Accounting for Context

Separating Monetary and (Uncertain) Social Incentives

Bergh, Andreas; Wichardt, Philipp C.

Published in:
Journal of Behavioral and Experimental Economics

DOI:
10.1016/j.socec.2017.11.002

2018

Document Version:
Peer reviewed version (aka post-print)

Accounting for Context: Separating Monetary and (Uncertain) Social Incentives*

Andreas Bergh†

Dept. of Economics, University of Lund;
The Research Institute of Industrial Economics (IFN), Stockholm

Philipp C. Wichardt‡

Kiel Institute for the World Economy; Dept. of Economics, University of Lund;
Dept. of Economics, University of Rostock; CESifo Munich

This Version: August 18, 2017

Abstract

This paper proposes a simple framework to model social preferences in a way that explicitly separates economic incentives from social (context) effects and allows for uncertainty also about the latter. Moreover, it allows non-economic cost associated with the deviation from some norm to be more discriminatory than just "right" or "wrong." We refer to existing evidence on dictator game giving to demonstrate how intermediate behaviours (giving some) as well as payments to change the context (e.g. exiting the game) can be accounted for. Furthermore, the framework is used to exemplify both theoretically and empirically how contextual variables such as social norms can worsen a social dilemma or possibly make it disappear. The empirical results of a classroom experiment suggest that women are more responsive to such contextual effects.

Keywords: Context Effects, Efficiency, Social Norms, Social Preferences, Utility

JEL codes: D03, D63, Z10

*We are grateful to Tore Ellingsen, Håkan J. Holm, Martin Kocher and Erik Wengström for helpful comments and discussions. Moreover, we have benefited considerably from the comments of two anonymous reviewers. Financial support from Swedish National Science Foundation (Bergh) and the Arne Ryde Foundation (Wichardt) is gratefully acknowledged. Wichardt thanks the Department of Economics at Lund University for its hospitality.

†Department of Economics, Lund University, PO Box 7082, SE-22007 Lund, Sweden; tel.: +46-(0)46-2224643 ; e-mail: andreas.bergh@ifn.se.

‡Department of Economics, University of Rostock, Ulmenstraße 69, D-18057 Rostock, Germany; tel.: +49-(0)381-4984486; e-mail: philipp.wichardt@uni-rostock.de.

1 Introduction

One of the fundamental tenets of economics is that people respond to incentives. Traditionally, the incentives are assumed to be material and related to the individual's own consumption. By now, however, the tendency of people to deviate from the predictions of simple self-centered utility-maximization, where utility is understood in terms of economic benefits, is well documented in the literature (e.g. Bowles and Gintis, 2011; Bolton et al., 2008; Gintis et al., 2005; Hoffman et al., 1994; Rabin, 1993; Camerer, 2003, provide a general review of various experimental results).

In response to these observations, a variety of models of social preferences have been proposed (see Sobel, 2005, for an illuminating review). In some of these, social preferences are modelled by adding a preference for the (monetary) utility of others (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). In a similar vein, albeit with a stronger focus on the impact of social norms, Fehr, Kirchsteiger and Riedl (1993) show that reciprocity can generate persistent noncompetitive outcomes in a competitive market (see also Fehr et al., 1998; or Dufwenberg and Kirchsteiger, 2004). Moreover, Lopez-Perez (2008) or Kranz (2010) analyse the effects on equilibrium outcomes if (some fraction of) agents experience a non-monetary disutility when deviating from some specific norm-prescribed behaviour.

However, a common drawback of these models is that they leave little room to account for specific context effects such as context-driven changes in preferences or uncertainty about the size or direction of social incentives.[1] Yet, empirical evidence strongly suggest that such effects have indeed a tangible influence on individual behaviour. Liberman et al. (2004) document large behavioral changes resulting from changing the name of a game from Wall street game to community game, despite payoffs being the same. Dana et al. (2006) provide evidence showing that potential 'dictators' are willing to forsake part of their potential benefits (1 of 10) in order to opt out of the interaction (see also Andreoni et al., 2017). In a similar vein, Frey and Bohnet (1995) find that in a dictator game social control, i.e. being identifiable as a dictator, leads to transfers shifting towards half of the pie (see also Engel, 2011, for a discussion). Moreover, Croson and Shang (2008) report that social references influence donations to a public radio station in the direction of the reference. And while there is abundant evidence for cooperation in social dilemmas (cf. Camerer, 2003) there is also evidence for people behaving perfectly rational – and selfish –

[1] The importance of finding useful models that allow preferences to vary with the context is also emphasised by Sobel (2005).

when playing a game of tennis (cf. Walker and Wooders, 2001).

Starting from this observation, we present a simple game theoretic framework in which utility is modelled as a combination of economic and social preferences both of which can be exposed to uncertainty. We explicitly allow for uncertainty about social incentives as we believe that certain patterns in behaviour, i.e. some intermediate or seemingly "safe" half-way choices, are a response to such uncertainty; we come back to this point in the discussion of our model in the Section 3. Yet, we do not directly incorporate the utility of others into the utility function of the agent. This choice was made for two reasons: (1) to retain the assumption of behaviour eventually being selfish, and (2) not to obscure the analysis with additional potentially problematic variables (such as interdependent utility).

