, J., WEBER, R. A. and KUANG, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, **33** (1), 67–80.
- ERKUT, H., NOSENZO, D. and SEFTON, M. (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters*, **130**, 28–31.
- FEHR, E. and SCHMIDT, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, **114** (3), 817–868.
- FISCHBACHER, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, **10** (2), 171–178.
- HOFFMAN, E., MCCABE, K., SHACHAT, K. and SMITH, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, **7** (3), 346–380.
- , — and SMITH, V. L. (1996). Social distance and other-regarding behavior in dictator games. *American Economic Review*, **86** (3), 653–60.
- KESSLER, J. B. and LEIDER, S. (2012). Norms and contracting. *Management Science*, **58** (1), 62–77.
- KIMBROUGH, E. O., MILLER, J. B. and VOSTROKNUTOV, A. (2014). Norms, frames and prosocial behavior in games, mimeo.
- and VOSTROKNUTOV, A. (2015). The social and ecological determinants of common pool resource sustainability. *Journal of Environmental Economics and Management*, **72**, 38–53.
- and — (2016). Norms make preferences social. *Journal of the European Economic Association*, **14** (3).
- KRUPKA, E. L. and WEBER, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, **11** (3), 495–524.
- LAZEAR, E. P., MALMENDIER, U. and WEBER, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, **4** (1), 136–163.
- LIST, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, **115** (3), 482–493.
- LÓPEZ-PÉREZ, R. (2008). Aversion to norm-breaking: A model. *Games and Economic Behavior*, **64** (1), 237–267.
- SMITH, A. (1759). *The Theory of Moral Sentiments*. Liberty Fund: Indianapolis (1982).
- WILSON, B. J. (2008). Language games of reciprocity. *Journal of Economic Behavior & Organization*, **68** (2), 365–377.
- (2010). Social preferences aren't preferences. *Journal of Economic Behavior & Organization*, **73** (1), 77–82.
- (2012). Contra private fairness. *American Journal of Economics and Sociology*, **71** (2), 407–435.

# Appendix (for online publication)

## A Demographics Regression

<b>Rule Following</b>	$\beta$	Std.Err.
Social Sciences	-0.043	(0.051)
Arts & Humanities	0.110**	(0.054)
Natural Sciences	0.005	(0.041)
“American”	0.163	(0.272)
“Canadian”	0.204	(0.269)
White	0.192	(0.273)
Hispanic	0.275	(0.279)
Black	0.118	(0.278)
Asian	0.164	(0.267)
Male	-0.083**	(0.033)
Years of Study	0.000	(0.011)
Age	0.001	(0.004)
USA	0.151**	(0.072)
Canada	0.114	(0.080)
Cons	0.278	(0.289)
<i>N</i>	600	

Table A1: OLS regression of the rule following propensity. Errors are robust. \* –  $p < 0.1$ ; \*\* –  $p < 0.05$ ; \*\*\* –  $p < 0.01$ .

<b>Field of Study</b>	<b>Definition</b>
Economics (baseline)	Economics, finance, business, international business, business economics, accounting, marketing, econometrics (and operations research), human decision sciences, supply chain management, banking, actuarial, economics and psychology, business and psychology, human resource management, public administration
Social Sciences	Psychology, sociology, cognitive science, hospitality management, anthropology, law, political science
Arts & Humanities	Arts, journalism, University College Maastricht, music, languages, european studies, architecture, sports studies, higher education, international relations, apparel studies, history, geography
Natural Sciences	Computer science, informational systems, biology, animal science, environmental sciences, engineering of all types, (bio)chemistry, physics, mathematics, biomedical research, health sciences, communication disorders, kinesiology, dietetics, nutrition studies, nursing
<b>Ethnicity</b>	<b>Definition</b>
N/A (baseline)	Not specified
“American”	Ethnicity specified as “American”
“Canadian”	Ethnicity specified as “Canadian”
White	All White Europeans including Spanish and Portuguese, Caucasian or white Americans and Canadians
Hispanic	All Spanish (Portuguese) speaking non-Europeans (Latin and Southern America)
Black	African Americans, African Canadians and Africans
Asian	All nations from Middle East and Asia and Native Americans

Table A2: Classification of fields of study and ethnicities.

