Personal norms — and not only social norms — shape economic behavior*

Zvonimir Bašić and Eugenio Verrina†

November 8, 2021

Abstract

While social norms have received great attention as an explanation of various economic behaviors, little is known about personal norms. We propose a simple utility framework and design a novel two-part experiment to study the relevance of personal norms across various economic games and settings. We show that personal norms — together with social norms and monetary payoff — are highly predictive of individuals' behavior. Moreover, they are: *i)* distinct from social norms across a series of economic contexts, *ii)* robust to an exogenous increase in the salience of social norms, and *iii)* complementary to social norms in predicting behavior. Our findings support personal norms as a key driver of economic behavior.

Keywords: Personal norms, social norms, social image, elicitation method, normative conflict

JEL Classification: C91, D01, D63, D64, D91

*We thank Peter Andre, Valerio Capraro, Gary Charness, Eugen Dimant, Christoph Engel, Fabio Galeotti, Johannes Haushofer, Matthias Heinz, Adrian Hillenbrand, Lukas Kiessling, Nathan Maddix, Frédéric Moisan, Daniele Nosenzo, Sebastian Schneider, Matthias Sutter, Alexander Vostroknutov, Roberto Weber, and various seminar and conference audiences for helpful comments; and Viet A Nguyen for excellent research support. Financial support through the Max Planck Institute for Research on Collective Goods and the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. The main experiment was conducted at the Cologne Laboratory for Economic Research (CLER) of the University of Cologne, and it followed the IRB block approval (Application No. 2018_3) granted by the Ethics Council of the Max Planck Society. The additional experiment was conducted at the BonnEconLab of the University of Bonn, and it followed the ethical guidelines for study procedures from the BonnEconLab. The authors declare no conflict of interest.

†Zvonimir Bašić: Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, 53113 Bonn, Germany; E-mail: basic@coll.mpg.de. Eugenio Verrina: Groupe d'Analyse et de Théorie Economique (GATE), UMR5824, Univ Lyon, CNRS, F-69130 Ecully, France; E-mail: verrina@gate.cnrs.fr.

1 Introduction

For decades, social norms have been used as a central assumption of economic models, and have helped to explain a large variety of phenomena, such as prosocial behavior (Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009; Krupka and Weber, 2013; Bénabou et al., 2019), lying aversion (Gächter and Schulz, 2016; Abeler et al., 2019), costly punishment (Fehr and Gächter, 2000; Fehr and Fischbacher, 2004), all the way to labor market outcomes (Akerlof, 1980; Fehr et al., 1998; Lindbeck et al., 1999), and the effect of incentive schemes and contracts (Sliwka, 2007; Kessler and Leider, 2012; Huck et al., 2012; Danilov and Sliwka, 2017). Yet, economic decisions do not have to be guided only by *social norms*, but they can also be guided by *personal norms*. Both social and personal norms represent perceptions of appropriate behavior, i.e., behavior that is considered as moral or correct. However, they differ with respect to whose perceptions they capture — while social norms capture *societal* perceptions, personal norms capture *individually* held perceptions.[1] Scientists in neighboring fields have long argued that personal norms are an important driver of behavior (see, e.g., Schwartz, 1973, 1977; Schwartz and Fleishman, 1978; Cialdini et al., 1991; Bicchieri, 2005), yet — in contrast to social norms — they have received very little attention in the field of economics.

While these two normative prescriptions can coincide, they can also be in conflict in many economically relevant contexts. For example, someone who disapproves of wealth redistribution might have a different personal norm about tax compliance compared to the social norm of the social welfare-oriented society she lives in. Likewise, the normative beliefs of a person who supports universal equal rights can be in conflict with those of a society that openly discriminates against members of other ethnic or socioeconomic groups. Further, the norms that are promoted through organizational values (organizational "culture") of a company, e.g, on shirking, whistle blowing, or taking advantage of informational asymmetries between sellers and consumers, can be at odds with the personal norms of its employees. Such discrepancies of normative beliefs can exist in principle for any economic behavior that is governed by norms.[2] If people's behavior

[1] In this paper, we are interested in the clash between personal and social perceptions of *appropriate* behavior. Hence we focus on injunctive social norms, i.e., prescriptions of which behavior is socially appropriate, and not descriptive social norms, i.e., which behavior people usually do. While both norms can influence behavior (see, e.g., Cialdini et al., 1990; Bicchieri and Xiao, 2009; Krupka and Weber, 2009; Bartke et al., 2017), the former provide the ideal conceptual counterpart to personal norms and allow us to analyze how the two normative perceptions of appropriate behavior shape economic decisions.

[2] Similar conflicts can arise in the realm of social norms due to pluralistic ignorance (Katz et al., 1931), which describes a situation in which people are privately against a social norm, but wrongly believe that others are in

is not only driven by social but also by personal norms, they are then confronted with two (potentially competing) normative principles that can both determine behavior, and neglecting the latter would be detrimental to the understanding of economic decision making. Furthermore, it could lead to dangerous pitfalls in the design of normative interventions, and it may provoke frictions and affect the success of organizations.[3]

Taking this as our starting point, we propose a simple utility framework and design a novel experiment to study the relevance of personal norms across various economic games and settings. We estimate our utility framework with the collected data, and demonstrate that personal norms — while taking monetary payoff and social norms into account — are *strong predictors* of economic behavior. Our findings show that personal norms are: *i) distinct* from social norms across a wide range of economic contexts, *ii) robust* to an exogenous increase in the salience of social norms, which heightens the predictive value of social norms, but does not weaken that of personal norms, and *iii) complementary* to social norms in predicting behavior, as a model with both personal and social norms outperforms a model with only one of the two norms.

As a first step, we present a simple utility framework in which people care about their monetary payoff, social norms and personal norms. More precisely, we assume people care about the money they earn from an action, the degree to which this action complies with their beliefs about social norms, i.e., what society finds appropriate, and the degree to which this action complies with their own private perception about what is appropriate.[4] This captures a decision-making process in which two normative principles — one imposed from within the person and the other one by society — are decisive for behavior.[5] We then design a two-part experiment which allows us to investigate the predictive value of personal norms as well as social norms across four economic games.

In the first part of the experiment, we elicit both social and personal norms in an online session for four games: dictator game, dictator game with tax, ultimatum game and third-party punishment game. To do so, we design a simple method to elicit beliefs about personal and social

favor of it. An example is that of racial segregation in the USA (O'Gorman, 1975). Recently, Bursztyn et al. (2020) show how correcting these beliefs can lead to important changes in behavior.

[3]We return to these issues when we discuss our results in Section 5.

[4]Note that personal norms are distinct from preferences. We discuss the conceptual distinction in Section 2 and bolster this with empirical evidence in Section 4.3

[5]This aspect marks the key distinction of our framework in comparison to classic approaches to modeling prosocial behavior, as it distinguishes between *who* is dictating the fairness principle — the individual or the society (see Section 2 for a discussion).

norms with a symmetric procedure. Subjects go through an adapted version of the Krupka and Weber (2013) social norms elicitation method and a symmetric procedure for eliciting personal norms, in randomized order. The main difference between the two procedures is the following: subjects evaluate all possible actions *i)* according to the opinion of *society* for social norms, or *ii)* according to their *own* opinion for personal norms. We demonstrate that the two norms elicited with this procedure are correlated, but that there is substantial heterogeneity at the individual level in all four games, indicating that the two norms are empirically distinct.

In the second part of our experiment, we invite the same subjects to the lab approximately four weeks after the norm elicitation took place. In the lab, they play the four games from which we elicited the norms. Importantly, this was not revealed to them prior to the lab experiment. We then connect subjects' behavior elicited in the lab to their answers on personal and social norms elicited in the online experiment. We estimate our utility framework by using a conditional (fixed-effect) logit choice model, and show that personal norms — while taking social norms and monetary payoffs into account — are highly predictive of individuals' behavior. This finding holds across all four games individually, as well as when analyzing them together. In addition to personal norms, our results also reveal that the other two components of our framework — social norms and monetary payoff — are predictive of behavior. Having demonstrated the strong relation between personal norms and economic behavior in a treatment where decisions remain private (Private), we then analyze the results of a treatment in which we *exogenously* increase the salience of social norms (Social). As economic decisions are rarely taken in a social vacuum, we investigate the predictive value of the two norms when subjects' actions are under the scrutiny of others. Following the reasoning of Bicchieri (2005), we hypothesize that this manipulation will increase subjects' concerns for their social image, leading them to act more in line with the views held by society. Hence, we expect the relation between social norms and behavior to become stronger. We find that, on average, the relation between social norms and behavior is strengthened by our manipulation. The weight people put on social norms significantly increases for two out of four games, as well as when pooling all four games together. This change, however, does not come at the expense of personal norms, since their relation with behavior not only survives, but the weight people put on personal norms remains stable between the two treatments. Together, these results show that personal norms are strong predictors of behavior across different contexts, and support their role as a fundamental behavioral motive.

To substantiate further the importance of personal norms and to validate our two-norm utility framework, we pit it against two alternative ones: one in which subjects only care about social norms and their monetary payoff (see, e.g., Krupka and Weber, 2013) and the other one in which they only care about personal norms and their monetary payoff. If our fundamental assumption is true that people's decision-making is not only influenced by social norms, but also by personal norms, the inclusion of both norms should lead to an improvement in the predictive fit of the estimated models. Comparing Log-likelihood ratios between the models, we find that adding personal norms significantly increases the predictive fit for almost all games, across both the PRIVATE and the SOCIAL treatment. We then also preform this comparison with the Akaike and Bayesian information criteria, which penalize the inclusion of additional predictors. Again, for both measures, the majority of comparisons favors the two-norm model (in almost all other cases, it is actually the model with only personal norms that prevails). This model comparison exercise supports the central assumption of our framework, and shows that the inclusion of personal norms leads to an increase in our ability to predict economic behavior.

After establishing our main results, we provide several robustness checks to further support our findings and counter alternative explanations. First, we argue and provide evidence for why our results cannot be explained by a preference for consistency (Falk and Zimmermann, 2018). We minimize this concern through features of our design, most importantly, a long time lag between the online and the lab session (approx. 4 weeks). Moreover, we perform a robustness check and show that restricting our sample to subjects who report having an imperfect recall of the online part leaves our results intact. Second, we argue and provide multiple robustness checks against the concern that our results might be due to subjects rating the actions according to their preferences, i.e., according to which action maximizes their utility, instead of answering according to their personal norms. We tackle this issue already in our experimental design, most importantly, by carefully constructing the elicitation of personal norms in a way that clearly distinguishes them from preferences. We then show that these concerns are not supported in our data: *i)* monetary payoff and social norms bear strong predictive value and increase the goodness-of-fit of our models, which is at odds with the alternative explanation; *ii)* the actions which according to the alternative explanation should be among the least preferred ones are often chosen, and personal norms remain a robust predictor when we restrict our attention to these actions; and *iii)* a non-negligible proportion of actions which are rated as most personally appropriate are inconsistent with the alternative explanation, and for those decisions personal

norms remain a significant predictor. Third, we test whether our results are affected by the operationalization we use for social norms. Instead of using subjects' beliefs about the social norm, we assign them a common social norm by computing the average beliefs across all subjects — in line with Krupka and Weber (2013) — and re-run our main regressions. Also in this case, our results stay robust. Finally, we test whether the order in which the norms were elicited has any effect on our results. We utilize the fact that some subjects faced the social norm elicitation while others faced the personal norm elicitation first and test whether our results remain robust when looking at each order separately. Moreover, we re-estimate our utility framework by adding an interaction term between each of the two norms and elicitation order. Again, our results remain robust.

As a last step, we go beyond our main analysis and complement it with the results of an additional experiment run with a different set of subjects. In this experiment, we elicit the personal and social norms for: *i)* the same four games as in the main experiment, and *ii)* seven additional games as well as a battery of vignettes representing economic situations in everyday life or at a workplace, e.g., nepotism, tax evasion, shirking, and misreporting the amount of hours worked. We successfully replicate the distributions of the two norms as well as the results on their relation (correlation and individual-level differences) across the four games, indicating that these findings are stable across comparable populations. Moreover, we find substantial heterogeneity between the two norms in all the additional games and vignettes, showing that the differences between the two norms are not restricted to the four games in our main experiment. In combination with our main findings, this suggests that personal norms are a relevant behavioral predictor across a wide range of economic contexts.

Our results contribute to the literature investigating the effect of social norms on economic behavior (see, e.g., Krupka and Weber, 2009; Kessler and Leider, 2012; Krupka and Weber, 2013; Gächter et al., 2013; Banerjee, 2016; Kimbrough and Vostroknutov, 2016; Agerström et al., 2016; Krupka et al., 2017; Gächter et al., 2017; Barr et al. 2018; Boonmanunt et al., 2020; Bicchieri et al., 2020; Erkut, forthcoming). We show that, in addition to social norms, personal norms have a strong and robust relation to economic behavior. Importantly, we show that they *complement* social norms in predicting behavior.

Our study is closely connected to a nascent stream of papers which explore the idea that — alongside social norms — people also care about some type of private values or private normative beliefs. Michaeli and Spiro (2015) utilize this idea to explain theoretically how

strictness in societies can affect publicly stated opinions and te Velde (2019) to show how using social incentives might backfire. Closer to our study, Burks and Krupka (2012) study social norms across different groups in a financial firm. They also elicit personal norms, and report that a summary measure of overall misalignment between the two norms in a vignette on whistle-blowing is correlated to job satisfaction and behavior in an "advice game". Moreover, Alempaki et al. (2021) study how using a foreign language affects lying behavior, and elicit both personal and social norms after the lying task. They report that both norms correlate with lying behavior.[6] These studies point to the potential that personal norms hold; however, they do not aim (and hence are not designed) to empirically isolate the relevance of personal norms. In our study, we employ a novel design with two experimental parts, separated by a time lag. This allows us to cleanly *estimate and establish* a relation between both social and personal norms, on the one side and behavior, on the other, across various experimental games and settings.

Our findings also advance the literature on image concerns, in particular in the domain of prosociality. We connect to studies on self-image (see, e.g. Dana et al., 2007; Gneezy et al., 2012; Grossman and Van Der Weele, 2017; Adena and Huck, 2020; Bašić et al., 2020; Falk, 2021), as personal norms hinge on inner enforcement mechanisms which rely on the image or concept one has of herself. Moreover, we contribute to the understanding of social image concerns (see, e.g., Andreoni and Petrie, 2004; Alpizar et al., 2008; Ariely et al., 2009; Andreoni and Bernheim, 2009; Schram and Charness, 2015; Krysowski and Tremewan, 2020), as we report evidence underscoring the relevance of social norms in settings in which social image concerns are high.

Finally, our findings also relate to signaling models which capture the relations between image concerns, social norms and behavior (see, e.g., Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009; Bénabou and Tirole, 2011). Bénabou et al. (2019) consider a setting with multiple audiences that have conflicting normative values. Our results speak to the relevance of such settings, as the presence of an internal audience (judging according to personal norms) and an external audience (judging according to social norms) is common to many economic decisions.

[6]There is also evidence showing that nudging people to think about what they personally believe is moral before making a decision affects their behavior (Capraro et al., 2019; Bilancini et al., 2020).

2 Social and personal norm-dependent utility framework

We start by defining the two concepts that build the cornerstones of our framework, social and personal norms. Regarding social norms, we closely follow the approach of Krupka and Weber (2013) and stay in line with other seminal work on the topic (Elster, 1989; Ostrom, 2000; Bicchieri, 2005). We conceptualize social norms as collective perceptions among members of a group or society regarding the appropriateness of different actions in a given situation, where the standard of appropriateness is judged according to what is perceived as moral or correct behavior. In this sense, they represent shared understandings about actions that are permitted or not. They hinge on expectations of others and can be enforced by external sanctions or the threat thereof (see Bicchieri, 2005). Importantly, the definition we use implies that, for each potential action in a given situation, it is possible to attach a socially accepted value which indicates how appropriate the action is perceived to be from the viewpoint of the respective group or society.

In contrast to social norms, personal norms represent one's *private* perceptions about the appropriateness of different actions in a given situation. To define them, we follow Schwartz (1973, 1977) who argue that personal norms come from internalized values and deviations from them are subject to intrinsic sanctioning tied to the self-concept, e.g., guilt or self-depreciation (see also Schwartz and Fleishman, 1978; Elster, 1989; Cialdini et al., 1991). In this sense, personal norms do not hinge on others' expectations to follow them (see Bicchieri, 2005), and can diverge from social norms. They arise from one's conviction of what is the right thing to do, reflecting the views on what one personally considers as moral or correct behavior (see Schwartz, 1973, 1977; Bicchieri, 2005).[7] In this respect, it is important to stress the conceptual difference between personal norms and preferences. While the latter answer the question of what one *prefers* or *wants*, and are a product of a utility maximization process, the former answer the question of what one personally perceives to be appropriate in terms of *moral* or *correct* behavior, and are a product of one's inner convictions. Following this line of reasoning, we define *personal norms* as a person's individual perception regarding the appropriateness of different actions in a given situation, where the standard of appropriateness is judged according to what an individual perceives as moral or correct behavior, irrespective of the opinion of others. Consequently, we assume that it is possible to attach a personal value to how appropriate each

[7] Bicchieri (2005) also refers to these type of norms as moral norms.

action is in a given situation.[8]

Concerning the relation between the two norms, since personal norms are a product of internalized values, they can originate from (and hence be congruent with) social norms. However, they can also diverge from them via different processes (see Schwartz, 1973). For instance, an individual could engage in introspective moral reasoning or be exposed to different (norm-relevant) information compared to others (see Tremewan and Vostroknutov, 2020), which (over time) could lead her to adopt personal values that are different from the socially prescribed ones. Moreover, she could adopt a new social norm after moving to another society or after a new social norm was formed, but still personally hold the old norm. Or, also, this discrepancy could originate at an early age due the transmission of parental personal values that differ from socially prescribed ones. While we do not focus on the sources of such discrepancies, we do rely on their existence to be able to disentangle the relation between the two norms and economic behavior.