Similar in spirit but focusing on extensive games without uncertainty about contextual incentives, Lopez-Perez (2008) and Kranz (2010) already demonstrate nicely how accounting for a non-monetary cost of norm deviant behaviour for some agents can change equilibrium outcomes.[2] The present paper adds to these approaches by allowing the cost to vary between different behaviours in a more fine-grained way, i.e. to identify more than "right" (no cost) or "wrong" (cost), and by emphasising the effects of a possible variance in the interpretation of a certain situation. In order to exemplify the benefits of our approach and to clarify the distinction to other approaches, we discuss how the model offers a way to account for various context effects regarding altruistic giving in the dictator game, such as the willingess to pay for not entering the game as a dictator (Dana et al., 2006), in Section 3.

Finally, we use the framework to demonstrate how social norms may also induce negative economic consequences, namely if the behaviour which is socially recommended for some instances happens to point towards economic inefficiency.[3] We also exemplify the possibility to analyse such *social failure*, as we will call it, with a small classroom experiment in which subjects face different decision situations – two Prisoner's Dilemmas with identical payoffs but reversed labels – after first having jointly agreed on some general recommendation for behaviour. Having done so, they are asked to play two versions of the Prisoner's Dilemma with identical payoffs but reversed labels. Thus, we create two situations: one in which the newly agreed norm[4] is

[2]For a discussion of why such a cost may indeed be evolutionary advantageous, see Wichardt (2011).

[3]When referring to *economic* efficiency, we refer to utility as generated from self-centered preferences focused on material outcomes.

[4]Referring to the agreed recommendation as a norm, of course, is optimistic. In fact, we would not expect (and do not find) its effect to be very strong. This notwithstanding, we do not expect

in line with social efficiency and one where it is just opposed. The results show that the creation of a weak social norm affects mainly women but that for them defection becomes indeed more likely once defection corresponds to the suggested behaviour. As a possible explanation for the gender effect, the data suggest a higer social sensitivity for women.

The rest of the paper is structured as follows: In Section 2, we introduce our model and provide some brief motivation for our modelling choices. A discussion of the model, including a comparison with existing models on fairness preferences and some illustrating examples, is given in Section 3. In Section 4, we report on a simple classroom experiment and its results. Section 5 concludes.

2 The Model

Consider a standard normal form game, G, given by a finite set of Players N, a finite set of strategies S_i and a utility function $u_i : \times_{i \in N} S_i \mapsto \mathbb{R}$ reflecting player i's preferences over outcomes for each player i, $i \in N$.

In addition, assume that prior to the play of the game Nature chooses the state of the world θ, with $\theta \in \mathcal{C} := \{\mathbb{E} = \mathbb{S}^0, \mathbb{S}^1, \ldots, \mathbb{S}^n\}$; $\theta = \mathbb{S}^k$ here can be thought of as indicating a certain (type of) social context,[5] whereas $\mathbb{E}$, henceforth referred to as $\mathbb{S}^0$, indicates a purely economic context. In order to make things interesting, assume that the exact type of context is not observable to the players and that only probabilities are. The probability of some state of the world $\mathbb{S}^k \in \mathcal{C}$ is given by p_k, $k = 0, \ldots, n$, and $\sum_{k=0}^{n} p_k = 1$.

Moreover, for each player i, $i \in N$, and each state of the world $\mathbb{S}^k$, $k = 0, \ldots, n$, let there be a partition $\mathbb{P}_i^k$ of player i's set of pure strategies S_i and a function $\phi_{i,G}^k : \mathbb{P}_i^k \to \mathbb{R}$.[6] Overall utility for player i, then, is given by

$$U_i(s_i, s_{-i}) = u_i(s_i, s_{-i}) + \sum_{k=0}^{n} p_k \phi_{i,G}^k(s_i)$$

for all $\mathbb{S}^k$ and all $i \in N$, where, slightly abusing notation, we define $\phi_{i,G}^k(s_i) := \phi_{i,G}^k(\tilde{S}_i)$

the wording here to trigger any misleading intuitions.

[5] When referring to a context, we of course mean classes of contexts such as "meeting colleagues" or "family." If the individuation of the context went any further, the framework would become tautological.

[6] For evidence on how contextual variations affect different subjects differently see, for example, Lönnqvist et al. (2009).

with $\tilde{S}_i \in \mathbb{P}_i^k$ such that $s_i \in S_i$.

As is customary, we use s_i and s_{-i} to denote the strategy of i and the strategy of agents other than i. Moreover, if $\theta = \mathbb{S}^0 = \mathbb{E}$, we assume $\phi_i^k(s_i) = 0$ for all $i \in N$ and all $s_i \in S_i$, i.e. in purely economic contexts only economic outcomes count.

