Kiel
Working Papers

Kiel Institute
for the World Economy

# Accounting for Context: Separating Monetary and Social Incentives

by

Andres Bergh and Philipp Wichardt

**No. 1971**| November 2014

Web: www.ifw-kiel.de

Kiel Working Paper No. 1971| November 2014

# Accounting for Context: Separating Monetary and Social Incentives*

Andreas Bergh[a]
Dept. of Economics, University of Lund;
The Research Institute for Industrial Economics (IFN), Stockholm

Philipp C. Wichardt[b]
Kiel Institute for the World Economy; Dept. of Economics, University of Lund;
Dept. of Economics, University of Rostock; CESifo Munich

This Version: October 7, 2014

## Abstract

This paper proposes a simple framework to model social preferences in a game theoretic framework which explicitly separates economic incentives from social (context) effects. It is argued that such a perspective makes it easier to analyse contextual effects. Moreover, 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 randomised controlled classroom experiment show that women are more responsive to such contextual effects and that social agreements can also worsen economic inefficiencies.

**Keywords:** Context Effects, Efficiency, Social Norms, Social Preferences, Utility
**JEL:** D03, D63, Z10

*We are grateful to Tore Ellingsen, Håkan J. Holm, Martin Kocher and Erik Wengström for helpful comments and discussions. 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.
[a]Department of Economics, Lund University, PO Box 7082, SE-22007 Lund, Sweden; tel.: +46-(0)46-2224643; e-mail: drbergh@gmail.com.
[b]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).

In response to these observations, a variety of models of social preferences have been proposed in which social preferences are modelled by adding a preference for the (monetary) utility of others (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). A common drawback of these models is that they leave little room to account for contextual effects which could, for example, explain why people may sometimes cooperate in social dilemmas but still behave perfectly rational – and selfish – when playing a game of tennis (cf. Walker and Wooders, 2001).

Starting from this observation, we present a simple framework which allows us to model utility as a combination of economic and social preferences in a game theoretic setting. In particular, we separate economic incentives from potential social (context) effects and avoid to directly incorporate the utility of others into the utility function of an agent. The latter choice is 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). As we will demonstrate, the framework also allows us to analyse the possibly negative economic consequences of social norms which may arise if the behaviour which is socially recommended happens to point towards economic inefficiency.[1]

Moreover, we exemplify the possibility to analyse such *social failure*, as we will call it, with a small classroom experiment. In the experiment, subjects first have to agree 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[2] is in line with social efficiency and one where it is just opposed. The results show that once

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

[2]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 the wording here to trigger any misleading intuitions.

defection corresponds to the suggested behaviour, behaviour indeed becomes more defective and, hence, less efficient.

# 2 The Model

In the sequel, we introduce a simple formal framework to model social concerns in strategic interaction and provide an illustrating example.

### Technical Aspects

Consider a standard normal form game given by a finite set of Players $N$, as well as 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 $\theta$, with $\theta \in \{\mathbb{E}, \mathbb{S}^1, \ldots, \mathbb{S}^n\}$; $\theta = \mathbb{S}^k$ here can be thought of as indicating a certain (type of) social context,[3] whereas $\mathbb{E}$ indicates a purely economic one. The probability of $\mathbb{E}$ is given by $p_0$ and the probability of $\mathbb{S}^k$ is given by $p_k$, $k = 1, \ldots, n$, and $\sum_{k=0}^{n} p_k = 1$. Moreover, for each player $i$, $i \in N$, and each state of the world $\mathbb{S}^k$, $k = 1, \ldots, n$, let there be a distinguished subset of pure strategies $\hat{S}_i^k \subsetneq S_i$, $\hat{S}_i^k \neq \emptyset$, such that if $\theta = \mathbb{S}^k$ utility for player $i$ is given by

$$U_i(s_i, s_{-i}) = \begin{cases} u_i(s_i, s_{-i}) - \phi_i(\mathbb{S}^k) & \text{if } s_i \notin \hat{S}_i^k \\ u_i(s_i, s_{-i}) + \xi_i(\mathbb{S}^k) & \text{if } s_i \in \hat{S}_i^k \end{cases}$$

with $\xi_i(\mathbb{S}^k), \phi_i(\mathbb{S}^k) \in \mathbb{R}_0^+$ for all $\mathbb{S}^k$ and all $i \in N$; if $\theta = \mathbb{E}$ this would correspond to the case $\xi_i(\mathbb{E}) = \phi_i(\mathbb{E}) = 0$.

Intuitively, one can think of $\hat{S}_i^k$ as the socially desired or accepted behavior, e.g. the behavior prescribed by some social norm, in the corresponding context. Thus, if the context is social in nature, players obtain a, possibly idiosyncratic, additional social payoff $\xi_i$ from complying to what is socially desired and a, possibly idiosyncratic, additional cost $\phi_i$ from not doing so. Positive aspects may, for example, cover a warm glow of giving (Andreoni, 1990), while effects such as guilt (e.g. Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007) or cognitive dissonance (e.g. Wichardt et al., 2011) may hide behind the cost.

[3] 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.

**Example – Transforming a Social Dilemma**

As an illustrating example, consider the a standard Prisoner's Dilemma game as depicted in Figure 1.

| | C | D |
|---|---|---|
| C | $9, 9$ | $0, 10$ |
| D | $10, 0$ | $4, 4$ |

Figure 1: A common Prisoner's Dilemma game.