Before turning to our utility framework, we note that our normative definitions imply that social norms are shared values, whereas personal norms can differ across people. While an individual should have more or less perfect insight into what her own personal norm is, she can only rely on her belief about the social norm. Elicitations of social norms often reveal that some individuals fail to guess the normative belief of the majority, i.e., their belief about the social norm is inaccurate (see, e.g., Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016; Krupka et al., 2017). Thus, if individuals act upon their beliefs and not upon the commonly recognized social norm, using the latter could potentially misidentify its relation with behavior. For this reason, in our utility framework and analysis, we rely on what subjects *think* is appropriate from the viewpoint of society (belief about the social norm), and what they themselves perceive as appropriate (personal norm). We find this to be a more natural way to compare personal and social norms. We later relax this assumption and repeat our entire analysis by assuming that people care about the commonly recognized social norm instead of their belief about the social norm.

[8]The two normative concepts can also be viewed through the prism of the literature on morality. Following a descriptive definition, morality refers to "certain codes of conduct put forward by a society or a group (e.g., a religion), or accepted by an individual for her own behavior", and the latter does not have to be congruent with the former (Gert and Gert, 2020). In this sense, our study can be viewed as contrasting these two notions of moral behavior. As such, our work also connects to the literature on morality and economic decision making, for example, studies on moral motivation and behavior (see, e.g., Dana et al., 2007; Cappelen et al., 2017; Falk, forthcoming) and on modeling of moral actions (see, e.g., Bénabou and Tirole, 2011; Bénabou et al., 2019).

To construct our utility framework, we build on recent social norm models (see, e.g., Kessler and Leider, 2012; Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016; Krupka et al., 2017; Gächter et al., 2017; Barr et al., 2018), which explain social behavior by positing that — alongside monetary payoff — people care about adhering to social norms. In contrast to classic social preferences models, in which the fairness principle people care about is part of the assumptions — for example, equality (e.g., Fehr and Schmidt, 1999), or reciprocity (e.g., Falk and Fischbacher, 2006) —, here it arises through a collectively recognized social norm, which can change depending on the context and can be measured empirically (see Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016, for a discussion). These models have recently gained a lot of attention, as they possess strong predictive capabilities and are successful in a wide array of contexts, also in those in which social preferences models commonly fail, for example, when subjects can exploit moral wiggle room (e.g., Dana et al., 2007), or when subtle contextual changes alter behavior (e.g., Krupka and Weber, 2013). Here, we adapt the social norm models by positing that people also care about adhering to their personal norms. Guided by a long-standing literature in neighboring fields (see, e.g., Schwartz, 1973, 1977; Schwartz and Fleishman, 1978; Cialdini et al., 1991; Bicchieri, 2005), we introduce another (potentially conflicting) normative principle which can also determine behavior. Analogously to social norms, personal norms are distinct from classic social preferences, as they capture normative principles that — instead of being part of the assumption — arise from private normative values, which can change depending on the context and can be measured empirically. Importantly, they can differ at the individual level, as their key assumption is that they are *privately* held. This assumption is what gives rise to the key novelty of our approach. Instead of focusing on *how* the fairness principle is modeled and captured (which differentiates social norm models and classic social preferences models), our framework sets to separate *who* is dictating the fairness principle — the society or the individual.

We now describe our utility framework. An individual i takes an action a_k from a set of possible actions $A = [a_1, ..., a_K]$. She cares about: *i)* the monetary payoff $\pi(a_k)$ she gets from the action, *ii)* her belief about the appropriateness of the action from society's view $S_i(a_k)$, *iii)* and her own private perception about the appropriateness of the action $P_i(a_k)$.[9] $S_i(a_k)$ and $P_i(a_k)$ are functions that assign an appropriateness score in an interval $[-1, 1]$ to each action. $S_i(a_k)$ represents the perception about the commonly held view in society and, hence,

[9] A similar concept is also presented in Burks and Krupka (2012).

describes the subjects' beliefs about how socially appropriate or inappropriate it is to perform a certain action. Similarly, $P_i(a_k)$ describes the subjects' perception about how appropriate or inappropriate it is to perform an action from their own viewpoint. In both cases, a negative score means that the action is perceived as inappropriate, whereas, if the score is positive, the action is considered to be appropriate. The utility function of an individual is then simply given by:

$$U_i(a_k) = V(\pi(a_k)) + \gamma S_i(a_k) + \delta P_i(a_k). \tag{1}$$

Here, $V(\cdot)$ is the utility derived from money. The two parameters $\gamma, \delta \geq 0$ represent the tendency or concern of an individual to follow the social and personal norm. They are zero for an individual who is entirely untroubled by the two. The larger they are, the more an individual is influenced by the respective appropriateness ratings. While an individual might want to follow both norms, she could also be highly concerned by the social appropriateness of an action and not by the personal appropriateness, or the other way around.[10] We assume that the two parameters are determined by one's preferences to follow the respective norms, but can also be affected by other, external factors. Specifically, we follow Bicchieri (2005) in arguing that "situational factors may increase the effect of norms on behavior by making a norm salient" (p. 46); hence, we assume that γ and δ can be affected by the environment (see also Berkowitz and Daniels, 1964; Schwartz and Fleishman, 1978; Rutkowski et al., 1983; Cialdini et al., 1991). We use this assumption for our manipulation of social norms salience (see Section 3.5).[11]

[10] The assumption of additive separability between the two normative utility components is derived from the literature, as personal norms are generally conceptionalized as separate and independent drivers of behavior from social norms within the decision-making process (see, e.g., Schwartz, 1973, 1977). Importantly, we also test this assumption using our dataset, and find no evidence of any interaction effects between the two norms (see footnote 24 in Section 4.2).

[11] One could formalize this more explicitly by assuming that γ (δ) is a function of subjects' context independent preferences to follow social (personal) norms, and the context dependent salience of social (personal) norms. This reflects a similar intuition to the one used when modeling prosocial behavior through reputational concerns in signaling models. For example, Bénabou and Tirole (2006) assume that a single parameter captures one's concern for prosocial reputation, and that this parameter can be separated into one's context independent concern to be perceived positively, and the context dependent visibility or salience of one's actions.

3 Experimental design and predictions

Our experimental design consists of two parts: an online experiment and a laboratory experiment. Each subject participated in both the online and the lab part, which were separated by a considerable time lag. In both parts, subjects went through four different games. In this section, we first give an overview of our four games (see Section 3.1). We then describe the online experiment in which we elicited subjects' social and personal norms for the four games along with other variables (see Section 3.2). Following that, we illustrate the design of the lab experiment, where subjects played the four games, either in a PRIVATE or a SOCIAL treatment in a between-subjects design (see Section 3.3). We conclude by detailing the experimental procedure (see Section 3.4) and state our predictions derived from the theoretical framework (see Section 3.5).

3.1 Games

We chose the following four games: dictator game, dictator game with tax, ultimatum game, and third-party punishment game. The dictator game (Kahneman et al., 1986; Forsythe et al., 1994) is one of the most widely studied experimental setups, which captures individuals' prosocial behavior in the absence of strategic interaction. The dictator game with tax (Andreoni and Miller, 2002) extends this setup to a broader range of motives, as it introduces a conflict between competing fairness principles. The ultimatum game (Güth et al., 1982) is a widely-used paradigm that (in contrast to the first two games) investigates fairness concerns in a strategic setting. Finally, the third-party punishment game (Fehr and Fischbacher, 2004) is a more recent, but highly influential setup that studies norm enforcement and altruistic punishment. We chose these four games to demonstrate that our results apply to a variety of economically relevant settings. We aimed to capture: *i)* a broad range of motives that are present as drivers in various economic decisions, and *ii)* important economic contexts, which have gained a lot of attention in previous research. The complete structure of the games was kept constant both when we elicited norms in the online experiment and when subjects played the games in the lab experiment.

Dictator game In the dictator game (DG), two participants are randomly matched together. We implement role uncertainty: participants do not know their role at the beginning and both have to decide how they would split an endowment of €10 (in intervals of €1), if they were

assigned to the role of dictator.[12] This decision is private and both decide without knowing what the other participant would choose. The decision of the actual dictator is then implemented.

Dictator game with tax The dictator game with tax (DGT) is identical to the DG above, except that the endowment to be split is of €12, and any amount sent to the recipient is reduced by 40% (the tax). Subjects can send amounts in €1.50 increments (€1.50, €3, ..., €12). Note that sending €0 maximizes the sum of payoffs, while sending €7.5 ensures equal earnings for both players (€4.5) and sending €6 equalizes the two shares before taxation.

Ultimatum game In the ultimatum game (UG), two participants are randomly matched together and assigned the roles of proposer and responder. The proposer gets €10 and can offer any integer amount from €0 to €10 to the responder. If the responder accepts the offer, the €10 are divided as suggested by the proposer. If she rejects the offer, both participants earn nothing. We elicit the responder's choice using the strategy method (Selten, 1965): the responder has to state the minimum offer she would accept. Any offer greater or equal to the declared amount is accepted, while those below are rejected. The payoffs are determined by matching the proposer's actual offer with the choice of the responder. In this game, we are interested in responders' rejection behavior.[13]

Third-party punishment game In the third-party punishment game (TPP), three subjects are randomly matched together. One of them is assigned the role of dictator. The other two subjects both have to indicate how they would decide if assigned the role of third party. The dictator gets €10 and can give either €0, €2 or €5 to the recipient. The third party can punish the dictator. She gets €5 and can reduce the dictator's payoff by €3 for each punishment point she assigns, with the dictator's payoff being bounded below by €0. Each punishment point costs her €1 and she can assign at most 2 punishment points. We elicit the third party's choice using the strategy method: the third party has to assign punishment points for each possible choice of the dictator (€0, €2, or €5). The decisions are private and all three subjects decide without knowing what the other subjects have decided. Punishment points are then assigned

[12] We used role uncertainty in games where there is a passive player: dictator game, dictator game with tax, and third-party punishment game.

[13] While the behavior of proposer is also interesting, the effects of norms are not straightforward to identify. In particular, the behavior depends on the proposer's personal and social normative perceptions, as well as her beliefs about the responder's behavior, which is also driven by her personal and social normative perceptions of the situation. Hence, staying close to our theoretical framework, we focus on the behavior of responders.

according to the actual choice of the dictator and the punishment choice of the actual third party. In this game, we are interested in third-parties' behavior.

3.2 Online experiment

Subjects received a link to access the online experiment immediately after subscribing to the laboratory experiment. This occurred four weeks before the first session of the lab experiment, which took place on three consecutive days. Subjects had six days to complete the online experiment, which means they completed it between 23 and 30 days before their lab session. This long time lag was specifically chosen to reduce subjects' recollection of the online tasks and of their exact answers, once they came to the lab. At the beginning of the online session, participants generated a code which we used to match their data between the online and the lab session. Then, they proceeded to the main task: the elicitation of their beliefs about social and personal norms in the four games, as described below. The elicitation of norms was organized in two blocks: a block with personal norms, and a block with social norms. The order of the two blocks as well as the order of the games within the blocks was randomized at the individual level. Each block started with an explanation of the task and an example. While completing the first block of norm elicitations, subjects were unaware of the upcoming second block. For example, if they faced the personal norm elicitation first, they were not aware that afterwards they would be facing the social norm elicitation. After both blocks, we collected some demographic variables.[14]

Social norms We elicited social norms using an adapted version of the widely used Krupka and Weber (2013) elicitation method. We carefully phrased the text in a manner allowing us to directly contrast personal and social norms. Subjects had to rate how socially appropriate they believed each action to be on a 6-point Likert scale, ranging from very inappropriate to very appropriate. In particular, we used the following text: *"For each action, evaluate according to the opinion of the society and independently of your own opinion, whether it is appropriate or not to choose it. "Appropriate" behavior means the behavior that you consider most people would agree upon as being "correct" or "moral"."* (See Appendix B.1 for the full instructions). We rescaled the answers to an interval from -1 to 1 for the subsequent analysis. Subjects

[14] We asked for subjects' gender, age, field of study, number of siblings, favorite food and favorite movie. The last two variables were an additional safeguard to be able to distinguish subjects if they had the same code, which was never the case.

received €0.30 for each answer that matched what most other subjects had chosen, earning up to €12 from this task. This provides an incentive to coordinate on the social norm (for further discussion, see Krupka and Weber, 2013).

Personal norms We elicited personal norms with a symmetric procedure to the one just described for social norms. However, instead of asking for the social appropriateness, we asked subjects to rate how personally appropriate they believed each action to be, irrespective of the others' views. In particular, we used the following text: *"For each action, evaluate according to your own opinion and independently of the opinion of others, whether it is appropriate or not to choose it. "Appropriate" behavior means the behavior that you personally consider to be "correct" or "moral".*" Subjects answered on a 6-point Likert scale, ranging from very inappropriate to very appropriate, and were asked to answer as precisely as possible with their honest opinion. Answers are re-scaled between -1 to 1 for the subsequent analysis. This elicitation was not incentivized, as personal norms are by definition an individual value and cannot be matched to others' personal norms (see Burks and Krupka (2012) for a similar method).

3.3 Laboratory experiment

The main purpose of the lab experiment was to elicit subjects' behavior in the four games. Each subject played all games and their order was randomized at the individual level. We imposed perfect stranger matching, i.e., each subject could only be matched once with another given subject across the four games. One game was randomly selected to determine the payoff. The outcomes of the games as well as the payoff and the role assignment were revealed only after all subjects went through all four games.

Subjects were randomly assigned to one of two treatments. The treatment assignment was done at the session level, i.e., all subjects in one session were in the same treatment. In the PRIVATE treatment, subjects made the decisions for the four games in an anonymous setting. In the SOCIAL treatment, we *exogenously* manipulated the visibility of subjects' actions in order to increase their social image concerns. To this end, subjects were informed at the very beginning of the experiment that, after all participants had completed all tasks, they all had to stand up so that everyone could see and hear everyone else. A laboratory assistant would subsequently call up each participant one after the other. Participants would then have to say their first name and what they had chosen in each of the four games. Specifically, they would have to read verbatim

a text displayed on their screen containing all information regarding all the decisions they had taken. Importantly, this approach ensured that the environment during the decision-making stage was kept constant across the two treatments, and the only difference was the information about whether their behavior would become publicly known or not (see Ariely et al. (2009) and Ewers and Zimmermann (2015) for similar manipulations, and Appendix B.2 for experimental instructions).

Before the start of each game, subjects had to answer control questions to make sure they had understood the experimental instructions correctly. Once subjects had completed the main part of the experiment, they went through a short series of questionnaires. This included a measurement of participants' reputational concerns (adapted from Romano and Balliet, 2017) and questions about their recollection of the online experiment and some sociodemographics.

3.4 Procedure

The experiment was conducted at the Cologne Laboratory for Economic Research (CLER) of the University of Cologne between October and November 2019. The online experiment was conducted using Qualtrics, while the laboratory experiment was programmed in z-Tree (Fischbacher, 2007). Subjects were recruited via Orsee (Greiner, 2015). The invitations contained the information that subjects' decisions in the study might be disclosed to other participants. Our sample consists of 250 subjects that took part in both the online and the lab experiment (62% female, average age 25.8 years). In Appendix A.1, we show that there was no systematic attrition between the online and lab experiment. Out of 250 subjects, 127 participated in the PRIVATE treatment and 123 in the SOCIAL treatment.[15] All subjects received a show-up fee of €8, plus their earnings from the online experiment and their earnings from the laboratory experiment. Overall, subjects received a payment of €17.3, on average. The online experiment lasted between 20 and 35 minutes, while the laboratory experiment took on average 50 minutes.

3.5 Predictions

We have three main predictions for the results of our experiment. First, as highlighted in Section 2, we expect personal and social norms to be related, since personal norms represent internalized values which may originate from the society; however, they do not need to be identical. In fact,

[15] 13 subjects participated in the lab experiment without having completed the online experiment. We exclude those lab observations, as all our analyses rely on subjects' ratings of personal and social norms, which was done in the online part.

many economic settings contain a multitude of normative principles (e.g., equality, altruism, payoff maximization, efficiency) that could give rise to discrepancies between the two norms. This heterogeneity represents a *conditio sine qua non* for identifying the differential relation between the two norms and behavior.

Hypothesis 1 *Perceptions of social and personal normative appropriateness are correlated; however, there is non-negligible heterogeneity between the two at the individual level.*

Second, while it is well-established that social norms and monetary payoff influence economic behavior, we conjecture that personal norms are also a driver of behavior; thus, we expect them to play an important role in explaining subjects' actions across the four games.