Intuitively, one can think of $\phi_{i,G}^k(\tilde{S}_i)$ as the individual non-monetary rewards reflecting the socially desirability of strategies in the subset $\tilde{S}_i \subseteq S_i$ in context $\mathbb{S}^k$, e.g. in how far the respective behavior corresponds to or deviates from the social norm in the corresponding context. In fact, the context dependent partitioning of the strategy sets allows to potentially classify strategies into more than "right-wrong" categories, such as "entirely appropriate," "well...okay" and "far off the mark" and to assign cost and benefits accordingly (see the discussion of the dictator game in the next section). Moreover, we focus on pure strategies as we believe that what matters most for the conscience to be affected is actual behaviour and not plans I might have entertained.

Note, however, that the exact interpretation of the additional payoff is not crucial here. Positive aspects may include, for example, warm glow effects (Andreoni, 1990), while effects such as guilt (e.g. Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007), Identity concerns (Akerlof and Kranton, 2000, 2005; Wichardt, 2008) or cognitive dissonance (e.g. Wichardt, 2011) may hide behind the cost. Moreover, the scale of the parameters will, of course, depend on the game itself, e.g. whom one plays with or what the overall level of economic payoffs is; for notational convenience, the subscript indicating this dependence is dropped in the sequel, though.

3 Discussion and Examples

Having introduced the technical framework, we move on to a comparison with existing models and a discussion of some illustrating examples.

Comparison with other Approaches

As pointed out by Lopez-Perez (2008), the idea to capture social preferences in the modelling of the agent's utility function dates back at least to Edgeworth (1881) who already proposes to model individual utility as a weighted sum of the economic utility of all individuals. Later very influential models in a similar spirit have then been proposed, for example, by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). In this type of model, the idea is to account for pro-social behaviour by integrating the utility of others into the agents own utility function. As argued by Blanco et

al. (2011), these models tend to perform reasonably well when it comes to aggregate behaviour but are problematic on an individual level.

With respect to the present discussion, the most important point is that these models focus on general equality concerns and do not specifically account for the context as agents are assumed to care for others in a situation transcendent way. Once an agent is associated with a specific utility function, say expressing concerns for equality in the spirit of Fehr and Schmidt (1999), usually nothing is said about how this function may change with the context. Of course, the experienced researcher will supposedly have a reasonable guess about when the model is appropriate. But there is nothing in the model that accounts for such potential changes of context and, hence, effects which may be due to uncertainty about contextual aspects cannot be accounted for within the model.

The present framework, therefore, tries to broaden the perspective a bit and, hence, does not presuppose any general concern for others. Instead it requires additional (external) information about possible interpretations of the decision context and potentially relevant corresponding norms.[7] In certain instances these norms may well prescribe other-regarding behaviour in a way that could also be captured by the aforementioned models. In general, however, this will not be the case as the examples discussed below will clarify. Importantly, in some instances what is socially at stake may be uncertain. The primary intention of the present paper, therefore, is to propose a framework which allows to captuer such uncertainty.

A further strand of literature that deals with fairness concerns was started by Rabin (1993) who emphasised the importance of intentions ascribed to others. Later approaches extending the work by Rabin are, for example, Dufwenberg and Kirchsteiger (2004) or Falk and Fischbacher (2006) who provide models of reciprocal behaviour. While we do not deny the relevance of intentions or reciprocity concerns, which we believe to be nicely captured in this discussion, the focus of our argument is on the non-strategic aspects of social preferences in the sense of a general desire to abide by social norms (irrespective of the intentions of others, which of course will have additional influence). Yet, we believe that the ideas presented here, in particular the uncertainty about socially appropriate behaviour, can easily be adapted to arguments about reciprocity.

Closer to the present discussion, in fact, are those models which allow for (some) agents to respond not only to monetary incentives but to be also affected by some non-

[7]Note that also the standard models of inequity aversion have to be calibrated.

monetary "fine" in case they do not comply with some kind of norm (e.g. Lopez-Perez, 2008; Kranz, 2010; or Wichardt, 2011). More specifically, Lopez-Perez (2008) and Kranz (2010) provide illuminating discussions of equilibrium play in extensive form games once some agents are internally motivated to follow some social conventions.[8] Wichardt (2011), in turn, is concerned with the evolutionary benefit of having the ability to develop a guilty conscience in connection with social norms.

Again, what is important for the present discussion is that in these models there is no uncertainty about the context and what is socially desired. By contrast, the focus of the present discussion lies on motivating patterns in individual behaviour in response to changes in the social interpretation of the context or uncertainty about it. The discussion of the dictator game below shall illustrate this point.