Moreover, assume that the context is social and that $\xi_i(C), \phi_i(C) > 0$, $\hat{S}_1 = \hat{S}_2 = \{C\}$ and $p_0 = 0$, e.g. because players are observed by peers who think that $C$ is the nicer action. Also, assume that $\xi_1 = \xi_2 = \xi$ and $\phi_1 = \phi_2 = \phi$. Then, if the social incentives $\xi + \phi$ are sufficiently strong, i.e. $\xi + \phi > 4$, $C$ 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 $\xi + \phi \in [1, 4]$ both $(C, C)$ and $(D, D)$ are Nash equilibria. And, for a sufficiently low social sensitivity, $\xi + \phi < 1$, $(D, D)$ 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.

| $\xi = \phi = 0$ | | |
|---|---|---|
| | C | D |
| C | $9, 9$ | $0, 10$ |
| D | $10, 0$ | $\mathbf{4, 4}$ |

| $\xi = 3, \phi = 0$ | | |
|---|---|---|
| | C | D |
| C | $\mathbf{12, 12}$ | $3, 10$ |
| D | $10, 3$ | $\mathbf{4, 4}$ |

| $\xi = 3, \phi = 3$ | | |
|---|---|---|
| | C | D |
| C | $\mathbf{12, 12}$ | $3, 7$ |
| D | $7, 3$ | $1, 1$ |

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

Finally, assume that $\xi + \phi = 5$ 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 $p_0 \leq 0.8$ 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 desirable, $\xi + \phi > 4$.

# 3 The Classroom Experiment

In this section, we present the results from a small classroom experiment with a randomized control group. The experiment was designed so as 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).[4]

| | 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.

*Procedures*

The (classroom) experiment was conducted at the end of a lecture of a first year micro

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

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 (41 subjects) and a treatment group (165 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 before that. Once everyone had made their four decisions, answer sheets were collected and the 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.

Putting things into the proposed framework, we get the following: If subjects perceive the situation as purely economic, they should defect as $u_i(D, s_{-i}) > u_i(C, s_{-i})$ for all $s_{-i} \in \{C, D\}$. If subjects perceive the situation as possibly having a social component, they should assign positive probability, $\rho^C > 0$, to the case where cooperation

has some intrinsic value, $\xi_i(C), \phi_i(C) > 0$,[5] because common social norms would suggest so and provided recommended behaviour does not conflict with this. Depending on the size of $\rho^C \cdot (\xi_i(C) + \phi_i(C))$ 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:

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

Thus, the condition for cooperation to be the preferable action is given by

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

This is essentially saying that, once the expected loss from defection in terms of social utility, $\rho^C(\xi_i(C) + \phi_i(C))$, 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 $\rho^C > 0$ we have $\xi_i(C), \phi(C) > 0$ and with with probability $\rho^D > 0$ we have $\xi_i(D), \phi(D) > 0$. All other things equal (and assuming $\rho^C, \xi_i(D), \phi(D)$ to be idiosyncratic), the additional chance of $D$ being the contextually appropriate action should therefore decrease the aggregate tendency towards cooperation; the effect is 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*

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

In order to test for the effects of manipulating the context on behaviour, we analyse a standard linear regression model with defection as dependent variable. The results

[5]Slightly abusing notation, we here refer to the type of context by the respectively desired behaviour, i.e. $\xi_i(C)$ ($\phi_i(C)$) is the benefit (cost) from (not) acting as desired.

for the case in which the social norm recommended defection are summarised in Table 1; if recommended behaviour is in line with common social norms, we do not find any statistically significant effects.

| VARIABLES | Defection |
|---|---|
| treatment | -0.0219 |
| | (0.108) |
| Female | -0.267* |
| | (0.155) |
| Treatment*Female | 0.289* |
| | (0.173) |
| Constant | 0.667*** |
| | (0.0925) |
| Observations | 196 |
| R-squared | 0.021 |

Table 1: Determinants of defective behaviour in the Prisoner's Dilemma when the agreed recommendation in the treatment coincides with defection; OLS regression analysis. Standard errors 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.

Thus, the results of the classroom experiment show that, socially desirable cooperation decreases once the context is manipulated in a way that provides some external justification / benefit for actually defecting. 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.

# References


**Andreoni,** J. (1990). Impure altruism and donations to public goods: A theory of warm glow giving. *Economic Journal*, 100(401):464–477.

**Battigalli,** P. and Dufwenberg, M. (2007). Guilt in games. *American Economic Review* 97: 170-176.

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

**Bolton,** G. E., Brandts, J., Katok, E., Ockenfels, A., and Zwick, R. (2008). Testing theories of other-regarding behavior: A sequence of four laboratory studies. In Charles, R. P. and Vernon, L. S., editors, *Handbook of Experimental Economics Results*. Elsevier.

**Bowles,** S. and Gintis, H. (2011). *A cooperative species*. Princeton University Press.

**Charness,** G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica* 74: 1579-1601.

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

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

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

**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.

**Rabin,** M. (1993). Incorporating fairness into game theory andeconomics. *The American Economic Review*, 83:1281–1302.

**Walker,** M. and Wooders, J. (2001). Minmax Play at Wimbledon. *The American Economic Review* 91(5):1521-1538.

**Wichardt,** P. C. (2011). Identity, Utility, and Cooperative Behaviour: An Evolutionary Perspective. *Scandinavian Journal of Economics*, 113(2):418–443.