Hypothesis 2 *Personal norms play a substantial role in explaining behavior:* $\delta > 0$ *(Equation (1)).*

Third, as described in Section 2, we also posit that the weight put on social and personal norms might differ across situations. Our treatment manipulation in the SOCIAL treatment is aimed at making *only* social norms more salient. Since social norms, in contrast to personal norms, are subject to others' expectations of following them (Bicchieri, 2005), we conjecture that increasing the visibility of actions, i.e., social image concerns, will make subjects more concerned about the opinion of others. If there is an expectation to follow the social norm, the manipulation should raise the influence of social norms on behavior; thus, we expect the social norms' parameter to increase.[16]

Hypothesis 3 *Social norms play a more important role in the* SOCIAL *treatment compared to the* PRIVATE *treatment in explaining behavior:* $\gamma_{SOCIAL} > \gamma_{PRIVATE}$ *(Equation (1)).*

The increase in observability should not affect the influence personal norms have on behavior directly. However, if social norms become more salient, the presence of a strong competing normative principle could "override" the effect of personal norms (see Bicchieri, 2010). Thus,

[16]Note that, within our framework, costs and benefits of social image are fully captured by $\gamma S_i(a_k)$. As an example, think of an individual who has a binary choice between two actions: a_m and a_n, where $S_i(a_m) > S_i(a_n)$. The individual decides for a_m, so the utility she gains from the social norm component of her utility for choosing a_m over a_n is $\gamma(S_i(a_m) - S_i(a_n))$. Now imagine that she opts for the same action in the social image treatment as the exogenous manipulation (only) changes γ to $\widetilde{\gamma}$, where $0 \leq \gamma < \widetilde{\gamma}$. Her gain in utility from choosing a_m over a_n is now larger due to the increased weight she puts on the social norm component of her utility, and is captured by $(\widetilde{\gamma} - \gamma)(S_i(a_m) - S_i(a_n))$.

when analyzing how the SOCIAL treatment affects social norms, we will also test for potential indirect ("crowding-out") effects on personal norms.

4 Results

Our results are structured in the following way. We first give an overview of personal and social norms across the four games and provide evidence for their heterogeneity (see Section 4.1). Then, we move to our main results and analyze how personal and social norms are related to behavior (see Section 4.2). Here, we establish the predictive power of personal norms in the PRIVATE treatment and investigate how the weights of personal and social norms change in the SOCIAL treatment. Moreover, we pit our model against two competing models where subjects only care about one of the two norms to compare their predictive power. Following that, we perform a series of robustness checks to validate our main results (see Section 4.3). Finally, we report data from an additional experiment in which we replicate the patterns of social and personal norms in our main games, and perform the same exercise for seven additional games and ten vignettes depicting real-life economic situations (see Section 4.4).

4.1 Overview and heterogeneity of personal and social norms

We start by providing evidence for the heterogeneity of personal and social norms. As argued in Hypothesis 1, we expect social and personal norms to be related, but also sufficiently distinct from each other. In line with our conjecture, we find that appropriateness ratings of personal and social norms have a strong relationship. Specifically, we observe strong and significant correlations across all four games: 0.72 for the DG, 0.65 for the DGT, 0.74 for the UG, and 0.76 for the TPP ($p < 0.001$ for all correlations; Pearson product-moment correlation). However, this strong relation masks important heterogeneity. To investigate the differences at the individual level, we look at the personal and social appropriateness ratings of the available actions in each of the four games and check whether and to what extent the two ratings differ. We visualize this information in Figure 1. For each individual, we subtract her personal-norm appropriateness rating from her social-norm appropriateness rating for all possible actions across the four games. The difference can range from -2 to 2. A difference of 0 means that the two ratings are the same.

One can easily notice that, while a difference of 0 is frequent, for a substantial amount of

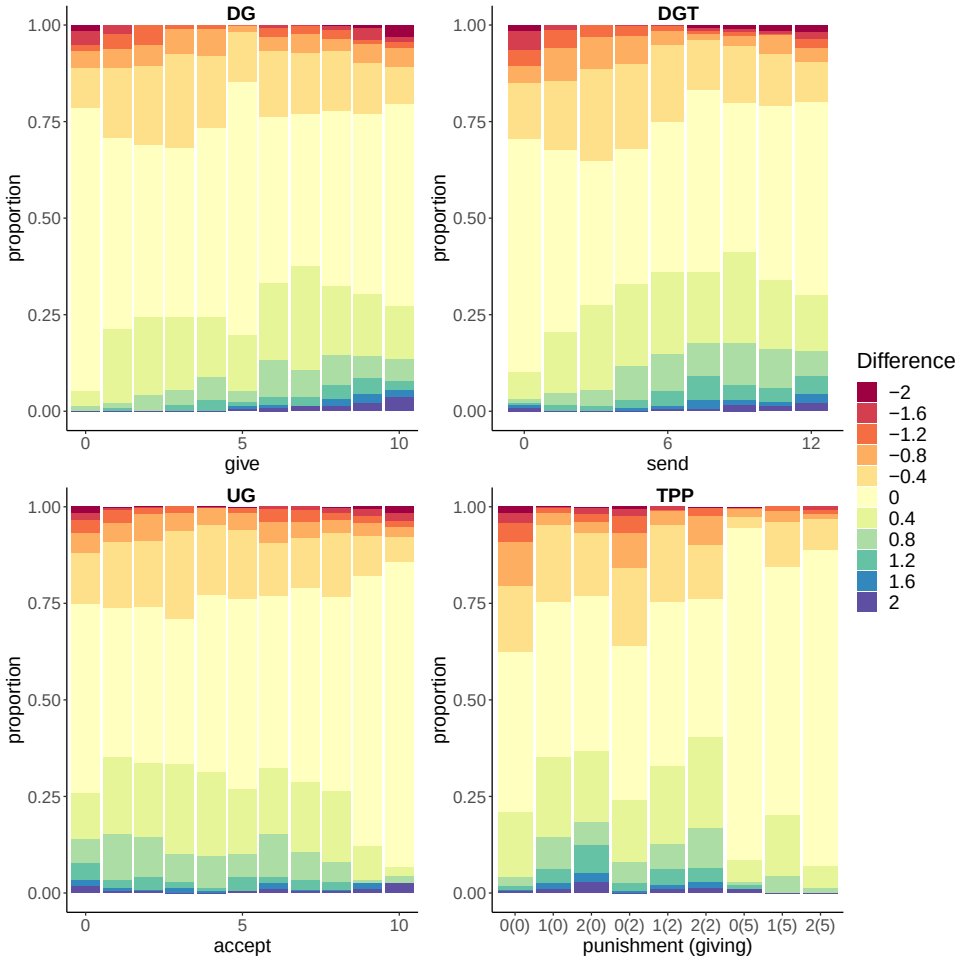


Figure 1: Individual difference between appropriateness ratings of social and personal norms

Note: The difference is calculated by subtracting an individual's personal appropriateness rating from her social appropriateness rating. The proportion of each difference is displayed for each action in a given game. In the DG, a subject can send from €0 to €10. In the DGT, a subject can send from €0 to €12. In the UG, a subject chooses the minimum offer she is ready to accept (from €0 to €10). In the TPP, a subject decides how many punishment points she wants to assign $(0, 1, 2)$ depending on how many euros the dictator gives $(0, 2, 5)$.

cases there is indeed a difference in the ratings of social and personal norms. In fact, a difference is present for 49.89% of the cases in DG, 55.64% in DGT, 49.67% in UG, and 47.56% in TPP.[17] All proportions are significantly different from 0 (the 99% asymptotic binomial confidence interval does not contain 0 in any comparison).[18] This confirms our conjecture and constitutes

[17]Note that the frequency of positive and negative differences is fairly similar across actions and games, indicating that personal norms are not equal to a biased version of social norms or vice versa.

[18]These results are not driven by a small subset of individuals. To show this, we calculate the proportion of non-zero differences between the two ratings for each individual, across all the actions in the four games. We find very few subjects whose proportion of non-zero differences is negligible, as well as very few who (almost) always exhibit a difference. Around 90% of subjects show non-zero differences between the two ratings in 25% to 75% of cases (the average across all individuals is 51%, $sd = 15.83\%$).

an excellent precondition to study the importance of personal and social norms for people's behavior.

Result 1 *While social norms and personal norms are correlated, there is substantial heterogeneity at the individual level across all four games.*

4.2 Personal norms, social norms, and behavior

We now join the data regarding personal and social norms from the online experiment with the behavioral data from the lab. This allows us to find out whether personal norms are predictors of behavior in the four games, as conjectured by Hypothesis 2.

To estimate our utility framework (Equation (1)) and capture the predictiveness of the two norm ratings, we follow the approach of the current literature on social norms (see, e.g., Krupka and Weber, 2013; Gächter et al., 2013; Krupka et al., 2017), and employ a conditional (fixed-effect) logit choice model (McFadden, 1973). In this regression model, the dependent variable is a dummy variable indicating whether a subject chose a given action, and the independent variables are the characteristics of that potential action: the monetary payoff attached to the action[19], the individual's social appropriateness rating of that action, and her personal appropriateness rating of that action. The obtained coefficients provide estimates for the weights of our utility framework (for more details, see Appendix A.2).

Table 1 provides the estimates of our model in the PRIVATE treatment. First, we look at the personal norm ratings. We find that personal norms have sizable and significant positive coefficients across all four of our games. Pooling the four games together, we observe that the personal norm coefficient remains large and significant. Turning to social appropriateness ratings, we find a significant coefficient in all games except in the UG. Looking at the pooled dataset, we observe that social norms have a significant positive relation with behavior. Finally, in line with previous findings and standard economic theory, we also find that monetary payoffs are a strong and significant predictor of behavior.[20]

[19] In line with previous work, we assume a linear restriction on the utility derived from money $V(\cdot)$, such that, for any payoff, $\pi(a_k)$, $V(\pi(a_k)) = \beta\pi(a_k)$ (see, e.g., Krupka and Weber, 2013; Gächter et al., 2013; Krupka et al., 2017). Thus, we estimate β which captures the weight subjects place on monetary payoff.

[20] As two of our main predictors — personal norm rating and social norm rating — are strongly correlated, we calculate the Variance Inflation Factor (VIF) to test for potential multicollinearity issues. We calculate the VIF for each independent variable in each of the regression models reported in Table 1. We find that all values are below 5; hence, we do not find any indication that multicollinearity is a concern for our results (see Marquaridt, 1970; Hair Jr et al., 1995). The same holds for all regressions reported in Table 2.

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.727***	0.338***	0.514***	0.989***	0.443***
	(0.103)	(0.051)	(0.128)	(0.158)	(0.034)
Social norm rating	0.734**	0.628**	0.561	0.628***	0.514***
	(0.365)	(0.255)	(0.358)	(0.227)	(0.130)
Personal norm rating	1.399***	0.765***	0.819**	0.712***	0.933***
	(0.323)	(0.213)	(0.338)	(0.222)	(0.124)
Observations	1,397	1,143	704	504	3,748

Note: Estimation of conditional logit choice models with dummy variable indicating whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 1: Conditional logit estimation of choice determinants in PRIVATE treatment

Result 2 *Personal norms are a strong predictor of behavior across our four games.*

In the SOCIAL treatment, we made subjects' choices observable to others in order to increase their social image concerns.[21] According to our predictions, this manipulation should increase the importance of social norms for behavior (Hypothesis 3). If so, this could also have an indirect detrimental effect on the relation between personal norms and behavior.

Table 2 provides the estimates of our model where we test Hypothesis 3. We find that the coefficient of the interaction between social norm ratings and SOCIAL is positive and highly significant in the DG and the DGT. For the UG and the TPP, we do not find a significant interaction effect. Considering the pooled dataset, we observe that the interaction coefficient is positive and highly significant. Overall, while we observe differences across individual games, on average we find that social norms become more important when subjects' social image concerns are increased. Turning to the interaction between personal norm ratings and SOCIAL, we do not observe a significant effect in any of the four games nor when pooling the dataset together.[22] Indeed, if we look at the predictive value of personal norms in a regression that estimates

[21] To further confirm the validity of our SOCIAL manipulation, we also elicited responses to an adapted version of reputation concerns questionnaire by Balliet et al. (2009), which measures subjects' concerns about the opinion of others and thus represents a proxy for social image concerns. We find that subjects in SOCIAL are indeed more concerned about others' opinions than subjects in PRIVATE (two-sided t-test, $p < 0.001$, $N = 250$).

[22] The fact that the interaction between SOCIAL and the social norm rating is significant, while the interaction between SOCIAL and the personal norm rating is not, also underscores the notion that the two norms are distinct, complementing our Result 1.

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.763***	0.234***	0.561***	0.823***	0.358***
	(0.078)	(0.030)	(0.103)	(0.104)	(0.023)
Social norm rating	0.804**	0.313	0.585	0.527**	0.371***
	(0.343)	(0.218)	(0.358)	(0.206)	(0.120)
Personal norm rating	1.424***	0.709***	0.820**	0.750***	0.895***
	(0.323)	(0.203)	(0.339)	(0.215)	(0.119)
Social norm rating	1.259***	0.893***	0.316	-0.251	0.748***
× Social	(0.426)	(0.286)	(0.503)	(0.316)	(0.173)
Personal norm rating	0.182	-0.000	-0.176	0.372	0.091
× Social	(0.455)	(0.293)	(0.478)	(0.338)	(0.179)
Observations	2,750	2,250	1,397	990	7,387

Note: Estimation of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as well as an interaction term between personal and social norms ratings and a dummy for the Social treatment as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2: Conditional logit estimation of choice determinants interacted with Social treatment

their effect in the Social treatment (see Table A2 in Appendix A.3), the coefficients remain significant and comparable to coefficients in Private across all games.[23] This shows that the relation between personal norms and behavior remains strong and stable.[24]

Result 3 *The relation between social norms and behavior is on average stronger in the Social in comparison to the Private treatment. The relation between personal norms and behavior remains stable.*

Next, we put our utility framework to a further test. The central assumption of our social- and personal-norm dependent utility framework, is that people are not only influenced by social norms, but that two (potentially competing) normative principles guide their behavior. Hence, we pit our two-norm framework against two other frameworks. One in which subjects care

[23]Given the caveats in interpreting interaction terms in non-linear models (see Ai and Norton, 2003), as a robustness check, we re-estimate regressions from Table 2 with a linear probability model. The reported results remain robust.

[24]As our utility function assumes additive separability between the two normative utility components, we additionally test for potential interaction effects between the two norms. To do so, we add an interaction term between personal and social norms to our framework, and we re-estimate it for each of the four games and the pooled dataset, both in Private and Social. Across the 10 regressions, we find no evidence of interaction effects between personal and social norms ($p > 0.258$ for each regression).

only about their monetary payoff and social norms, which reflects the usual modeling approach taken by the social norms literature (see, e.g., Krupka and Weber, 2013), and another in which subjects care only about their monetary payoff and personal norms. We carry out a model comparison exercise to evaluate the predictive fit of these three models in each of the four games and for the pooled dataset, for both the PRIVATE and the SOCIAL treatment, using three complementary approaches. First, we perform a pairwise comparison of the Log-likelihood measures of each of the two single-norm models (with either the personal or the social norm) with the two-norm model and report the corresponding Likelihood ratio tests. Second, we perform a direct comparison of the three models by using the Akaike information criterion (AIC), and finally, by using the Bayesian information criterion (BIC). Both the AIC and BIC penalize for an increase in the amount of predictors.[25] Table 3 contains the comparisons. Overall, we find strong support for our utility framework. All comparisons of Log-likelihoods, both for the PRIVATE and the SOCIAL treatment across all four games and for the pooled dataset (20 pairwise comparisons), favor our two-norm model over the other two one-norm models. These differences are significant in 90% of cases (Log-likelihood ratio test). Turning to the AIC, results are strongly in favor of the two-norm model (9 out of 10 three-way comparisons). The social norm model never prevails, while the personal norm model arises as the winner in just one case. When using the BIC, the two-norm model is again the most successful one (5 out of 10 three-way comparisons). For the cases in which this model does not prevail, the comparisons primarily support the model with only personal norms (4 out of 10 three-way comparisons), and only once the model with only social norms.[26] To sum up, we find coherent evidence supporting our claim that personal and social norms are complements in predicting behavior. In the few cases in which this is not true, the model comparisons predominantly favor the model with only personal norms.

Result 4 *Personal norms and social norms complement each other in predicting behavior.*

[25] For the AIC and BIC there is no clear testing procedure to determine whether one model is better than the other, but the differences are to be interpreted in an ordinal way. In general, the greater the difference the stronger the support for one model over the other (Burnham et al., 2011).

[26] If we also look at which of the two one-norm models prevails in a direct comparison (without the two-norm model), both the AIC and the BIC results again suggest that both norms matter, as the personal norm model prevails 6 times, and the social norm model 4 times, for both measures. Complementary to our Result 3, the personal norm model always prevails in PRIVATE (5/5 comparisons), and the social norm model almost always does so in SOCIAL (4/5 comparisons; for both BIC and AIC), suggesting that personal norms might be more informative in an private setting, while social norms might be more informative in a public setting.