The Dictator Game Revisited

Despite its simplicity, the dictator game – one player distributing a fixed amount of money between themselves and a second player – has been widely used to study altruistic giving (see, for example, Engel, 2011, for a review and a meta-analysis of the existing evidence; see also Camerer, 2003). A common finding for this game is that individual transfers have a peak at zero and a minor one at giving half (cf. Engel, 2011, p. 589). However, over the years, a variety of deviation from these common patterns have been observed. For example, as demonstrated by Oxoby and Spraggon (2008) inducing a feeling of entitlement for dictators (or receivers) shifts transfers towards the entitled player.[9] Moreover, Frey and Bohnet (1995) demonstrate that being identifiable as a dictator leads to transfers shifting towards 50%-50%, which is the common social sharing norm.[10] And, last but not least, as shown by Dana et al. (2006), many potential 'dictators' turn out to be willing to forsake some of their possible gain (1 of 10) in order to avoid the actual interaction (see also Andreoni et al., 2017).

In order to see how these results can (qualitatively) be accounted for in the present framework, consider a standard dictator game, i.e. a situation where one player has to decide how to allocate 10 Euro between themselves and a receiver such that only

[8] An interesting aspect of the model by Lopez-Perez (2008), which we abstract from in the present discussion, is that agents disutility from deviant behaviour depends on how many agents of a given group comply with the norm.

[9] Entitlement for the receiver is induced by letting them earn the money and allowing dictators to take some amount of it (Oxoby and Spraggon, 2008).

[10] There are, of course, many other interesting findings regarding altruistic giving in the dictator game. In order to make our point, however, we choose to focus on the ones stated.

divisions up to a full Euro are possible.[11] The strategy set of the dictator, then, is given by $S = \{0, 1, 2, \ldots, 10\}$.

Moreover, for the sake of argument, assume that standard outcome utility corresponds one-to-one to the monetary value received and that, whatever the social context, there is only one partition of possible strategies for which the assignment of non-monetary utility may vary with the context, though: $\mathbb{P} = \{S_1, ..., S_6\}$ with $S_1 = \{0\}, S_2 = \{1, 2\}, S_3 = \{3, 4\}, S_4 = \{5\}, S_5 = \{6, 7, 8, 9\}, S_6 = \{10\}$. Moreover, regarding the context, consider three possible scenarios:

1. $\mathbb{S}^0$ - purely economic. The situation is such that no social conventions have to be considered and $\phi^0_m = 0$, $m = 1, \ldots, 6$, where ϕ^0_m is shorthand for $\phi^0(S_m)$.

2. $\mathbb{S}^1$ - standard social, i.e. the convention is giving half, and the dictatorâ€™s social sensitivity reflects this in the following way: $\phi^1_1 = -6$, $\phi^1_2 = -3$, $\phi^1_3 = 0$, $\phi^1_4 = 3$, $\phi^1_5 = 3.5$, $\phi^1_6 = 4$. The motivation for these values could, for example, be social observability of the dictator (Frey and Bohnet, 1995).

3. $\mathbb{S}^2$ - the receiver is entitled to the money, e.g. as in Oxoby and Spraggon (2008), and the dictator's social sensitivity reflects this as follows: $\phi^2_1 = -20$, $\phi^2_2 = -15$, $\phi^2_3 = -10$, $\phi^2_4 = -5$, $\phi^2_5 = -4$, $\phi^2_6 = 0$.

Then, the following observations are immediate:

1. If $p^0 = 1$, i.e. $\mathbb{S}^0$, it is optimal for the agent to keep all.

2. If $p^1 = 1$, i.e. $\mathbb{S}^1$, it is optimal for the agent to keep 5.

3. If $p^2 = 1$, i.e. $\mathbb{S}^2$, it is optimal for the agent to give all.

Note, however, that if the agent becomes uncertain about the demands of the context, intermediate transfers become optimal. For example, given the above choice of parameters, it is optimal for a player who considers both the pure economic and the social context equally likely, i.e. $p^0 = p^1 = 0.5$, to give 1 (which leads to a payoff of 7.5). And giving 3 would actually be as good as giving nothing (both leading to a payoff of 7). Similarly, if the agent is uncertain about who is entitled, i.e. $p^0 = p^2 = 0.5$, giving 5 leads to an expected payoff of 2.5 which is the best the agent can do.

The specific results, of course, depend on the choice of parameters, e.g. the marginal utility of the additional Euro or the marginal cost of deviation further from what is

[11] 10 Euro are chosen for expositional purposes only. The subsequent argument can easily be adjusted to other amounts or currencies.

expected. In fact, allowing also for variation of the partition, it would be technically easy to motivate any other intermediate amount, although the plausibility of the assumed parameters could suffer.

The point of the example, however, are not the specific results. Rather, the example was chosen to clarify how different interpretation of a context can lead to different behaviours and, in particular, how uncertainty about the interpretation of the context can lead to intermediate behaviour. Casual evidence suggests that there often is a tendency to "go half way" once people become uncertain about what is expected. Of course, if what is expected is revealed in the end, half way is never optimal ex post. Yet, it may be ex ante and it may even stay so if the uncertainty remains ex post.