## B Instructions

### B.1 RF Task Instructions

**General information** You are now participating in a decision making experiment. If you follow the instructions carefully, you can earn a considerable amount of money depending on your decisions and the decisions of the other participants. Your earnings will be paid to you in CASH at the end of the experiment

This set of instructions is for your private use only. During the experiment you are not allowed to communicate with anybody. In case of questions, please raise your hand. Then we will come to you and answer your questions privately. Any violation of this rule excludes you immediately from the experiment.

**Part I** In Part I of this experiment, you will decide how to allocate 100 [50] balls between two buckets.

Your task is to put each of the balls, one-by-one, into one of the two buckets: the blue bucket or the yellow bucket. The balls will appear in the center of your screen, and you can allocate each ball by clicking and dragging it to the bucket of your choice. For each ball you put in the blue bucket, you will receive 5 cents, and for each ball you put in the yellow bucket, you will receive 10 cents.

The rule is to put the balls in the blue bucket.

Once the experiment begins, you will have 10 minutes to put the balls into the buckets. When you are finished, please wait quietly until the end of the 10-minute period.

Your payment from Part I will be based on your decisions: it is the sum of payments from the blue and yellow buckets.

This is the end of the instructions for Part I. If you have any questions, please raise your hand and an experimenter will answer them privately. Otherwise, please wait quietly for the experiment to begin.

## B.2 DG Instructions (USA)

**Part II**

In Part II of this experiment, you are matched with one other person in this room. Each of you will decide how to allocate 100 balls between two buckets.

Your task is to put each of the balls, one-by-one, into one of the two buckets in front of you.

Each person has received 50 balls, each worth \$0.15, and they are in the buckets.

You can give balls to the other person by clicking and dragging from the green bucket to the orange bucket. You can keep balls in your (green) bucket by clicking them once.

You can take balls from the other person's bucket to your bucket by clicking and dragging them from the orange bucket into the green bucket. You can leave balls in the other person's (orange) bucket by clicking them once.


For each ball that ends up in the green bucket you get **15 cents**, and for each ball that ends up in the orange bucket, the other person keeps **15 cents**.

Once the experiment begins, you will have 10 minutes to put the balls into the buckets. When you are finished, please wait quietly until the end of the 10-minute period.

At that time, the computer will randomly choose one person from each pair whose decision is implemented. Your payment for Part II will be based on the randomly chosen person's decisions.

This is the end of the instructions for Part II. If you have any questions, please raise your hand and an experimenter will answer them privately. Otherwise, please wait quietly for the experiment to begin.

Click **OK** when you are ready to go on.



## B.3 Norm Elicitation Instructions (USA)

**Instructions 1**

In the following task you will be asked to evaluate the different situations and decide whether taking certain action would be "socially appropriate" and "consistent with moral or proper social behavior" or "socially inappropriate" and "inconsistent with moral or proper social behavior." By socially appropriate, we mean behavior that most people agree is the "correct" or "ethical" thing to do.


On the following screens, you will read descriptions of a series of scenarios. These descriptions correspond to situations in which one person, "Individual A" or "Individual B", must make a decision.

For each situation, you will be given a description of the decision faced by Individual A or B. This description will include several possible choices available to this Individual. After you read the description of the decision, you will be asked to evaluate the different possible choices available to Individual A and to decide, for each of the possible actions the level of social appropriateness or inappropriateness of that action. Another way to think about what we mean is that if Individual A or B were to select a socially inappropriate choice, then someone else might be angry at Individual A or B for doing so.

In each of your responses, we would like you to answer as truthfully as possible, based on your opinions of what constitutes socially appropriate or socially inappropriate behavior.

To give you an idea of how the experiment will proceed, we will go through an example and show you how you will indicate your responses. On the next screen you will see an example of a situation.

Click **OK** when you are ready to go on.





**Instructions 2**

In what follows, you will read a description of each scenario. Then, you will indicate your appropriateness rating by placing a check mark in the corresponding box.

At the end of the experiment today, the computer will randomly select one of the situations. For this situation, the computer will also randomly select one of the possible choices that the Individual could make. Thus, we will select both a situation and one possible choice at random. Each choice is equally likely to be selected.

For the choice selected, we will determine which response was selected by the most people here today.

**If you give the same response as that most frequently given by other people, then you will receive an additional \$10.** This amount will be paid to you, in cash, at the conclusion of the experiment.