	DG			DGT			UG			TPP			All games		
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)	(5a)	(5b)	(5c)
	Sn	Pn	Sn + Pn	Sn	Pn	Sn + Pn	Sn	Pn	Sn + Pn	Sn	Pn	Sn + Pn	Sn	Pn	Sn + Pn
Private															
Log-likelihood	-240.988	-231.757	-229.699	-238.166	-234.569	-231.457	-119.641	-117.866	-116.645	-116.294	-114.734	-110.744	-738.654	-715.751	-707.913
Likelihood test Sn vs. Sn+Pn	22.58 (<0.001)			13.42 (<0.001)			5.99 (0.014)			11.10 (<0.001)			61.48 (<0.001)		
Likelihood test Pn vs. Sn+Pn	4.12 (0.042)			6.22 (0.013)			2.44 (0.118)			7.981 (0.005)			15.68 (<0.001)		
Akaike IC	485.976	467.514	465.398	480.331	473.137	468.914	243.281	239.732	239.290	236.587	233.468	227.487	1481.309	1435.503	1421.827
Bayesian IC	496.460	477.998	481.124	490.414	483.219	484.038	252.395	248.846	252.960	245.032	241.913	240.155	1493.767	1447.961	1440.514
Social															
Log-likelihood	-178.437	-186.443	-165.716	-230.950	-235.503	-225.354	-109.375	-110.965	-107.570	-126.056	-116.940	-116.378	-686.112	-687.040	-659.378
Likelihood test Sn vs. Sn+Pn	25.44 (<0.001)			11.19 (<0.001)			3.61 (0.057)			19.36 (<0.001)			53.47 (<0.001)		
Likelihood test Pn vs. Sn+Pn	41.54 (<0.001)			20.3 (<0.001)			6.79 (0.009)			1.13 (0.288)			55.32 (<0.001)		
Akaike IC	360.875	376.886	337.432	465.899	475.006	456.709	222.750	225.929	221.139	256.111	237.880	238.756	1376.225	1378.080	1324.756
Bayesian IC	371.295	387.306	353.062	475.918	485.025	471.737	231.832	235.011	234.762	264.484	246.253	251.314	1388.624	1390.479	1343.354

Note: Comparisons of log-likelihoods, Bayesian information criteria, and Akaike information criteria between models which include monetary payoff and social norm as predictors (Sn columns), models which include monetary payoff and personal norm as predictors (Pn columns), and models which include monetary payoff, and both social norm and personal norm as predictors (Sn + Pn columns). Comparisons are accompanied by likelihood ratio tests (p value in brackets) which are reported for the estimation of all individual games (Columns 1a - 4c) and all games together (5a - 5c), separately for Private and Social treatment.

Table 3: Model comparison

4.3 Robustness checks

After having established our main findings, we check their robustness in four ways. First, we argue and provide evidence against the conjecture that the predictive value of personal norms is due to a preference to behave consistently with the answers given in the online experiment. Second, we argue and report several checks against the argument that our findings might be due to subjects stating their preferences instead of personal norms. Third, we confirm that our results do not depend on the operationalization we use for social norms. In particular, we re-run our analysis and, instead of using an individual's belief about the social norm, we take the average across everyone's beliefs. Fourth, we show that the order in which the two norms were elicited does not influence our results.

Consistency. One could argue that our novel predictor, personal norms, is related to behavior due to a desire to act consistently. In particular, if people have a preference for consistency (see Falk and Zimmermann, 2018), they might want to behave in line with what they stated to be the personally most appropriate behavior in the online experiment. Our experiment was designed to minimize such concerns. During the online session, subjects answered to more than 80 items, including both personal and social norms, as well as the post-experimental questionnaire. After answering these questions, there was a time lag of approximately 4 weeks until the lab experiment. Hence, it is unlikely that subjects had a precise recollection of the specific answers given in the online session when making their decisions in the lab. Nevertheless, to remove any further concerns, we asked subjects at the end of the lab experiment how well they remembered the online experiment on a scale from 1 (not at all) to 7 (extremely well). We observe strong heterogeneity of reported recollection (mean answer = 4.04, $sd = 1.69$). We test whether the predictive value of personal norms stays robust when removing those who have a good recollection of the online experiment. To this end, we re-estimate our utility framework without subjects who claimed they had a good recollection (answers 6 or 7 on the Likert scale; see Table A3 in Appendix A.4). Here, we pool the PRIVATE and SOCIAL treatment together for reasons of statistical power.[27] Furthermore, we also take a more extreme approach, and keep only those who reported having trouble remembering the online experiment (answers below the midpoint of the Likert scale). As this strongly reduces the sample size, we only estimate our

[27] Note that the coefficients estimating the relation between personal norms and behavior do not differ between PRIVATE and SOCIAL treatment (see Table 2).

framework using the entire dataset. Our results remain robust in all regressions.

Most preferred action. Another potential concern is that, instead of answering according to their personal norms, subjects simply stated their preferences. Although we made a clear conceptual distinction that is rooted in the literature between one's preferences and personal norms (see Section 2), one might still argue that subjects do not actually care about their personal norms and may have answered according to what maximizes their utility function, leading to a spurious relation between our measure of personal norms and behavior. We address this concern through several design features and robustness checks. Our elicitation method is constructed to clearly follow the definition of personal norms and mitigate these concerns. Subjects rated the actions according to what they personally consider to be *correct* or *moral* behavior and not according to what they *want* or *prefer* to do. Moreover, they were rating the actions of hypothetical other individuals and were not informed that they would face these decisions in the lab. Nevertheless, as we cannot completely rule out this concern *a priori*, we turn to our dataset.

First, we look at our results from the main analysis. According to the alternative explanation, the personally most preferred action should also be the one that subjects chose in the lab. This would leave little scope for social norms and monetary payoff to bear weight in our analysis. However, both these predictors are robust and sizeable across our regressions. Furthermore, both predictors increase the predictive fit of our models, showing that they contribute together with personal norms to explaining behavior.[28]

Second, we directly check to what extent this alternative explanation is borne out by our data. While subjects should have always chosen the personally most appropriate action if they had rated actions according to their preferences, this actually happened in 52% of cases (similarly, in 47% of cases the action with the highest social appropriateness rating was chosen). Assuming some error, subjects could have chosen the action with the second or even the third highest personal appropriateness rating; however, we observe that they often opted for actions that ranked even lower, i.e., those that were supposedly among the least preferred. In particular, the actions ranked as fourth, fifth, and sixth were *each* chosen in 10% of cases.[29] Moreover,

[28]For each game, both in PRIVATE and SOCIAL, adding monetary payoff to a model with personal and social norms increases the predictive fit according to BIC and AIC, and significantly so according to the Log-likelihood ratio test. Likewise, adding social (personal) norms to a model with personal (social) norms and payoff also increases the predictive fit (see Result 4).

[29]To calculate the proportion of x ranked choice, we take into account decisions where subjects can choose

we re-estimate our utility framework by progressively excluding the actions with the highest, second highest, and third highest personal appropriateness rating from our dataset (see Table A4 in Appendix A.4). To maintain sufficient power, we join PRIVATE and SOCIAL together when excluding the action with the highest personal appropriateness rating, and pool the entire sample across games when excluding further actions. Personal norms remain a significant predictor across all degrees of exclusion.

Third, we check whether the personal norms ratings themselves are in line with the alternative explanation. If subjects care only about monetary payoffs and social norms, then the utility-maximizing action should always fall within the interval of the most profitable action and the action with the highest social appropriateness (see Appendix A.4 for more details). Yet, we observe that across our four games, the action with the highest personal appropriateness is strictly outside of that interval in 13% of cases.[30] We take this sample and re-estimate our utility framework. We find that personal norms remain a significant predictor, comparable in size with the values of our main regressions (see Table A5 in Appendix A.4). Taken together, our robustness results are inconsistent with the alternative explanation.

Average social norm rating. As we argued in the utility framework, we believe that a person will act upon her belief about the social norm, rather than upon the commonly recognized social norm (which she might fail to guess). Indeed, our dataset was conceived to obtain individual values for both the personal and social norms, allowing us to contrast the two. One might still argue, however, in favor of using the commonly held social norm. For example, since the social norm, and hence its correct guess, depends on the beliefs of others, it is possible that the inherent uncertainty might cause some to fail in their guess, although they actually also possess a good understanding of what is socially appropriate. Here, we take a different approach and — in line with Krupka and Weber (2013) — assume that people care about the commonly recognized social norm, which we calculate as the mean of all individual social appropriateness

rank x, i.e., cases in which there are at least x available ranks. We also employ an alternative check and look at the proportion of cases where subjects choose between differently ranked actions and decide for one which is strictly in the lower half (e.g., if there are 5 ranks then only 4th and 5th ranks correspond to the lower half). Again, we find strong evidence that subjects often go for lower ranked actions, as they choose an action from the lower half of ranks in 23% of cases.

[30]The percentage ranges from 10.9% in the TPP to 20.5% in the UG. Also in the additional games (see Section 4.4), the action with the highest personal appropriateness is frequently strictly outside of the interval reaching from the most profitable action to the action with the highest social appropriateness. Across the seven additional games this happens between 10.4% to 50.75% of the times.

ratings (see also, e.g., Gächter et al., 2013; Kimbrough and Vostroknutov, 2016). We repeat our complete regression analyses in Appendix A.4. All our results remain unaffected.

Order effect. As our experiment required the elicitation of both personal and social norms at the individual level, one might be concerned that our estimates are affected by the order of the norm elicitation, especially in the case of self-reported personal norms. To account for this potential concern, first, we randomized the norm elicitation order at the individual level. Second, we use the randomization to test whether the predictive value of a particular norm remains robust conditioning on the order (see Tables A10 and A11 in Appendix A.4). As this check decreases our sample size by half, we join the two treatments together. Both personal and social norms remain strongly predictive when the personal norm was elicited first (and subjects were still unaware of the second norm elicitation), and when it was elicited second.[31] Finally, we *directly* test whether the estimated norm coefficients in our regression analyses were affected by the order. To do so, we re-estimate our utility framework (both in PRIVATE and SOCIAL) while interacting the variables for the two norms with a dummy variable for the elicitation order (see Tables A12 and A13 in Appendix A.4). We find no significant interaction between the two norms and the order in any of the regressions. Altogether, our reported results stay robust.

4.4 Further evidence on personal and social norms

In this section, we report evidence from an additional lab experiment in which we test the following two questions with a different set of subjects ($n = 160$). First, while we have shown that our main results are robust, it would be reassuring for the empirical value of personal norms (as well as for our elicitation method) to show that the patterns of the two norms and in particular their relation are a robust finding. To investigate this, we elicit social and personal norms for our four games from a different sample, and compare them to the ones we obtained in our main sample. Second, it is unclear at this point whether the heterogeneity between personal and social norms only applies to our four games, or it also extends to other economic contexts. If so, this would give strong support for a broad applicability of our findings. Hence, we elicit the two norms for seven additional games and ten vignettes representing real-life economic

[31] As the estimates of the social norms coefficient depend on the treatment (see Result 3), we also inspect them by repeating the robustness check for PRIVATE and SOCIAL separately. Since this strongly reduces our sample size, we repeat the regressions only for the pooled dataset. We find that regardless of the treatment or order, social norm coefficient is always significant ($p < 0.027$ for the social norm coefficient in each of the four regressions).

situations. For more information about the procedure of these experiments, see Appendix A.5.

Replication. First, we examine the results for our four main games. The appropriateness ratings for social and personal norms again display strong and significant correlations, similar to those of our main sample. Correlation coefficients for this and the main sample are 0.72 and 0.72 in DG, 0.58 and 0.65 in DGT, 0.68 and 0.74 in UG, and 0.68 and 0.76 in TPP ($p < 0.001$ for each correlation in the new sample, Pearson product-moment correlation). Turning to the patterns of heterogeneity, we observe that the proportions of non-zero differences between the individual-level appropriateness ratings for personal and social norms are again large, and are rather similar to the main sample. The difference for this and the main sample is non-zero in 53.75% and 49.89% of cases in DG, 60.28% and 55.64% in DGT, 50.99% and 49.67% in UG, and 54.61% and 47.56% in TPP. Again, all proportions in the new sample are significantly different from 0 (the 99% asymptotic binomial confidence interval does not contain 0 in any comparison).[32] Finally, we check whether the distribution of personal and social norm ratings for each action in the four games differs across the two samples. For a total of 80 tests, we find that the two distributions differ only in a single case, revealing a very consistent pattern for both normative perceptions.[33] Overall, we observe a high level of consistency with the results from our main sample.

Result 5 *The distribution of personal and social norm ratings as well as their relation stay consistent in a replication with a different set of subjects.*

Additional games. We elicited personal and social norms in seven additional games: lying (die-roll) game, trust game, public-goods game, charitable giving game, charitable giving game with entitlement, dictator game with entitlement and ultimatum game with computer first move. Some of these games study important realms of economic behavior, not captured by our four games, such as lying (see Fischbacher and Föllmi-Heusi, 2013), trustworthiness (see Berg et al., 1995), or cooperation (see Ledyard, 1995), while others are variants of our four games

[32] We also compare the proportions by running a probit regression for each game with clustered standard errors at the individual level. The dependent variable is a dummy for whether the two ratings differ for a given action. As the independent variable, we use a dummy for whether the observation comes from the main or the new sample. We find that the difference is insignificant in DG, DGT, and UG ($p > 0.206$ for each comparison), and significant only in TPP ($p = 0.047$).

[33] We run Chi-squared tests with Monte Carlo-simulated p-values over 10.000 replications, and use the Bonferroni-Holm correction to account for multiple hypotheses testing at the game level for personal and social norms separately.

in which we introduce new motives with variations which are widely used in the literature, such as entitlement over the endowment (see Cherry et al., 2002), playing with a charity (see Eckel and Grossman, 1996), or eliminating intentions by randomly determining the first mover's choice (see Falk et al., 2008). In addition to these seven games, we also elicited personal and social norms in ten vignettes capturing real-life situations. These vignettes represent common economic interactions that people encounter in everyday life or at a workplace. For example, "a colleague working from home claims to have worked for more hours than she actually did", "your neighbor pays a painter under the table and thus pays no taxes", or "an employee of a firm calls in sick to prolong his holiday". A full description of the games and the list of all vignettes can be found in Appendix A.5.

Figure 2 depicts the differences in social and personal appropriateness ratings for the additional games and vignettes. The correlation between the two ratings is highest in the public-goods game (0.75) and lowest in the trust game (0.13). It takes the value of 0.53 in the dictator game with entitlement, 0.53 in the charitable giving game, 0.54 in the charitable giving game with entitlement, 0.56 in the ultimatum game with computer first move, and 0.65 in the lying game ($p = 0.004$ for the trust game, $p < 0.001$ for all other games; Pearson product-moment correlation). For the vignettes, we observe an overall correlation of 0.51 ($p < 0.001$), ranging from 0.1 to 0.59 for individual vignettes ($p < 0.013$ for 5 out of 10 vignettes, $p > 0.162$ for the remaining 5 vignettes). The proportion of non-zero differences between the two norms is again substantial in each game. In line with the findings above, it is the lowest in the public-goods game (47.87%) and the highest in the trust game (76.74%). In the other games, the proportion of non-zero differences is always higher then 50%. It takes the value of 50.25% in the lying game, 58.62% in the ultimatum game with computer first move, 60.35% in the charitable giving game, 61.41% in the charitable giving game with entitlement, and 62.42% in the dictator game with entitlement. Turning to the the vignettes, the proportion of non-zero differences ranges from 44.68% to 70.21% across the individual vignettes. All reported proportions are significantly different than 0 (the 99% asymptotic binomial confidence interval does not contain 0 in any comparison for both the games and the vignettes).

Overall, these data show that the presence of heterogeneity between personal and social norms is common to a wide array of economics interactions. Together with our main findings, this supports personal norms as a relevant predictor of behavior across a wide range of economic contexts.