Result 1 *Uncertainty about the demands of the context can lead to giving of intermediate amounts in the dictator game.*

As final remark on the dictator game, let us return to the observations made by Dana et al. (2006) who found subjects willing to forsake some of the typical 10 dollar dictator endowment in order not to enter the actual interaction. While standard models of inequity averion have difficulties giving reasonable explanations for such behaviour, the present framework does not. The point to note is simply that if not giving any in the dictator game leads to a certain disutility, e.g. due to a gulity conscience,[12] while the non-monetary benefits of giving what seems socially acceptable do not balance with the loss in monetary utility, then paying to not enter the context will indeed be optimal.

Result 2 *For people who are affected by social aspects of the context but benefit only weakly from following social norms, it may be optimal to foresake some monetary benefits in order to not enter the respective contex. Referring to avoiding the context as* outsideoption*, this holds strictly if for all* $s_i \in S_i, s_{-i} \in S_{-i}$ *we have* $u_i(outsideoption) > U_i(s_i, s_{-i}) = u_i(s_i, s_{-i}) + \sum_{k=0}^{n} p_k \phi_{i,G}^k(s_i)$.

More specifically even and in line with the discussion by Andreoni et al. (2017), our model suggests that it will in fact be those who, ceteris paribus, experience comparatively low non-monetary benefits from giving who should be willing to pay to opt out.

The advantage of the present framework, as we see it, is that allows to account for such phenomena and to offer potential explanations for their occurrence without leaving the model at hand.

[12]Note that the exact interpretation of the non-monetary aspects is not crucial here.

Transforming a Social Dilemma

As a further illustrating example, which we will use to demonstrate how social conventions can occasionally lead to inefficient outcomes, consider a standard Prisoner's Dilemma game as depicted in Figure 1.

	C	D
C	$10, 10$	$0, 14$
D	$14, 0$	$4, 4$

Figure 1: A common Prisoner's Dilemma game.

Assume that if the context is social we have $\mathbb{P}_i^1 = \{\{C\}, \{D\}\}$, $i = 1, 2$, as well as $\phi_1(\{C\}) = \phi_2(\{C\}) =: \phi^+ > 0$ and $\phi_1(\{D\}) = \phi_2(\{D\}) = \phi^- < 0$. Then, if the context is social with certainty, i.e. $p_0 = 0$, and social incentives are sufficiently strong, i.e. $\phi^+ - \phi^- > 4$, C becomes the dominant action regardless of the other agent's action thereby making (C, C) the unique Nash equilibrium of the "social Prisoner's Dilemma." For $\phi^+ - \phi^- = 4$ both (C, C) and (D, D) are Nash equilibria. And, for a sufficiently low social sensitivity, $\phi^+ - \phi^- < 4$, (D, D) remains the unique Nash equilibrium.

Accordingly, a sufficiently strong incentive to follow the socially desired can transform the Prisoner's Dilemma into a situation where cooperation is individually rational – a line of argument which is often implicitly taken in models of social preferences but without referring to the context. Figure 2 illustrates this point.

$\phi^+ = \phi^- = 0$		
	C	D
C	$10, 10$	$0, 14$
D	$14, 0$	$\mathbf{4, 4}$

$\phi^+ = 4, \phi^- = 0$		
	C	D
C	$\mathbf{14, 14}$	$4, 10$
D	$14, 4$	$\mathbf{4, 4}$

$\phi^+ = 4, \phi^- = -4$		
	C	D
C	$\mathbf{14, 14}$	$4, 10$
D	$10, 4$	$0, 0$

Figure 2: Transforming the Prisoner's Dilemma when cooperation is socially demanded; Nash equilibria are marked in bold.

Finally, assume that $\phi^+ - \phi^- = 8$ but that there is some uncertainty as to whether the context is really social, in which case C is a strictly dominant action, or not, in which case D is strictly dominant. A straightforward calculation shows that already for $p_0 \leq 0.5$ both players playing C becomes a Nash equilibrium. Thus, even if the connotation of the context in question is uncertain – as might be the case for many lab

experiments – cooperation may be the dominant behaviour in the Prisoner's Dilemma; the only requirement to be met is that players (subjects) subjectively consider the situation to be sufficiently likely to be a social one in which cooperation is socially desired with $\phi^+ - \phi^- > 4$.

4 The Classroom Experiment

In this section, we present the results from a small classroom experiment to exemplify how we believe that the above effects can influence behaviour. In particular, the experiment was designed to demonstrate both the interplay of conflicting context effects and how the framework introduced above offers a comparably simple way to account for and study such effects.

Design and Procedures

Design

The experiment consisted of a brief introductory questionnaire asking subjects about some personal characteristics. After that subjects had to (simultaneously) decide on their behavior in two the Prisoner's Dilemma with identical payoffs but reversed labels (see Figure 3).[13]

	A	B
A	$100, 100$	$0, 140$
B	$140, 0$	$40, 40$

	A	B
A	$40, 40$	$140, 0$
B	$0, 140$	$100, 100$

Figure 3: The Prisoner's Dilemmas.

Before the questionnaire was handed out, all subjects were told that, once the questionnaire was finished, they would have to indicate how they would behave in some 2-by-2 games in which they could choose between A and B. In the treatment group, a weak social norm was created by having the participants first vote on a collective non-binding recommendation for choosing A or B, described as "potentially simplifying later decision making." Obedience with the norm, however, was neither enforced nor monitored in any way.