For instance, if we were to select the example situation from the last screen and the possible choice "Leave the wallet where it is," and if your response had been "somewhat socially inappropriate," then you would receive \$10, in addition to the \$7 participation fee, if this was the response selected by most other people in today's session. Otherwise you would receive only the \$7 participation fee.

Your payment depends only on whether your response corresponds to the most common response to the chosen and DOES NOT depend on the decisions made by individual A or B.

Please click **OK** when you are ready to go on.

**If you have any questions, please raise your hand and wait for the experimenter.**

OK

**Example Situation**

Individual A is at a café. While there, Individual A notices that someone has left a wallet at one of the tables. Individual A must decide what to do. Individual A has four possible choices: take the wallet, ask others nearby if the wallet belongs to them, leave the wallet where it is, or give the wallet to the bartender.

Individual A can choose only one of these four options. The table on the right presents a list of the possible actions available to Individual A. For each of the actions, please indicate whether you believe choosing that option is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate. To indicate your response, please click on the corresponding cell.

Individual A's Action	Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate
Take the wallet				✓
Leave the wallet where it is			✓	
Give the wallet to the bartender		✓		
Ask others nearby if the wallet belongs to them	✓			

Please make sure you make an assessment for each possible choice in each row of the table.

In what follows, you will be asked to assess the appropriateness of the actions in two other scenarios. For each action please indicate the extent to which you believe taking that action would be "socially appropriate" and "consistent with moral or proper social behavior" or "socially inappropriate" and "inconsistent with moral or proper social behavior." By socially appropriate we mean behavior that **most people** agree is the "correct" or "ethical" thing to do.

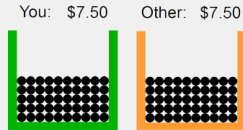
OK

**Scenario 1**

Individual A has been invited to an experiment and paired with another anonymous individual B so that neither individual will ever know the identity of the other individual with whom he/she is paired.

In the experiment, Individual A will make a choice, the experimenter will record this choice, and then both individuals will be informed of the choice and paid money based on the choice made by Individual A, as well as a small participation fee. Suppose that neither individual will receive any other money for participating in the experiment.

Individual A observes two buckets on the computer screen. One of the buckets is green and labeled "You" and one of the buckets is orange and labeled "Other". Each bucket contains 50 balls.



Individual A reads the following instructions: "Each person has received 50 balls, each worth \$0.15, and they are in the buckets. You can give balls to the other person by clicking and dragging from the green bucket to the orange bucket. You can keep balls in your (green) bucket by clicking them once. You can take balls from the other person's bucket to your bucket by clicking and dragging them from the orange bucket into the green bucket. You can leave balls in the other person's (orange) bucket by clicking them once. For each ball that ends up in the green bucket you get 15 cents, and for each ball that ends up in the orange bucket, the other person keeps 15 cents."

Now, look at the table on the right side of the screen and consider eleven possible actions that Individual A could take.

For each of the actions, please indicate whether you believe choosing that action is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate.

**Remember: when we select a scenario and an action for payment, you will only receive the additional \$10 if your response is the same as the most frequent response made by other people.**

Individual A's Action	Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate
Take 50 balls to the green bucket, leave 0 balls in the orange bucket: A gets \$15, B gets \$0			✓	
Take 40 balls to the green bucket, leave 10 balls in the orange bucket: A gets \$13.5, B gets \$1.5				✓
Take 30 balls to the green bucket, leave 20 balls in the orange bucket: A gets \$12, B gets \$3			✓	
Take 20 balls to the green bucket, leave 30 balls in the orange bucket: A gets \$10.5, B gets \$4.5		✓		
Take 10 balls to the green bucket, leave 40 balls in the orange bucket: A gets \$9, B gets \$6	✓			
Keep 50 balls in the green bucket, leave 50 balls in the orange bucket: A gets \$7.5, gets B \$7.5		✓		
Keep 40 balls in the green bucket, Give 10 balls to the orange bucket: A gets \$6, B gets \$9			✓	
Keep 30 balls in the green bucket, Give 20 balls to the orange bucket: A gets \$4.5, B gets \$10.5				✓
Keep 20 balls in the green bucket, Give 30 balls to the orange bucket: A gets \$3, B gets \$12			✓	
Keep 10 balls in the green bucket, Give 40 balls to the orange bucket: A gets \$1.5, B gets \$13.5		✓		
Keep 0 balls in the green bucket, Give 50 balls to the orange bucket: A gets \$0, B gets \$15	✓			

OK