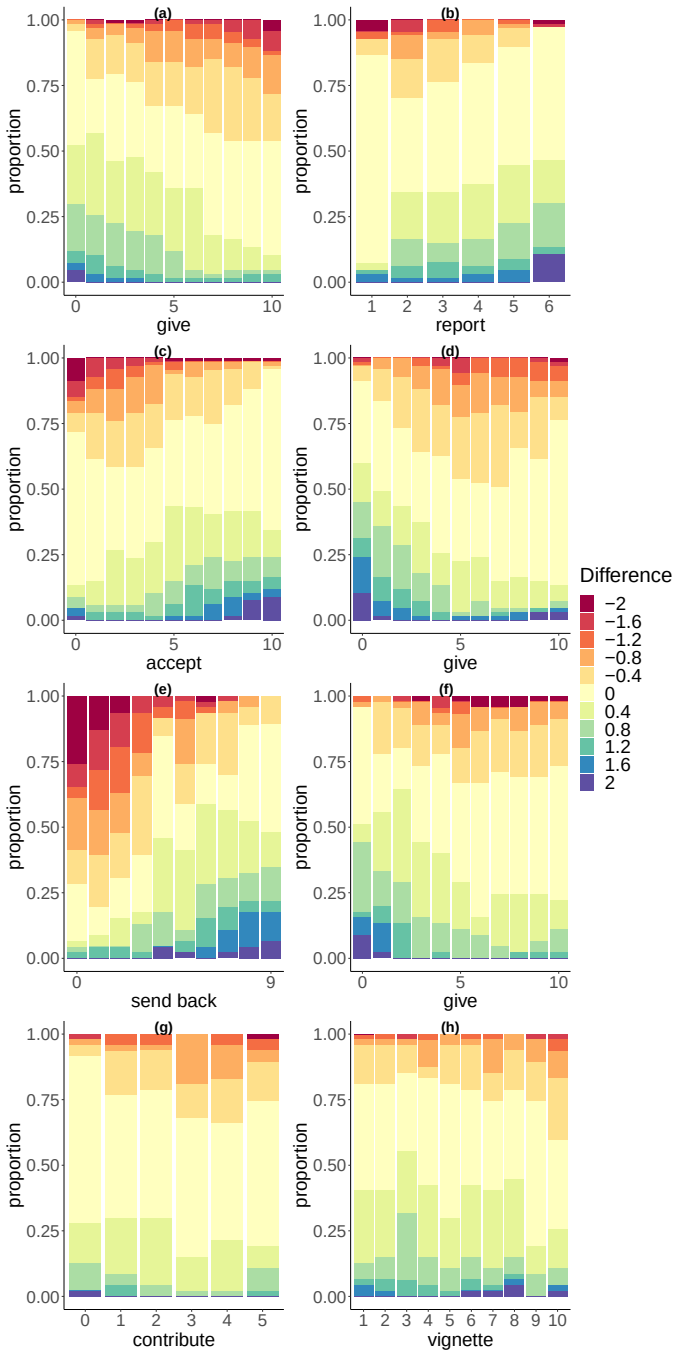


Figure 2: Individual difference between appropriateness ratings of social and personal norms in additional games.

Note: The difference is calculated by subtracting an individual's personal appropriateness rating from her social appropriateness rating. The proportion of each difference is displayed for each action in a given game. The following games are displayed: (a) charitable giving game, (b) lying game, (c) ultimatum game with computer first move, (d) dictator game with entitlement, (e) trust game, (f) charitable giving game with entitlement, (g) public-goods game, and (h) 10 different vignettes.

Result 6 *There is substantial individual heterogeneity between personal and social norms across seven additional games and ten vignettes describing real-life economic situations.*

5 Discussion and conclusion

In this study, we propose that people, in addition to caring about social norms and their monetary payoff, also care about personal norms. We offer a simple utility framework that captures these relations and design a novel two-part experiment to estimate it.

We establish that personal and social norms are related, but that there is substantial heterogeneity between the two at the individual level. We then estimate our framework and show robust evidence that personal norms — while taking social norms and monetary payoff into account — are strong predictors of economic behavior across four different economic games, both in a PRIVATE treatment where decisions are anonymous and in a SOCIAL treatment where social image concerns are made salient. In line with our predictions, the increase in social image concerns on average strengthens the relation between social norms and behavior; however, this does not come at the expense of personal norms. We show that our two-norms framework has higher predictive power in contrast to a framework where people only care about the social norm (see, e.g., Krupka and Weber, 2013) or the personal norm. Finally, we successfully replicate our findings regarding the relation and the heterogeneity between the two norms for our main games, and show that this heterogeneity exists also across seven additional games and a battery of vignettes that capture economic situations in everyday life and at a workplace.

The findings we present in this study offer strong evidence on the relevance of personal norms for economic behavior. They show that personal norms are powerful predictors of behavior in economic settings, and they support them as a key motive of economic decision-making. Given that we observe that personal norms are distinct from social norms across a large variety of games and vignettes, the implications of these findings are likely to extend to a wide array of real-world economic decisions, especially so as these are often embedded in complex environments that provide a great scope for different normative judgements to arise.

While our findings highlight the relevance of personal norms, it is important to stress that we do not belittle the role played by social norms. On the contrary, our results take both norms into account and provide insights on how the two norms interact and how they relate to behavior. In line with the existing literature, we find that social norms play an important role; however, the estimations of our framework with both norms have higher predictive power in contrast to frameworks that take only one of the two norms into account, showing that personal norms complement social norms in predicting behavior. In the few cases in which the

two-norm model does not outperform the others, interestingly, it is mostly the framework with only personal norms that prevails.

Apart from offering support to our utility framework, the findings from the model comparison imply that by ignoring personal norms and focusing only on social norms we are worse off in forecasting how people will behave in economic settings. This can have important implications for policy-makers. If people's behavior is co-determined by personal and social norms, an intervention targeting only social norms might lack effectiveness or even fail completely. This connects directly to the reasoning of Bicchieri and Dimant (2019), who argue that when trying to change norms — especially those with a long history of failed interventions (e.g., child marriage) — knowing the behavioral driver is crucial, otherwise the intervention might easily fail. For example, if the behavior is driven by personal normative beliefs, then targeting social normative beliefs will not yield the desired result. Similar challenges are also shared by companies when trying to promote desired behavior. Relying on employees to follow social norms in line with the company's organizational values might not be as effective if people are driven by their personal norms. Furthermore, as the misalignment of social and personal norms was found to be related to employee dissatisfaction (Burks and Krupka, 2012), promoting norms which diverge from employee's personal norms might even cause further frictions. Our findings provide evidence for the existence of dangerous pitfalls when designing normative interventions or shaping desired behavior within organizations, and they underscore the importance of understanding personal norms in these situations.

Besides from offering evidence that both types of norms shape behavior, we also shed light on how they interact and how the focus can be shifted to a particular norm. Our findings from the Social treatment indicate that increasing social image concerns enhances the importance of social norms for behavior. This supports our conjecture that situational factors can make a particular norm salient (see Bicchieri, 2005; Berkowitz and Daniels, 1964; Schwartz and Fleishman, 1978; Rutkowski et al., 1983; Cialdini et al., 1991). While we cannot dismiss the possibility that stronger manipulations might decrease the influence of personal norms, the findings from our Social treatment suggest that personal norms are rather robust (see Bicchieri, 2010), and "overriding" this motive is far from trivial.

Taken as a whole, our results imply that future research should consider personal norms when investigating normative prescriptions and their effect on economic behavior. This opens up important new questions, such as how personal norms develop over time and what leads to

incongruences with social norms. As we have shown that personal norms' relation to behavior is robust, it would be highly beneficial — from both a theoretical and a practical perspective — to understand whether and how we can shape them in the long run.

References

Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for truth-telling. *Econometrica 87*(4), 1115–1153.

Adena, M. and S. Huck (2020). Online fundraising, self-image, and the long-term impact of ask avoidance. *Management Science 66*(2), 722–743.

Agerström, J., R. Carlsson, L. Nicklasson, and L. Guntell (2016). Using descriptive social norms to increase charitable giving: The power of local norms. *Journal of Economic Psychology 52*, 147–153.

Ai, C. and E. C. Norton (2003). Interaction terms in logit and probit models. *Economics Letters 80*(1), 123–129.

Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *The Quarterly Journal of Economics 94*(4), 749–775.

Alempaki, D., G. Doğan, and Y. Yang (2021). Lying in a foreign language? *Journal of Economic Behavior & Organization 185*, 946–961.

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

Andreoni, J. and B. D. Bernheim (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica 77*(5), 1607–1636.

Andreoni, J. and J. Miller (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

Andreoni, J. and R. Petrie (2004). Public goods experiments without confidentiality: A glimpse into fund-raising. *Journal of Public Economics 88*(7-8), 1605–1623.

Ariely, D., A. Bracha, and S. Meier (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review 99*(1), 544–55.

Balliet, D., C. Parks, and J. Joireman (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations 12*(4), 533–547.

Banerjee, R. (2016). On the interpretation of bribery in a laboratory corruption game: Moral frames and social norms. *Experimental Economics 19*(1), 240–267.

Barr, A., T. Lane, and D. Nosenzo (2018). On the social inappropriateness of discrimination. *Journal of Public Economics 164*, 153–164.

Bartke, S., A. Friedl, F. Gelhaar, and L. Reh (2017). Social comparison nudges — Guessing the norm increases charitable giving. *Economics Letters 152*, 73–75.

Bašić, Z., A. Falk, and S. Quercia (2020). Self-image, social image, and prosocial behavior. *Mimeo*.

Bénabou, R., A. Falk, and J. Tirole (2019). Narratives, imperatives, and moral persuasion. *NBER Working Paper*.

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics 126*(2), 805–855.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and Economic Behavior 10*(1), 122–142.

Berkowitz, L. and L. R. Daniels (1964). Affecting the salience of the social responsibility norm: Effects of past help on the response to dependency relationships. *The Journal of Abnormal and Social Psychology 68*(3), 275.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. (2010). Norms, preferences, and conditional behavior. *Politics, Philosophy & Economics 9*(3), 297–313.

Bicchieri, C. and E. Dimant (2019). Nudging with care: The risks and benefits of social information. *Public Choice*, 1–22.

Bicchieri, C., E. Dimant, S. Gächter, and D. Nosenzo (2020). Observability, social proximity, and the erosion of norm compliance. *CESifo Working Paper, No. 8212*.

Bicchieri, C. and E. Xiao (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making 22*(2), 191–208.

Bilancini, E., L. Boncinelli, V. Capraro, T. Celadin, and R. Di Paolo (2020). "Do the right thing" for whom? An experiment on ingroup favouritism, group assorting and moral suasion. *Judgment and Decision Making 15*(2), 182–193.

Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg registration and organization online tool. *European Economic Review 71*, 117–120.

Boonmanunt, S., A. Kajackaite, and S. Meier (2020). Does poverty negate the impact of social norms on cheating? *Games and Economic Behavior 124*, 569–578.

Burks, S. V. and E. L. Krupka (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science 58*(1), 203–217.

Burnham, K. P., D. R. Anderson, and K. P. Huyvaert (2011). Aic model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral ecology and sociobiology 65*(1), 23–35.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American Economic Review*.

Cappelen, A. W., T. Halvorsen, E. Ø. Sørensen, and B. Tungodden (2017). Face-saving or fair-minded: What motivates moral behavior? *Journal of the European Economic Association 15*(3), 540–557.

Capraro, V., G. Jagfeld, R. Klein, M. Mul, and I. Van De Pol (2019). Increasing altruistic and cooperative behaviour with simple moral nudges. *Scientific Reports 9*(1), 1–11.

Cherry, T. L., P. Frykblom, and J. F. Shogren (2002). Hardnose the dictator. *American Economic Review 92*(4), 1218–1221.

Cialdini, R. B., C. A. Kallgren, and R. R. Reno (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in Experimental Social Psychology*, Volume 24, pp. 201–234. Elsevier.

Cialdini, R. B., R. R. Reno, and C. A. Kallgren (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology 58*(6), 1015.

Dana, J., R. A. Weber, and J. X. Kuang (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory 33*(1), 67–80.

Danilov, A. and D. Sliwka (2017). Can contracts signal social norms? experimental evidence. *Management Science 63*(2), 459–476.

Eckel, C. C. and P. J. Grossman (1996). Altruism in anonymous dictator games. *Games and Economic Behavior 16*(2), 181–191.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives 3*(4), 99–117.

Erkut, H. (forthcoming). Social norms and preferences for generosity are domain dependent. *Games and Economic Behavior*.

Ewers, M. and F. Zimmermann (2015). Image and misreporting. *Journal of the European Economic Association 13*(2), 363–380.

Falk, A. (2021). Facing yourself–a note on self-image. *Journal of Economic Behavior & Organization 186*, 724–734.

Falk, A., E. Fehr, and U. Fischbacher (2008). Testing theories of fairness – Intentions matter. *Games and Economic Behavior 62*(1), 287–303.

Falk, A. and U. Fischbacher (2006). A theory of reciprocity. *Games and Economic Behavior 54*(2), 293–315.

Falk, A. and F. Zimmermann (2018). Information processing and commitment. *The Economic Journal 128*(613), 1983–2002.

Fehr, E. and U. Fischbacher (2004). Third-party punishment and social norms. *Evolution and Human Behavior 25*(2), 63–87.

Fehr, E. and S. Gächter (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives 14*(3), 159–181.

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), 324–351.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The 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.

Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in disguise – An experimental study on cheating. *Journal of the European Economic Association 11*(3), 525–547.

Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior 6*(3), 347–369.

Gächter, S., L. Gerhards, and D. Nosenzo (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review 97*, 72–86.

Gächter, S., D. Nosenzo, and M. Sefton (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association 11*(3), 548–573.

Gächter, S. and J. F. Schulz (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature 531*(7595), 496–499.

Gert, B. and J. Gert (2020). The definition of morality. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/morality-definition/. Accessed: 2021-01-22.

Gneezy, A., U. Gneezy, G. Riener, and L. D. Nelson (2012). Pay-what-you-want, identity, and self-signaling in markets. *Proceedings of the National Academy of Sciences 109*(19), 7236–7240.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Grossman, Z. and J. J. Van Der Weele (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association 15*(1), 173–217.

Güth, W., R. Schmittberger, and B. Schwarze (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization 3*(4), 367–388.

Hair Jr, J., R. Anderson, R. Tatham, and W. Black (1995). *Multivariate Data Analysis* (3 ed.). New York: Macmillan.

Huck, S., D. Kübler, and J. Weibull (2012). Social norms and economic incentives in firms. *Journal of Economic Behavior & Organization 83*(2), 173–185.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1986). Fairness and the assumptions of economics. *Journal of Business*, S285–S300.

Katz, D., F. H. Allport, and M. B. Jenness (1931). Students' attitudes: A report of the syracuse university reaction study.

Kessler, J. B. and S. Leider (2012). Norms and contracting. *Management Science 58*(1), 62–77.

Kimbrough, E. O. and A. Vostroknutov (2016). Norms make preferences social. *Journal of the European Economic Association 14*(3), 608–638.

Krupka, E. and R. A. Weber (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology 30*(3), 307–320.

Krupka, E. L., S. Leider, and M. Jiang (2017). A meeting of the minds: Informal agreements and social norms. *Management Science 63*(6), 1708–1729.

Krupka, E. L. and R. A. Weber (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association 11*(3), 495–524.

Krysowski, E. and J. Tremewan (2020). Why does anonymity make us misbehave: Different norms or less compliance? *Economic Inquiry*.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. Kagel and A. Roth (Eds.), *Handbook of Experimental Economics*, pp. 111–194. Princeton: Princeton University Press.

Lindbeck, A., S. Nyberg, and J. W. Weibull (1999). Social norms and economic incentives in the welfare state. *The Quarterly Journal of Economics 114*(1), 1–35.

Marquaridt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics 12*(3), 591–612.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.

Michaeli, M. and D. Spiro (2015). Norm conformity across societies. *Journal of Public Economics 132*, 51–65.

O'Gorman, H. J. (1975). Pluralistic ignorance and white estimates of white support for racial segregation. *Public Opinion Quarterly 39*(3), 313–330.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives 14*(3), 137–158.

Romano, A. and D. Balliet (2017). Reciprocity outperforms conformity to promote cooperation. *Psychological Science 28*(10), 1490–1502.

Rutkowski, G. K., C. L. Gruder, and D. Romer (1983). Group cohesiveness, social norms, and bystander intervention. *Journal of Personality and Social Psychology 44*(3), 545.

Schram, A. and G. Charness (2015). Inducing social norms in laboratory allocation choices. *Management Science 61*(7), 1531–1546.

Schwartz, S. H. (1973). Normative explanations of helping behavior: A critique, proposal, and empirical test. *Journal of Experimental Social Psychology 9*(4), 349–364.

Schwartz, S. H. (1977). Normative influences on altruism. *Advances in Experimental Social Psychology 10*(1), 221–279.

Schwartz, S. H. and J. A. Fleishman (1978). Personal norms and the mediation of legitimacy effects on helping. *Social Psychology*, 306–315.

Selten, R. (1965). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review 97*(3), 999–1012.

te Velde, V. L. (2019). Heterogeneous norms: Social image and social pressure when people disagree. *Discussion Paper*.

Tremewan, J. and A. Vostroknutov (2020). An informational framework for studying social norms: An extended version. http://www.vostroknutov.com/pdfs/tv-normsframework.pdf. Accessed: 2021-01-27.

A Appendix A

This Appendix contains further details, tables and graphs which complement our analysis.

A.1 Attrition

As subjects participated in an online and a lab session that were about 4 weeks apart, we observe a certain attrition (24%). Here, we check whether attrition was systematic, as this might threaten the validity of our results. First, we check whether attrition is correlated with any of the observable characteristics elicited in the online study. Table A1 shows the results of a probit regression in which the dependent variable is a dummy equal to one if the subjects came to the lab and zero if the subject attrited. None of the observable characteristics predicts attrition.

	(1)
Female (=1)	0.028
	(0.048)
Siblings	-0.023
	(0.022)
Age	0.002
	(0.004)
Study (=1)	0.036
	(0.104)
Observations	330

Note: Estimation of probit model with dummy variable for whether a subject also participated in the lab session or only in the online session as the dependent variable, and sociodemographic variables collected in the online session as independent variables. Coefficients represent average marginal effects. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A1: Probit model for attrition on observable characteristics

Second, we go one step further and check whether the personal and social norm ratings differ between those who participated in the lab and the online session, and those who participated only in the online session. We compare the distribution of the two appropriateness ratings across the two samples for each action in the four games, for both personal and social norms. We run altogether 80 Chi-squared tests with Monte Carlo simulated p-values over 10.000 replications and use the Bonferroni-Holm correction to account for multiple hypotheses testing at the game level for personal and social norms separately. Only one out of the 80 tests turns out significant.