A few studies have found that the act of voting seems to create norms with large behavioral effects. For example, Alm et al. (1999) find that voting on enforcement

[13]The details of the questionnaire are available from the authors on request.

schemes can have large effects on tax compliance, even under identical fiscal regimes. Similarly, Markussen et al. (2014) study experimentally a collective action dilemma and find that adoption by voting enhances the efficacy of sanctions. The effect we are looking for here can be expected to be much weaker as subjects decide on nothing but a letter not knowing what is to come.

Procedures

The (classroom) experiment was conducted at the end of a lecture of a first year micro course at Lund University on September 24 2013. After half of the lecture, students were invited to take part in an decision experiment in which they could earn money. In all, 206 students participated in the experiment. Once those students who preferred to leave had done so (very few chose to leave), the participants were randomly assigned to two groups: a control group (42 subjects) and a treatment group (166 subjects) and were taken to different rooms. They were told that 20 out of all subjects would be randomly chosen to be matched with someone from their group and would be payed 1:1 for all games according to their behavior.

Subjects in both groups were first asked to fill in a brief questionnaire; in the treatment group, part of the questionnaire was to vote for either A or B as a general but non-binding recommendation for behavior. Once everything was filled in (including votes), questionnaires were collected and subjects and were presented with the questions about behaviour. Subjects in the treatment group were publicly informed about the result of the vote (numbers per option) before that. Eventually, subjects to be payed were chosen by a public random draw and privately payed after the experiment. The experiment lasted 45 minutes.

Results

Theoretical Argument

Before presenting the empirical results, let us briefly summarise the incentives which are likely to affect the situation:

1. Given the PD (payment-)structure, monetary incentives should induce defection.

2. Standard evidence from PD studies suggests that a substantial fraction of subjects see cooperation in social dilemmas as desirable even in (partly) anonymous environments.

3. The jointly agreed recommendation for behaviour may work as a weak social norm. In case it does, it should create uncertainty about what is appropriate

once it recommends defection; standard social norms would suggest cooperation but the agreement would suggest otherwise.[14]

Putting things into the proposed framework, we get the following: If subjects perceive the situation as purely economic, $\mathbb{S} = \mathbb{S}^0$, they should defect as $u_i(D, s_{-i}) > u_i(C, s_{-i})$ for all $s_{-i} \in \{C, D\}$. Yet, if subjects perceive the situation as possibly having a social component, they should assign positive probability, $p^1 > 0$, to the case where cooperation has some intrinsic value, $\mathbb{S} = \sim^1$ with $\phi_i^1(C) > 0$ and/or $\phi_i^1(D) < 0$, because common social norms would suggest so and the recommended behaviour does not conflict with this. Depending on the size of $p^1 \cdot (\phi_i^1(C) - \phi_i^1(D))$ and expectations about the likely behaviour of their opponent, this may suffice to induce cooperation.

Taking the case of the empirical example, we thus obtain the following expression for player i's expected utility from playing C and D, respectively:

$$EU_i(C) = \sigma_{-i}(C)u(100) + (1 - \sigma_{-i}(C))u(0) + p^1\phi_i^1(C)$$
$$EU_i(D) = \sigma_{-i}(C)u(140) + (1 - \sigma_{-i}(C))u(40) + p^1\phi_i^1(D)$$

where $\sigma_{-i}(x)$, $x = C, D)$, denotes the probability assigned to action x by the other player. Thus, the condition for cooperation to be the preferable action is given by

$$p^1(\phi_i^1(C) - \phi_i^1(D)) > \sigma_{-i}(C)[u(140) - u(100)] + (1 - \sigma_{-i}(C))[u(40) - u(0)]$$

This is essentially saying that, once the expected loss from defection in terms of social utility, $p^1(\phi_i^1(C) - \phi_i^1(D))$, is large enough, cooperation will be the dominant action. In case economic utility is linear in money, the equation is saying that the social incentive has to overcome the utility differential of 40 SEK for cooperation to become the dominant action, which exactly corresponds to intuition.

If subjects perceive the situation as having a social component but suggestions from common cooperative social norms and recommended behaviour conflict, there should be uncertainty regarding which recommendation to follow, i.e. with probability $p^1 > 0$ we have $\phi_i^1(C) > 0, \phi_i^1(D) < 0$ and with with probability $p^2 > 0$ we have $\phi_i^2(C) < 0, \phi_i^2(D) > 0$. All other things equal (and assuming $p^k, \phi_i^k(C), \phi_i^k(D)$ to be idiosyncratic), the additional chance of D being the contextually appropriate action should therefore decrease the aggregate tendency towards cooperation; the effect is

[14] A possible alternative would be to view the agreement as a coordination device. This would not contradict our argument, though, as we believe that most social norms in fact include a coordinating intention. If it was *only* the coordination aspect which was relevant here, it should affect behaviour also if it recommends cooperation. As we will see, this is not the case, though.

likely to be small, though, as the induced norm is rather weak. Note that, while the overall prediction might indeed have been natural to expect, the proposed framework provides a simple way to clarify the different aspects of the argument.

Empirical Results

Our sample consists of 206 subjects (mean age 21.7; std.dev. 2.4; 47% women). 164 of our subjects were randomly selected and treated with the voting procedure. The vote (anonymous, using paper sheets) resulted in a 65 percent majority voting for A over B. The remaining 42 subjects were assigned to the control group, that did not vote.

As we show below, the empirical evidence of our classroom experiment is essentially in line with the above argument.

Looking at raw data, the rate of cooperation is in all cases very close to 40 percent (39.9, 40.5 and 42.9) except when the norm suggest defection, where cooperation drops to 34

VARIABLES	Probability	of	Defection
	(1)	(2)	(3)
Treatment	0.0802	0.0853	-0.0219
	(0.95)	(1.00)	(-0.25)
Female		-0.0362	-0.267*
		(-0.52)	(1.72)
Treatment*Female			0.289*
			(-1.66)
Constant	0.571***	0.584***	0.667***
	(7.67)	(7.42)	(7.19)
Observations	197	194	194

Table 1: Determinants of defective behaviour in the Prisoner's Dilemma when the agreed recommendation in the treatment coincides with defection; OLS regression analysis. t-statistics in parentheses; *** $p < .01$, ** $p < .05$, * $p < .10$.

The analysis shows that the treatment has a statistically significant effect on women but not on men. In particular, women are in general more cooperative but also more inclined to abandon their cooperativeness once the (weak) social norm suggests defection. In accordance with the above theoretical argument and in line with the findings of Ellingsen et al. (2013), we interpret theses findings as suggesting that women are more responsive to social incentives (cooperation in the PD) and, hence,

also the treatment that manipulates the social incentives by creating a weak artificial norm. In fact, further analysis shows that women report to care more about being liked by others.[15]

Thus, the results of the classroom experiment show that socially desirable cooperation may even decrease – here for women – once the context is manipulated in a way that provides some external social justification / benefit for actually defecting.[16] Put differently, social recommendations, be it norms or otherwise, need not be in line with efficiency to be effective. Moreover, the technical framework presented above offers a simple way to account for the different aspects of such effects.

5 Concluding Remarks

In the present paper, we have presented a simple framework to model individual behaviour once both economic and context dependent social incentives are at play. Extending ideas presented in earlier models (e.g. Lopez-Perez, 2008; or Kranz, 2010), we allow for contextual factors to be uncertain, too, and for the cost of norm-deviant behaviour to be more discriminatory between strategies. As we have shown, the combination of both effects can be used to motivate intermediate behaviour in dictator games (giving more than nothing but less than half – what would be the social norm) as well as instances where potential 'dictators' forsake some their prospective endowment in order not to enter the interaction. Finally, we have presented the results of a small classroom experiment designed to illustrated the general type of argument captured by our model.

All in all, we believe that the proposed framework offers and interesting additional perspective on social interaction. Of course, it is no news that social norms influence behaviour in a tangible way. And that changes in contextual factors can lead to different norms being salient also is little surprising. In fact, the importance of finding models that also for preferences to change with the context is already emphasised by Sobel (2005) in his review of models of interdependent preferences and reciprocity. However, so far attampts to capture such arguments formally in order to make them more amenable to further study are scant. With the present model, we therefore hope

[15]Results are available from the authors on request.

[16]Social references have previously been shown to affect behaviour. For example, DellaVigna et al. (2012) find that social pressure is an important determinant of door to door giving. Also, Regner (2015) studied a pay what you want online music store and found that those inclined to follow social norms were more likely to pay the recommended price (despite being allowed to pay less).

to tak a first step in this direction.

It should be clear from the present discussion, though, that the present model is not primarily designed to to make clear cut predictions. As we see it, too little still is know about how people trade off contextual (social) incentives with context-transcendent economic ones in order to design a powerful general model of economic behaviour. Yet, we believe that it is valuable also to have a framework which enables to structure ex post discussions about what might have driven behaviour in instances where obviously more was at stake than just monetary incentives and where this "more" is obviously highly context dependent. More generally, we are convinced that what is important to understand eventually is not how one type of incentive works but how different types of incentives – here contextual social and economic ones – interact. While certainly only a step in that direction, we hope that the arguments provided within the present discussion can help to "identify general properties of extended preferences" as suggested to be important by Sobel (2005, p. 432) and thereby shedfurther light onto some still darker parts of this puzzle.

References

Akerlof, G. and R. Kranton, 2000. "Economics and Identity." Quarterly Journal of Economics, 1115, 715-753.

Akerlof, G. and R. Kranton, 2005. "Identity and the Economics of Organizations." Journal of Economic Perspectives 16, 9-32.

Alm, J., G. McClelland and W. Schulze, 1999. "Changing the Social Norm of Tax Compliance by Voting." Kyklos 52, 141-172.