Thus, the norm ratings across the two samples are highly consistent. Altogether, the observed attrition does not seem to present an issue for the interpretation our results.

A.2 Estimation of the utility framework

To estimate our utility framework (Equation 1), we use a conditional (fixed-effect) logit choice model (see, e.g., Krupka and Weber, 2013; Gächter et al., 2013; Krupka et al., 2017). To estimate the model, we first reshape our dataset for each game.

For the DG, we expand each individual decision to the number of actions the subject in the role of dictator could choose from (give €0, €1, ..., €10; 11 observations in total). We then generate a new dependant variable which equals one if the subject chose the given action and zero if she did not. We regress this outcome on characteristics of that potential action, which are the three dependent variables from our utility framework. The first variable is the monetary payoff. In the DG, the monetary payoff is equal to the amount of euros a subject would receive by choosing the particular action. Here (as well as in the other games), we assume a linear restriction on the function $V(\cdot)$ from our utility framework, such that $V(\pi(a_k)) = \beta\pi(a_k)$. Hence, we estimate β which is the weight subjects place on monetary payoff. The second dependent variable is the social norm appropriateness rating assigned by the subject to that action. The third is the personal norm appropriateness rating assigned by the subject to that action. The regression takes into account that each of the 11 observations stems from one individual decision.

The same approach was taken for the other three games with necessary adjustments. In the DGT, there were eight potential actions, which translates into eight observations per decision. In the UG, receivers had eleven potential actions; hence, this translates into eleven observations per decision. To get the receivers' monetary payoff in the UG, we calculated their expected payoff for each rejection threshold (i.e., each potential action) using the distribution of all proposers' offers. Finally, in the TPP, each subject playing as a third-party made three decisions, as she had to indicate her punishment choice for each potential action of the dictator (strategy method). Each of these decisions consisted of three potential actions; hence, we expanded the dataset to 3 observations per decision, where each subject made 3 decisions.[34]

[34]During the first day of data collection, subjects in the TPP game were exposed to a non-obstructive software issue. To avoid any potential bias in our estimation, we do not include the data from the TPP game collected during the first day in the analysis. We also perform a robustness check in which we include this data and find that all reported results in the study remain robust to inclusion of this data.

A.3 Personal norms, social norms, and behavior

Here we report complementary information to our main results. Table A2 reports the estimation of our utility framework in the SOCIAL treatment. All coefficients on the personal norm ratings are significant across all games as well as in the pooled regression, confirming that Result 2 also holds in SOCIAL. The fact that personal norm coefficients are comparable with Table 1, and that social norm coefficients on average increase, reflects what we report in Result 3.

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.805***	0.146***	0.635***	0.676***	0.265***
	(0.118)	(0.039)	(0.176)	(0.138)	(0.031)
Social norm rating	2.149***	0.973***	0.922**	0.256	0.966***
	(0.381)	(0.223)	(0.363)	(0.240)	(0.134)
Personal norm rating	1.631***	0.691***	0.655*	1.076***	0.944***
	(0.335)	(0.206)	(0.340)	(0.253)	(0.130)
Observations	1,353	1,107	693	486	3,639

Note: Estimation of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table A2: Conditional logit estimation of choice determinants in SOCIAL treatment

A.4 Robustness checks

As described in Section 4.3, we provide four robustness checks for our results. Here, we report complementary information to those checks. In the first, we want to rule out consistency as a potential explanation of our results. The regressions reported in Table A3 confirm Result 2. Personal norms remain a strong and stable predictor of behavior. Regression (1) to (5) are performed with subjects who report a score below 6 when asked how well they remember the online experiment on a likert scale from 1 to 7. Data are pooled across the PRIVATE and SOCIAL treatment to guarantee enough power. In regression (6), we only include subjects that score below the midpoint of our scale. We only perform this regression pooling all our games together and not for each game separately, as the number of observations decreases significantly.

The second set of robustness checks is constructed to rule out that our results were due to subjects stating their preferences instead of answering according to their personal norms in our

	Memory < 6					Memory < 4
	DG	DGT	UG	TPP	All games	All games
	(1)	(2)	(3)	(4)	(5)	(6)
Monetary payoff	0.688***	0.226***	0.588***	0.799***	0.344***	0.343***
	(0.078)	(0.033)	(0.120)	(0.114)	(0.025)	(0.034)
Social norm rating	1.155***	0.746***	0.643**	0.390**	0.638***	0.333**
	(0.265)	(0.188)	(0.275)	(0.174)	(0.102)	(0.139)
Personal norm rating	1.462***	0.651***	0.961***	0.948***	0.987***	1.005***
	(0.241)	(0.160)	(0.264)	(0.185)	(0.098)	(0.135)
Observations	2,178	1,782	1,078	774	5,812	2,731

Note: Estimation of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as independent variables. The sample is restricted to subjects with a given score on the question of how well they remember the online session. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Conditional logit estimation of choice determinants for robustness check of consistency

elicitation. Table A4 shows the results of the regressions obtained by progressively excluding actions rated as personally most appropriate. We first exclude the most appropriate action (Top 1). For purposes of statistical power, we pool the data across Private and Social together. The coefficients of the personal norm rating remain significant across all but one game and highly significant for the pooled dataset. We then exclude the second and the third personally most appropriate actions (Top 2 and Top 3). As this strongly reduces the sample size, we only look at the pooled dataset. Again, the personal norm coefficient remains significant when excluding the second personally most appropriate action, and, although the sample is drastically reduced, it also remains significant when excluding the third personally most appropriate action (at the 10% level).

In table A5, we restrict the sample to the cases where the personal norm rating is not consistent with the alternative explanation. If subjects care only about monetary payoff and social norms, the most preferred action should never fall outside of the interval of the highest monetary payoff and the socially most appropriate action. In particular, in DG, DGT (and TPP), higher giving (punishment) strictly implies lower monetary payoff; hence, the action with the highest personal appropriateness rating should never be outside of the interval between zero giving (punishment) and the action with the highest social appropriateness, or, if there are more

Actions excluded	Top 1					Top 2	Top 3
	DG	DGT	UG	TPP	All games	All games	All games
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Monetary payoff	0.651***	0.177***	0.565***	0.893***	0.309***	0.348***	0.440***
	(0.102)	(0.032)	(0.133)	(0.239)	(0.029)	(0.043)	(0.078)
Social norm rating	1.550***	0.648***	0.554	0.090	0.665***	0.626***	-0.161
	(0.385)	(0.211)	(0.442)	(0.386)	(0.147)	(0.242)	(0.420)
Personal norm rating	0.561*	0.362	1.897***	1.285***	0.698***	0.823***	0.990*
	(0.320)	(0.229)	(0.497)	(0.455)	(0.155)	(0.275)	(0.577)
Observations	1,104	1,093	695	176	3,068	1,240	488

Note: Estimation of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as independent variables. The sample is restricted by progressively excluding actions with the highest, second highest, and third highest personal appropriateness rating. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Conditional logit estimation of choice determinants when excluding personally most preferred actions

such actions, the one with smaller payoff. In UG, we utilize the expected monetary payoff, and as the payoff is weakly decreasing in the minimum accepted offer, again the same argument holds. As the decrease is not strict, we also eliminate six decisions for which, due to payoff equivalence, one might choose the action which is above the one where social appropriateness is highest. As this limits the sample to 13% of our total observations, we pool all these decisions together. Also in this case, the coefficient of the personal norm rating remains significant in our regression.

	All games
	(1)
Monetary payoff	0.229***
	(0.051)
Social norm rating	0.637***
	(0.172)
Personal norm rating	0.773***
	(0.238)
Observations	1,013

Note: Estimation of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action as independent variables. The sample is restricted to cases which are incosistent with subjects caring only about social norms and monetary payoff: it consists of decisions where the personally most appropriate action is outside of the range defined by the socially most appropriate action and the payoff maximizing one, and it additionally excludes 6 decisions in UG where the personally most appropriate action falls outside of the interval but might still be accounted for by the alternative explanation. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Conditional logit estimation of choice determinants when restricting to decisions where personally most appropriate action is outside of the interval defined by socially most appropriate action and payoff maximizing action

In our third set of robustness checks, we use the average social norm rating for a given action instead of a subject's belief and re-run our complete analysis. In line with the literature using this approach (see, e.g., Krupka and Weber, 2013; Gächter et al., 2013; Krupka et al., 2017), we estimate a conditional (fixed-effect) logit choice model and calculate bootstrapped standard errors. More in detail, as the average social norm ratings may suffer from a sampling error, we bootstrap 500 replications to calculate the errors. For each replication, we resample (with replacement) from the norm rating data to calculate the average of the social norm for that particular replication, and then resample (with replacement) from our behavioral data to conduct the replication. Table A6 displays the results of these regressions for the Private treatment. This confirms Result 2, namely that personal norms are a strong and stable predictor of behavior.

In Table A7, we provide a robustness check of Result 3. The interaction between average social norm ratings (as constructed for this robustness check) and the Social treatment is significant for the DG and DGT (and at a 10% level for TPP), as well as for all games pooled together. Also, the interaction between average social norm ratings and the Private treatment

is insignificant in all regression models. Finally, we also report the estimations performed only for the SOCIAL treatment in Table A8. As expected, both personal and social norms ratings remain significant predictors of behavior, and the coefficients observed for personal norms remain comparable to those in the PRIVATE treatment.

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	1.179***	0.542	0.382	1.022***	0.520***
	(0.355)	(0.626)	(0.495)	(0.186)	(0.040)
Social norm rating (avg.)	2.649***	2.096	1.945	0.999***	1.101***
	(1.019)	(3.396)	(1.184)	(0.308)	(0.190)
Personal norm rating	0.986***	0.755***	0.721**	0.795***	0.880***
	(0.281)	(0.255)	(0.288)	(0.263)	(0.140)
Observations	1,397	1,143	704	504	3,748

Note: Estimation of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, average social appropriateness rating, and personal appropriateness rating of the action as independent variables. Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Conditional logit estimation of choice determinants in PRIVATE treatment using average social norm

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	1.357***	0.595	0.545	1.022***	0.512***
	(0.270)	(0.528)	(0.353)	(0.148)	(0.029)
Social norm rating (avg.)	3.161***	2.382	1.997*	0.999***	1.079***
	(0.799)	(2.863)	(1.029)	(0.323)	(0.190)
Personal norm rating	0.984***	0.757***	0.762***	0.795***	0.878***
	(0.286)	(0.257)	(0.289)	(0.260)	(0.138)
Social norm rating (avg.)	1.388***	2.141***	0.733	0.973*	1.832***
× Social	(0.414)	(0.501)	(0.836)	(0.540)	(0.313)
Personal norm rating	0.036	-0.102	-0.048	0.013	-0.067
× Social	(0.407)	(0.342)	(0.388)	(0.362)	(0.201)
Observations	2,750	2,250	1,397	990	7,387

Note: Estimations of conditional logit choice model with dummy variable for whether the subjects chose the action as dependent variable, and monetary payoff, social appropriateness rating, and personal appropriateness rating of the action, as well as an interaction term between personal and average social norm ratings and the Social treatment as independent variables. Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Conditional logit estimation of choice determinants interacted with Social treatment using average social norm

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	1.660***	0.682	1.432**	1.022***	0.500***
	(0.315)	(0.744)	(0.568)	(0.248)	(0.038)
Social norm rating (avg.)	5.352***	4.999	4.156***	1.972***	2.874***
	(0.926)	(3.915)	(1.294)	(0.444)	(0.259)
Personal norm rating	1.001***	0.661***	0.827***	0.808***	0.808***
	(0.303)	(0.238)	(0.319)	(0.245)	(0.148)
Observations	1,353	1,107	693	486	3,639

Note: Estimations of conditional logit choice model with dummy variable indicating whether the subjects chose the particular action as dependent variable, and monetary payoff, average social appropriateness rating, and personal appropriateness rating of the action as independent variables. Bootstrapped standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Conditional logit estimation of choice determinants in Social treatment using average social norm

In Table A9, we provide a robustness check of Result 4. Again, all Log-Likelihood comparisons favor the model that includes both personal and social norms over the two models that include only one of the two norms (all of the 20 Likelihood tests are highly significant). The same holds when looking at the AIC and the BIC, where the result becomes even stronger as the comparisons all support the model with two norms (19 out of 20 three-way comparisons), except from one case in which the personal norm model is favored.

	DG			DGT			UG			TPP			All games		
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)	(5a)	(5b)	(5c)
	Sn	Pn	Sn + Pn	Sn	Pn	Sn + Pn	Sn	Pn	Sn + Pn	Sn	Pn	Sn + Pn	Sn	Pn	Sn + Pn
Private															
Log-likelihood	-226.502	-231.757	-219.634	-234.683	-234.568	-227.035	-113.919	-117.866	-110.537	-118.050	-114.734	-108.396	-728.383	-715.751	-693.979
Likelihood test Sn vs. Sn+Pn	13.74 (<0.001)			15.3 (<0.001)			6.76 (0.009)			19.31 (<0.001)			68.82 (<0.001)		
Likelihood test Pn vs. Sn+Pn	-24.25 (<0.001)			15.07 (<0.001)			14.66 (<0.001)			12.68 (<0.001)			43.55 (<0.001)		
Akaike IC	457.0047	467.5138	445.268	473.366	473.137	460.070	231.837	239.732	227.075	240.010	233.468	222.792	1460.767	1435.503	1393.950
Bayesian IC	467.489	477.998	460.995	483.448	483.219	475.194	240.951	248.846	240.745	248.545	241.913	235.459	1473.225	1447.961	1412.636
Social															
Log-likelihood	-136.146	-186.443	-130.452	-194.441	-235.503	-188.423	-100.558	-110.965	-96.465	-107.921	-116.940	-99.611	-583.317	-687.040	-558.763
Likelihood test Sn vs. Sn+Pn	11.39 (<0.001)			12.04 (<0.001)			8.19 (0.004)			16.62 (<0.001)			49.11 (<0.001)		
Likelihood test Pn vs. Sn+Pn	111.98 (<0.001)			94.16 (<0.001)			29 (<0.001)			34.66 (<0.001)			256.55 (<0.001)		
Akaike IC	276.292	376.886	266.905	392.882	475.006	382.847	205.115	225.929	198.930	219.842	237.880	205.222	1170.634	1378.080	1123.526
Bayesian IC	286.713	387.306	282.535	402.901	485.025	397.875	214.198	235.011	212.553	228.214	246.253	217.781	1183.033	1390.479	1142.125

Note: Comparisons of log-likelihoods, Bayesian information criteria, and Akaike information criteria between models which include monetary payoff and average social norm as predictors (Sn columns), models which include monetary payoff and personal norm as predictors (Pn columns), and models which include monetary payoff and both average social norm and personal norm as predictors (Sn + Pn columns). Comparisons are accompanied by likelihood ratio tests (p value in brackets) which are reported for the estimation of all individual games (Columns 1a - 4c) and all games together (5a - 5c), separately for Private and Social treatment.

Table A9: Model comparison robustness check (avg. social norm rating)

Finally, we provide a fourth set of robustness checks to rule out a potential influence of the order in which social and personal norms were elicited on our results. First, we test whether the predictive value of the two norms remains robust across games and in the pooled dataset when looking at each elicitation order individually. Since this exercise halves our sample size, we join the PRIVATE and SOCIAL treatment together. We observe that the predictive value of both personal and social norms remains strong, regardless whether personal or social norms were elicited first (see Table A10 and Table A11, respectively). Second, we directly test whether the elicitation order affects any of the estimated norms coefficients in the estimations of our utility framework. Here, we re-estimate our regressions for all games and the pooled data set in PRIVATE and SOCIAL treatment separately, and add an interaction variable between each of the two norm variables and an order dummy variable (see Table A12 and Table A13, respectively). Also in this case, we find no evidence that the order of the norm elicitation affects our results.[35]

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.654***	0.236***	0.532***	0.859***	0.337***
	(0.094)	(0.042)	(0.150)	(0.151)	(0.030)
Social norm rating	1.172***	0.882***	0.711**	0.703***	0.750***
	(0.330)	(0.224)	(0.355)	(0.231)	(0.127)
Personal norm rating	1.337***	0.574***	0.602*	0.618**	0.822***
	(0.301)	(0.204)	(0.365)	(0.244)	(0.126)
Observations	1,342	1,098	627	513	3,580

Note: Estimations of conditional logit choice model with dummy variable indicating whether the subjects chose the particular action as dependent variable, and monetary payoff, average social appropriateness rating, and personal appropriateness rating of the action as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A10: Conditional logit estimation of choice determinants when personal norms were elicited first in the online experiment

[35] Given the caveats in interpreting interaction terms in non-linear models (see Ai and Norton, 2003), as a robustness check, we also re-estimate the regressions from Tables A12 and A13, with a linear probability model. Our conclusions stay the same when using a linear probability model. Out of ten models with two interaction terms each, only one interaction term in one model (interaction between personal norms and the order dummy variable in DG, SOCIAL treatment) yields a significant result (at the 10% level).