Andreoni, J. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm Glow Giving." Economic Journal, 100(401), 464-477.

Andreoni, J., J. Rao and H. Trachtman, 2017. "Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving." Journal of Political Economy, 125, 625-653.

Battigalli, P. and M. Dufwenberg, 2007. "Guilt in Games." American Economic Review 97: 170-176.

Blanco, M., D. Engelmann and H. T. Normann, 2011. "A Within-Subject Analysis of Other-Regarding Preferences." Games and Economic Behavior, 72, 321-338

Bolton, G. E. and A. Ockenfels, 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." American Economic Review 90(1):166-193.

Bolton, G. E., J. Brandts, E. Katok, A. Ockenfels and R. Zwick, 2008. "Testing Theories of Other-Regarding Behavior: A Sequence of Four Laboratory Studies." In Plott, C. R. and V. L. Smith (editors), Handbook of Experimental Economics Results. Elsevier.

Bowles, S. and H. Gintis, 2011. "A Cooperative Species." Princeton University Press.

Camerer, C. 2003. "Behavioral Game Theory." Princeton University Press, Princeton, New Jersey.

Charness, G. and M. Dufwenberg, 2006. "Promises and Partnership." Econometrica 74: 1579-1601.

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

Dana, J., D. Cain 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, 193-201.

DellaVigna, S., J. A. List, and U. Malmendier, 2012. "Testing for Altruism and Social Pressure in Charitable Giving." Quarterly Journal of Economics 127, 1-56.

Dufwenberg, M. and G. Kirchsteiger, 2004. "A theory of sequential reciprocity." Games and Economic Behavior 47, 268-298.

Edgeworth, F. Y., 1881. "Mathematical psychics: A essay on the application of mathematics to the moral sciences." Kegan Paul, London.

Ellingsen, T., M. Johannesson, J. Mollerstrom and S. Munkhammar, 2013. "Gender Differences in Social Framing Effects." Economics Letters 118, 470-472.

Engel, C. (2011). "Dictator Games: A Meta Study." Experimental Economics 14:583-610.

Falk, A. and U. Fischbacher, 2006. "A theory of reciprocity." Games and Economic Behavior 54:293-315.

Fehr, E., E. Kirchler, A. Weichbold and S. Gächter, 1998. When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets. Journal of Labor Economics 16(2), 324-51.

Fehr, E., G. Kirchsteiger and A. Riedl, 1993 "Does Fairness prevent Market Clearing? An Experimental Investigation." Quarterly Journal of Economics 108, 437-60.

Fehr, E. and K. M. Schmidt, 1999. "A theory of fairness, competition and cooperation." Quarterly Journal of Economics, 114: 817-868.

Frey, B.and I. Bohnet, 1995. "Institutions affect fairness." Journal of Institutional and Theoretical Economics 151, 286-303.

Gintis, H., S. Bowles, R. Boyd, and E. Fehr, 2005. "Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life." MIT Press.

Hoffman, E., K. McCabe, K. Shachat, and V. L. Smith, 1994. "Preferences, property rights, and anonymity in bargaining games." Games and Economic Behavior, 7, 346-380.

Kranz, S, 2010. "Moral norms in a partly compliant society." Games and Economic Behavior, 68, 255-274.

Liberman, V. S. M. Samuels, and L. Ross, 2004. "The Name of the Game: Predictive Power of Reputations versus Situational Labels in Determining Prisoner's Dilemma Game Moves." Personality and Social Psychology Bulletin, 30, 1175-1185.

Lopez-Perez, R. 2008. "Aversion to norm-breaking: A model." Games and Economic Behavior, 64, 237-267.

Lönnqvist, J.-E., G. Walkowitz, M. Lindeman, M. and M. Verkasalo, 2009. "The Moderating Effect of Conformism Values on the Relations between Other Personal Values, Social Norms, Moral Obligation, and Single Altruistic Behaviors." British Journal of Social Psychology, 48, 525-546.

Markussen, T., L. Putterman, J.-R. Tyran, 2014. "Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes. Review of Economic Studies" 81, 301-24.

Oxoby, R. J. and J. Spraggon 2008. "Mine and yours: Property rights in dictator games." Journal of Economic Behavior and Organization, 65, 703-713.

Rabin, M. 1993. "Incorporating fairness into game theory and economics." The American Economic Review, 83, 1281-1302.

Regner, T. 2015. "Why Consumers Pay Voluntarily: Evidence from Online Music." Journal of Behavioral and Experimental Economics, 57, 205-214.

Sobel, J. 2005. "Interdependent Preferences and Reciprocity." Journal of Economic Literature 43, 392-436.

Walker, M. and J. Wooders, 2001. "Minmax Play at Wimbledon." The American Economic Review 91, 1521-1538.

Wichardt, P. 2008. "Identity and Why We Cooperate With Those We Do." Journal of Economic Psychology 29, 129-137.

Wichardt, P. 2011. "Identity, Utility, and Cooperative Behaviour: An Evolutionary Perspective." Scandinavian Journal of Economics, 113, 418-443.