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.953***	0.244***	0.604***	0.842***	0.378***
	(0.125)	(0.044)	(0.144)	(0.150)	(0.033)
Social norm rating	1.788***	0.565**	0.795**	0.128	0.641***
	(0.385)	(0.244)	(0.359)	(0.239)	(0.135)
Personal norm rating	1.566***	0.920***	0.811**	1.189***	1.080***
	(0.323)	(0.213)	(0.317)	(0.243)	(0.125)
Observations	1,397	1,143	770	477	3,787

Note: Estimations of conditional logit choice model with dummy variable indicating whether the subjects chose the particular action as dependent variable, and monetary payoff, average social appropriateness rating, and personal appropriateness rating of the action as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A11: Conditional logit estimation of choice determinants when social norms were elicited first in the online experiment

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.734***	0.335***	0.517***	1.010***	0.444***
	(0.104)	(0.052)	(0.129)	(0.161)	(0.034)
Social norm rating	0.335	0.422	0.492	0.348	0.374**
	(0.498)	(0.349)	(0.470)	(0.360)	(0.188)
Personal norm rating	1.830***	1.033***	0.915**	0.878**	1.116***
	(0.485)	(0.312)	(0.425)	(0.353)	(0.180)
Social norm rating	0.751	0.341	0.185	0.441	0.252
× order	(0.622)	(0.412)	(0.707)	(0.442)	(0.237)
Personal norm rating	-0.855	-0.525	-0.259	-0.249	-0.363
× order	(0.641)	(0.430)	(0.691)	(0.460)	(0.247)
Observations	1,386	1,134	704	504	3,728

Note: Estimations of conditional logit choice model with dummy variable indicating whether the subjects chose the particular action as dependent variable, and monetary payoff, average social appropriateness rating, and personal appropriateness rating of the action, as well as an interaction term between personal and social norm rating and a dummy for the order of norms elicitation as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A12: Conditional logit estimation of choice determinants interacted with the order of norm elicitation in Private

	DG	DGT	UG	TPP	All games
	(1)	(2)	(3)	(4)	(5)
Monetary payoff	0.801***	0.145***	0.637***	0.733***	0.266***
	(0.117)	(0.039)	(0.176)	(0.145)	(0.031)
Social norm rating	2.429***	0.819***	1.132**	-0.013	0.900***
	(0.485)	(0.315)	(0.554)	(0.333)	(0.183)
Personal norm rating	1.256***	0.729**	0.647	1.369***	0.968***
	(0.441)	(0.295)	(0.482)	(0.330)	(0.174)
Social norm rating	-0.601	0.286	-0.369	0.747	0.129
× order	(0.593)	(0.409)	(0.730)	(0.531)	(0.252)
Personal norm rating	0.844	-0.015	-0.095	-0.880	-0.033
× order	(0.663)	(0.420)	(0.699)	(0.584)	(0.264)
Observations	1,353	1,107	693	486	3,639

Note: Estimations of conditional logit choice model with dummy variable indicating whether the subjects chose the particular action as dependent variable, and monetary payoff, average social appropriateness rating, and personal appropriateness rating of the action, as well as an interaction term between personal and social norm rating and a dummy for the order of norms elicitation as independent variables. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A13: Conditional logit estimation of choice determinants interacted with the order of norm elicitation in Social

A.5 Further evidence on personal and social norms

	No. of subjects	Games rated
Group 1	67	charitable giving game, dictator game with entitlement, lying game, ultimatum game with computer first move
Group 2	46	dictator game (DG), charitable giving game with entitlement, ultimatum game (UG), trust game
Group 3	47	dictator game with tax (DGT), third-party punishment game (TPP), public-goods game, vignettes

Table A14: Additional games.

The data for these additional experiments were collected during July and September 2017 at the BonnEconLab (University of Bonn). The experiment was programmed in z-Tree (Fischbacher, 2007). Subjects were recruited via hroot (Bock et al., 2014). Subjects were divided in three groups and each group faced the norm elicitation task for a subset of the games (see Table A14), after an unrelated experiment. Subjects had to rate the personal and social appropriateness of each action available to the individual in the game or the behavior described in the vignettes presented to them. As in our main experiment, subjects were incentivized to guess the most common social appropriateness ranking of the given action in the session, while no incentives were provided for stating one's personal appropriateness ranking. All games and vignettes are described below.

Charitable giving game. An individual is given €10 and has to decide how much to give to a charity. She can give any integer amount between €0 and €10. The charity was UNICEF, an internationally renowned organization dedicated to providing humanitarian and developmental aid to children worldwide.

Charitable giving game with entitlement. Also, here, an individual has to decide how much out of €10 she wants to give to UNICEF. However, in this case, she has earned the €10 by answering a questionnaire that lasted about 30 minutes.

Dictator game with entitlement. Similarly to the DG in the main analysis, an individual has €10 and can decide how much to give to another individual in the lab. Before this, however, both individuals had to work on a tiresome task for 20 minutes. They were given a series of matrices containing ones and zeros and had to count the number of zero in each matrix. The

one who managed to complete more of such matrices is given the €10 and the decision of how much thereof to give to the other individual.

Lying game. An individual is given a six-sided die and can privately roll it once. She gets the amount she reports in euros. In the case evaluated by participants, the individual rolls a one and can decide which number to report (from one to six).

Ultimatum game with computer first move. The structure of this game is analogous to that of the UG in our main experiment. An individual is given €10 and can offer any integer amount to another individual. If the individual accepts the offer, both get the proposed amounts. If she rejects it, they both earn nothing. However, the proposed amount is determined by a random device. The responder has to state the minimum offer she is willing to accept.

Trust game. An individual receives €4 and a second one €0. The first individual can send any integer amount to the second one. This amount is tripled. The second individual can then decide how much she wants to send back to the first one. In the case evaluated by participants, the first individual sends €3 and the second participant has to decide how much of the €9 she received she wants to send back.

Public-goods game. An individual is grouped together with three other people. They each receive €5. They then simultaneously decide how to allocate the €5 between a private and a common account. The individual can keep any money put in the private account, while the money in the common account is summed together, multiplied by two and shared equally amongst all members.

Vignettes.

1. "Your neighbor pays a painter under the table and thus pays no taxes."
2. "The chair of a commission at the university rejects a weak candidate to hire the daughter of a good friend."
3. "A woman who is moving out of her flat sells the couch she had paid €1500 for €2000."
4. "A freelancer eats at a restaurant with his friends for his birthday and deducts the check from his taxes."
5. "An employee of a firm calls in sick to prolong his holiday."
6. "A young man who finished university two years ago uses his old student card to use public transport."
7. "A customer at the supermarket notices that he has been given €5 too much, but keeps them."
8. "An acquaintance buys a highly polluting vintage car and drives it around just for fun in his free time."
9. "A colleague working from home claims to have worked for more hours than she actually did."
10. "An acquaintance who has purchased an insurance for his smartphone places his phone in water to get a new one just before the insurance expires."

B Appendix B

B.1 Instructions for the online experiment

These are the instructions used in the online experiment. The original text was in German and is available upon request.

Welcome

Welcome and thank you for your participation in this study.

This study is composed of two parts, today's online part (first part) and a part in the premises of the Cologne Laboratory for Economic Research (second part). This online part will take about 30-45 minutes. Please be aware that you need to complete this online part to take part in the second part. You will receive an email to remind you of this.

Please complete this part in one sitting, undisturbed and concentrated. If possible, please use a computer or a tablet. Please avoid other disturbances and complete this study alone. We reserve the right to exclude participants from the experiment who do not complete the study carefully.

All your decisions will be used only for scientific purposes and for determining your payment.

You will get a fixed payment of €8 for participation after you have completed both parts of the study. You have the opportunity to earn a further amount of money during this first part as well as during the second part. For this reason, please read the following instructions carefully.

The further earnings from the online part as well as the earnings from the part in the Cologne Laboratory for Economic Research and the fixed payment of 8€ will be paid out after the part in the Cologne Laboratory.

Please click on the arrow below to start with the study.

Code

In order to guarantee your payment, you have to generate a code below. You will generate the exact same code in the second part of the experiment. We will use your code to complete your payment anonymously.

Please insert your personal code in lower-case letters and without accents or other special symbols.

The code is composed by the following components:

SECOND letter of your own name

FIRST letter of your mother's name (if unknown insert "***")

FIRST letter of your father's name (if unknown insert "***")

SECOND letter of the name of your birthplace (if unknown insert "***")

Day of your birthday (e.g., 15 for 15/07 or 08 for 08/03)

Please type in the code in small letters and without accents or other special symbols.

Please do not use any umlaut. Write a instead of ä, o instead of ö and u instead of ü.

Example: Max Mustermann, son of Lisa and Paul, born in Bonn on the 27/04 the resulting code would be alpo27.

This online part is composed of three parts. You will obtain the corresponding instructions before each part and then complete that part.

Part 1

(*Elicitation of personal norms*)

In this part of the study, you will read the description of different situations. In each situation there is one person who has to make a choice between different actions.

After you read the description of each situation, you have to evaluate the different actions amongst which the person in that situation can choose from. For each action, evaluate according to your own opinion and independently of the opinion of others, whether it is appropriate or not to choose it. "Appropriate" behavior means the behavior that you personally consider to be "correct" or "moral". The standard is, hence, your personal opinion, independently of the opinion of others.

We kindly ask you to answer as precisely as possible with your own honest opinion. There is no right or wrong answer; you will not get any additional payment for your answers in this part.

Overall there are four different situations for which you have to evaluate the possible actions. To show you how the different actions can be evaluated we now give you an example.

Example

Person A is sitting in a cafe near the university. Person A notices that another person has left his wallet on the table. Person A has to decide what to do. Person A has to choose from four possible actions:

- Take the wallet and keep it;
- Ask other guests if the wallet belongs to one of them;
- Leave the wallet there;
- Give the wallet to the manager of the cafe.

For each action evaluate according to your own personal opinion and independently of the opinion of others, whether it is appropriate or not to choose it. “Appropriate” behavior means the behavior that you personally consider to be “correct” or “moral”.

You can choose from a scale with six points

- Very inappropriate
- Inappropriate
- Rather inappropriate
- Rather appropriate
- Appropriate
- Very appropriate

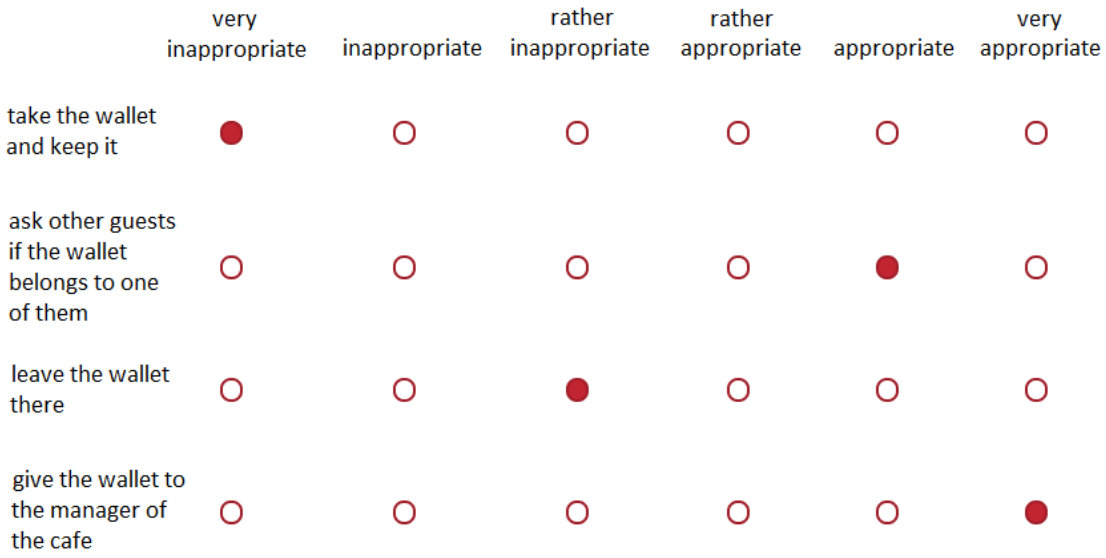
You will evaluate the actions using a table. To evaluate the behavior you have to mark the corresponding option. Please give an evaluation for each of the actions.

Assume, for example, that you evaluate

- Taking and keeping the wallet as very inappropriate,
- Asking other guests if the wallet belongs to them as appropriate,

- Leaving the wallet there as rather inappropriate,
- Giving the wallet to the manager of the cafe as very appropriate

You would insert the following evaluations.



After clicking on "next", the description of the actual situations that you have to evaluate will follow.

Description of the situations

Dictator game

In a study conducted at the economic laboratory, Person A is randomly matched with another participant, Person B. The matching is anonymous, hence no participant will ever learn about the identity of the other participants.

In this study, Person A takes a decision. Person B knows which decision Person A has to take. Person B also knows which consequences this decision has for the monetary payment and will know which decision Person A has taken.

Person A's decision

Person A gets €10 at the beginning of the task. Person A can then give any amount of this €10 to Person B.

Person A can, for example, give €0 to Person B. Person A would get €10 and Person B €0. Person A could also give €10. Person A would then get €0 and Person B €10. Similarly, Person A could give €1, €2, €3, ... or €9.

At first, both participants will take a decision in the role of Person A. This means that both will indicate how many euros they would give to Person B, in case they would be assigned the role of Person A. After both participants have taken their decision, they will learn who was assigned the role of Person A and of Person B. Both participants are paid according to the role assignment and the taken decision.

Please evaluate the possible actions of Person A.

Dictator game with tax

In a study conducted at the economic laboratory, Person A is randomly matched to another participant, Person B. The matching is anonymous, hence no participant will ever learn about the identity of the other participants.

In this study, Person A takes a decision. Person B knows which decision Person A has to take. Person B also knows which consequences this decision has for the monetary payment and will know which decision Person A has taken.

Person A's decision

Person A gets €12 at the beginning of the task. Person B gets €0. Person A can then send an amount of this €12 to Person B. Person B gets €0.90 for each €1.50 Person A sends to him. Hence, 40% of the amount sent gets lost.

Person A can, for example, send €0 to Person B. Person A would get €12 and Person B €0. Person A could also send €12. Person A would then get €0 and Person B €7.20. Similarly, Person A could send €1.50, €3, €4.50, ... or €10.50. You can find an overview of the possible actions and the corresponding earnings here:

At first, both participants take a decision in the role of Person A. This means that both have to indicate how many euros they would send to Person B, in case they would be assigned the

A sends	€0	€1.50	€3	€4.50	€6	€7.50	€9	€10.50	€12
hence Person A and Person B earn:									
A earns	€12	€10.50	€9	€7.50	€6	€4.50	€3	€1.50	€0
B earns	€0	€0.90	€1.80	€2.70	€3.60	€4.50	€5.40	€6.30	€7.20

role of Person A. After both participants have taken their decision, they will learn who was assigned the role of Person A and of Person B. Both participants are paid according to the role assignment and the taken decision.

Please evaluate the possible actions of Person A.

Ultimatum game

In a study conducted at the economic laboratory, Person A is randomly matched to another participant, Person B. The assignment is anonymous, hence no participant will ever learn about the identity of the other participants.

In this study, Person A and Person B take decisions simultaneously. Both know which decision the other has to take. They also know which consequences this decision has for the monetary payment and will know in the end which decision the other has taken. Here is a description of Person A's and Person B's decisions.

Person A gets €10 at the beginning of the task. Person B gets €0. Person A and Person B then take a simultaneous decision.

Person A's decision

Person A can propose any amount of the €10 to Person B. Person A hence decides how much of the €10 he wants to propose to Person B.

Person B's decision

Person B decides which proposals he is ready to accept. The two participants get the stipulated amounts only if Person B accepts the offer. If he rejects the offer, both get €0.

For this purpose, Person B chooses an amount between €0 and €10. This amount is the lowest proposal that Person B is still ready to accept. All proposals that are equal to or higher than this amount are accepted by Person B. All proposals that are lower than this amount are rejected by Person B.

Since the decisions are taken simultaneously, Person A does not know what the minimal amount of money Person B is willing to take at the point of his decision. Similarly, Person B does not know how much money Person A will actually propose at the point of his decision.

For example, Person B could accept proposals starting from 2 €. Proposals of €0 and €1 would be rejected. All other proposals would be accepted. Person B could also accept proposals starting from €8. Then, only proposals of €8, €9 or €10 would be accepted and all other offers would be rejected.

Please evaluate the possible actions of Person B.

Third-party punishment game

In a study conducted at the economic laboratory, Person C is randomly matched to two another participants, Person A and Person B. The matching is anonymous, hence no participant will ever learn about the identity of the other participants.

In this study, Person C and Person A take a decision. All three participants know which decision Person A and Person C have to take. They also know which consequences these decisions have for the monetary payment and will know which decisions have been taken.

Person A gets €10 at the beginning of the task. Person B gets €0. Person C gets €5.

Person A's decision

Person A can give Person B €0, €2, or €5 of his €10. Person A could give Person B €0. Then, Person A would get €10 and Person B €0. Person A could also give Person B €5. If Person A gave €5, then he would get €5 and Person B would get €5 as well. If Person A gave €2, then he would get €8 and Person B €2.

Person C's decision

Person C can assign deduction points to Person A depending on his decision. Person C can assign 0, 1 or 2 deduction points to Person A. The earnings of Person C are reduced by €1 and Person A by €3 for each deduction point assigned. The earning of Person A cannot, however, go below €0. This means that his earnings can be reduced only until €0. The assignment of deduction points has no consequence for Person B.

If Person C, for example, assigned 0 deduction points, then neither the earnings of Person C nor

those of Person A would be reduced. If Person C assigned 1 deduction point, then his earnings would be reduced by €1 and those of Person A by €3. If Person C assigned 2 deduction points, then his earnings would be reduced by €2 and those of Person A by €6.

Person C has to indicate how many deduction points he would assign Person A for each of his possible decisions (€0, €2, or €5). Only the decision of Person C that corresponds to the actual decision of Person A is implemented.

Example: Person A gives €2 to Person B. Person C indicated that in this case he would assign him 1 deduction point. Then, Person A would get a deduction of €3 and Person C of €1. In this case Person A would hence get (€8 - €3 =) €5, Person B €2, and Person C (€5 - €1 =) €4.

One of the participants is assigned the role of Person A. The other two both at first take a decision in the role of Person C. Both indicate how many deduction points they would assign to Person A in case they were Person C. The two participants will learn who was assigned the role of Person B and who to that of Person C only after they made their decision. Participants are paid according to role assignment and the decisions taken.

(1) Assume Person A decides to give Person B €0. He, hence, keeps €10 while Person B gets €0. Please evaluate the possible actions of Person C.

(2) Assume Person A decides to give Person B €2. He, hence, keeps €8 while Person B gets €2. Please evaluate the possible actions of Person C.

(3) Assume Person A decides to give Person B €5. He, hence, keeps €5 while Person B gets €5. Please evaluate the possible actions of Person C.

Elicitation

(*After each game description, subjects where first reminded of their task and then had to fill out the elicitation table for normative ratings. Here we show an example of the elicitation table from DG.*)

For each action, evaluate according to your own personal opinion and independently of the opinion of others, whether it is appropriate or not to choose it. "Appropriate" behavior means the behavior that you personally consider to be "correct" or "moral".

	very inappropriate	inappropriate	rather inappropriate	rather appropriate	appropriate	very appropriate
give €0	○	○	○	○	○	○
give €1	○	○	○	○	○	○
give €2	○	○	○	○	○	○
give €3	○	○	○	○	○	○
give €4	○	○	○	○	○	○
give €5	○	○	○	○	○	○
give €6	○	○	○	○	○	○
give €7	○	○	○	○	○	○
give €8	○	○	○	○	○	○
give €9	○	○	○	○	○	○
give €10	○	○	○	○	○	○

Part 2

(*Elicitation of social norms*)

In the following you will read the description of different situations. In each situation there is one person who has to make a choice between different actions.

After you have read the description of each situation, you have to evaluate the different actions amongst which the person in the situation can choose from. For each action, evaluate according to the opinion of the society and independently of your own opinion, whether it is appropriate or not to choose it. "Appropriate" behavior means the behavior that you consider most people would agree upon as being "correct" or "moral". The standard is, hence, not your personal opinion, but your assessment of the opinion of the society. We kindly ask you to answer as

precisely as possible.

In this part, you can earn up to €12 on top of your participation fee of €8, depending on your answers. The answers of the other participants will influence your payment in this part.

At the end of the study, we will determine for each action in each situation which answer most of the other participants gave. You will obtain €0.30 for each action for which you gave the same answer as most of the other participants.

Your payment is determined in the following way: you will evaluate the possible actions of a person according to the opinion of the society in 4 different situations. For each action in each situation the following holds: if your evaluation is exactly the same as the answer of most of the other participants, you will earn money. For each match you get €0.30. This means that you can earn up to €12 in addition to the fixed participation fee of €8. If, on the contrary, you never give the same answer as most of the other participants, then you will earn no money in this task. If, for example, you give the most frequent answer for 10 actions, you get €3 for this task.

Note: only the answers of other participants in this part count. All other participants have received the same instructions. Also, they get €0.30 for each action for which they give the same answer as most other participants.

Overall there are four different situations for which you have to evaluate the possible actions. To show you how the different actions can be evaluated, we now give you an example.

Example

Person A is sitting in a cafe near the university. Person A notices that another person has left his wallet on the table. Person A has to decide what to do. Person A has to choose from four possible actions:

- Take the wallet and keep it;
- Ask other guests if the wallet belongs to one of them;
- Leave the wallet there;
- Give the wallet to the manager of the cafe.

For each action, evaluate according to the opinion of the society and independently of your own opinion, whether it is appropriate or not choose it. “Appropriate” behavior means the behavior that you consider most people would agree upon as being “correct” or “moral”. Note: you earn €0.30 for each action for which your answer matches the most frequent answer of the other participants in this second part.

You can choose from a scale with six points

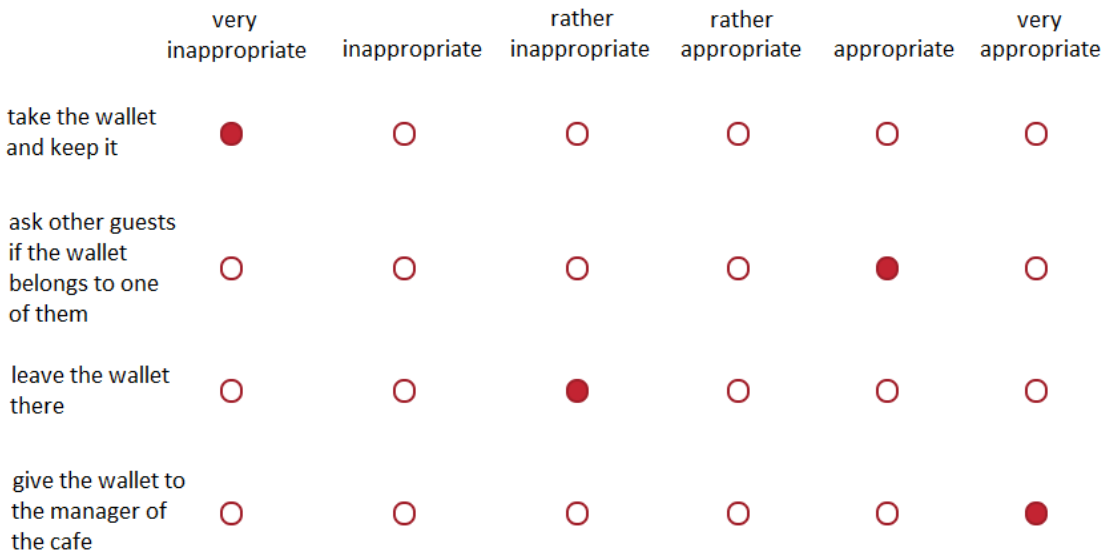
- Very inappropriate
- Inappropriate
- Rather inappropriate
- Rather appropriate
- Appropriate
- Very appropriate

You will evaluate the actions using a table. To evaluate the behavior you have to mark the corresponding option. Please give an evaluation for each of the actions.

Assume, for example, that you evaluate

- Taking and keeping the wallet as very inappropriate,
- Asking other guests if the wallet belongs to them as appropriate,
- Leaving the wallet there as rather inappropriate,
- Giving the wallet to the manager of the cafe as very appropriate.

You would insert following evaluations.



Assume the other participants gave the following evaluations. The table below shows for each action the percentage of other participants who gave a given evaluation. Obviously, you will not get this information in the actual situations. This example should help you understand how you can earn additional money.

Action	Very inappropriate	Inappropriate	Rather inappropriate	Rather appropriate	Appropriate	Very appropriate
Take the wallet and keep it	50%	30%	15%	5%	0%	0%
Ask other guests, if the wallet belongs to one of them	0%	5%	10%	40%	25%	20%
Leave the wallet there	15%	20%	40%	20%	0%	5%
Give the wallet to the manager of the cafe	0%	0%	0%	10%	30%	60%

How much additional money (in cent) would you get for this situation? (If, for example, the correct answer is €1.5, then write 150.)

After you have answered this question, the description of the actual situation that you have to evaluate will follow.

Description of the situations

Repetition of the situation descriptions (*see above*).

Elicitation

(*After each game description, subjects were first reminded of their task and then had to fill out the elicitation table for normative ratings.*)

For each action, evaluate according to the opinion of the society and independently of your own opinion, whether it is appropriate or not choose it. "Appropriate" behavior means the behavior that you consider most people would agree upon as being "correct" or "moral". Note: you earn €0.30 for each action for which your answer matches the most frequent answer of the other participants in this second part.

(*The elicitation tables were the same as when eliciting personal norms (see above).*)

B.2 Instructions for the laboratory experiment

These are the instructions used in the laboratory experiment. The original text was in German and is available upon request.

Welcome

Welcome to the second part of the study!

Today, you will take part in the second part of this study. You have already completed the first part online. You will be able to earn money in addition to the fixed amount of €8 and the amount you earned during the online study.

The size of this additional amount depends on your decisions, the decisions of other participants, and chance. Thus, please read the instructions carefully.

Please avoid any conversation with your neighbors. Switch off your mobile phone and remove any item you do not need for the study from your table. In case you have questions, raise your hand and we will answer your question at your seat.

Code

Please insert your code from the online study below so that we can carry out your payment correctly at the end of the study.

Reminder: The code is composed by the following components:

SECOND letter of you own name

FIRST letter of your mothers name (if unknown insert "***")

FIRST letter of your fathers name (if unknown insert "***")

SECOND name of your birthplace (if unknown insert "***")

Day of your birthday (e.g., 15 for 15/07 or 08 for 08/03)

Please type in the code in small letters and without accents or other special symbols.

Please do not use any umlaut. Write a instead of ä, o instead of ö and u instead of ü.

B.2.1 Instructions for experimental games

Instructions for PRIVATE treatment

Today's study is composed of four tasks. The tasks will be presented in a random order. You will receive the respective instructions before each task, and can then work on the task.

In these tasks you will be matched with other participants. You and other participants will take decisions during these tasks. You can be matched with each participant only once – it cannot happen that you are assigned to the same participant in two different tasks.

One of the tasks will be randomly selected for the payment of today's study. Since you will not know which task will be chosen until the end of the study, please go through the tasks carefully. At the end of this session you will receive the sum you earned during the whole study (€8 participation fee as well as the money from the online study and your payment from today's study) in cash.

Instructions for SOCIAL treatment

Today's study is composed of four tasks. The tasks will be presented in a random order. You will receive the respective instructions before each task, and can then work on the task.

In these tasks you will be matched with other participants. You and other participants will take decisions during these tasks. You can be matched with each participant only once — it cannot happen that you are assigned to the same participant in two different tasks.

One of the tasks will be randomly selected for the payment of today's study. Since you will not know which task will be chosen until the end of the study, please go through the tasks carefully. At the end of this session, you will receive the sum you earned during the whole study (€8 participation fee as well as the money from the online study and your payment from today's study) in cash.

When all participants in the session have completed the tasks, everyone will have to stand up (so that all participants can hear and see each other). An assistant will call the participants one after the other. Each participant will have to say his name and tell the other participants which choices he made in the tasks. For this purpose, a text will be displayed on your screen and you will have to read it verbatim. This means that all other participants will know your name and all the choices you have made in the tasks.

(*At the top of the decision screen in each of the four games, the following text was displayed:*)

Reminder: When you are done with all the tasks, you will have to stand up and tell all other participants which decision you made in this and the other tasks.

(*After the four games:*)

All participants have completed all tasks. Please stand up and wait until an assistant calls your cabin number. When you hear your cabin number, please read the following text verbatim.

Games

(*The four games were titled: task A, task B, task C, and task D.*)

Dictator game

In this task, you will be randomly matched to another participant. You will not find out neither before nor after the study who the other participant is.

You and the other participant will be assigned one of two roles: Person A or Person B.

Person A's decision

Person A gets €10 at the beginning of the task. Person A can then give any amount of this €10 to Person B. Person A can, for example, give €0 to Person B. Person A would get €10 and Person B €0. Person A could also give €10. Person A would then get €0 and Person B €10. Similarly, Person A could give €1, €2, €3, ... or €9.

At first, you and the other participant will both take a decision in the role of Person A. This means that you will indicate how many euros you would give to Person B, in case you would be assigned to the role of Person A. Both of you will learn which role you have been assigned (Person A or Person B) only at the end of the study. The earnings of both participants are calculated according to the assignment of roles and the decision taken by Person A.

Before you take your decision on the next page, please answer the following two questions.

1. How much does Person A earn, if Person A gives €3 to Person B?
2. How much does Person A earn, if Person A gives €1 to Person B?

Dictator game with tax

In this task, you will be randomly matched to another participant. You will not find out neither before nor after the study who the other participant is.

You and the other participant will be assigned one of two roles: Person A or Person B.

Person A's decision

Person A gets €12 at the beginning of the task. Person B gets €0. Person A can then send an amount of these €12 to Person B. Person B gets €0.90 for each €1.50 Person A sends to him. Hence, 40% of the amount sent gets lost.

Person A can, for example, send €0 to Person B. Person A would get €12 and Person B €0. Person A could also send €12. Person A would then get €0 and Person B €7.20. Similarly, Person A could send €1.50, €3, €4.50, ... or €10.50. You can find an overview of the possible actions and the corresponding earnings here:

At first, you and the other participant will both take a decision in the role of Person A. This means that you will indicate how many euros you would send to Person B, in case you would

A sends	€0	€1.50	€3	€4.50	€6	€7.50	€9	€10.50	€12
hence Person A and Person B earn:									
A earns	€12	€10.50	€9	€7.50	€6	€4.50	€3	€1.50	€0
B earns	€0	€0.90	€1.80	€2.70	€3.60	€4.50	€5.40	€6.30	€7.20

be assigned to the role of Person A. Both of you will learn which role you have been assigned (Person A or Person B) only at the end of the study. The earnings of both participants will be calculated based on the assignment of roles and the decision taken by Person A.

Before you take your decision on the next page, please answer the following two questions.

1. How much do Person A and Person B earn, if Person A sends €1.50 to Person B?
2. How much do Person A and Person B earn, if Person A sends €9 to Person B?

Ultimatum game

In this task, you will be randomly matched to another participant. You will not find out neither before nor after the study who the other participant is.

One participant is randomly assigned to the role of Person A and the other to the role of Person B.

Person A gets €10 at the beginning of the task. Person B gets €0. Person A and Person B then take a simultaneous decision.

Person A's decision

Person A can propose any amount of €10 to Person B. Person A hence decides how much of the €10 he wants to propose to Person B.

Person B's decision

Person B decides which proposals he is ready to accept. The two participants get the stipulated amounts only if Person B accepts the offer. If he rejects the offer, both get €0.

For this purpose, Person B chooses an amount between €0 and €10. This amount is the lowest proposal that Person B is still ready to accept. All proposals that are equal to or higher than this

amount are accepted by Person B. All proposals that are lower than this amount are rejected by Person B.

Person A does not know what the minimal amount of money Person B is willing to accept at the point of his decision. Similarly, Person B does not know how much money Person A will actually propose at the point of his decision.

For example, Person B could only accept proposals starting from €2. Proposals of €0 and €1 would be rejected. All other proposals would be accepted. Person B could also only accept proposals starting from €8. Then, only proposals of €8, €9, or €10 would be accepted and all other offers would be rejected.

(*Person A's text*)

You were assigned to the role of Person A. The other participant was assigned to the role of Person B.

(*Person B's text*)

You were assigned to the role of Person A. The other participant was assigned to the role of Person B.

(*Text for both participants*)

1. How much would Person A and Person B earn, if Person A offers Person B €4 and Person B accepts the offer?
2. How much would Person A and Person B earn, if Person A offers Person B €2 and Person B ...

   a ... accepts the offer?

   b ... rejects the offer?

Third-party punishment game

In this task, you will be randomly matched to two other participant. You will not find out neither before nor after the study who these other participants are.

One participant will be assigned to the role of Person A, another one to the role of Person B and a third one to the role of Person C. Person A gets €10 at the beginning of the task. Person B gets €0. Person C gets €5.

Persons A's decision

Person A can give Person B €0, €2, or €5. Person A could give Person B €0. Then, Person A would get €10 and Person B €0. Person A could also give Person B €5. If Person A gave €5, then he would get €5 and Person B would get €5 as well. If Person A gave €2, then he would get €8 and Person B €2.

Person C's decision

Person C can assign deduction points to Person A depending on his decision. Person C can assign 0, 1 or 2 deduction points to Person A. The earnings of Person C are reduced by €1 and Person A by €3 for each assigned deduction point. The earnings of Person A cannot, however, go below €0. This means that his earnings can be reduced only until €0. The assignment of deduction points has no consequence for Person B.

If Person C assigned 0 deduction points, for example, then neither the earnings of Person C nor those of Person A would be reduced. If Person C assigned 1 deduction points, then his earnings would be reduced by €1 and those of Person A by €3. If Person C assigned 2 deduction points, then his earnings would be reduced by €2 and those of Person A by €6.

Example: Person A gives €2 to Person B. Person C indicated that in this case he would assign him 1 deduction point. Then, Person A would be deducted €3 and Person C €1. In this case, Person A would hence get (€8 - €3 =) €5, Person B €2 and Person C (€5 - €1 =) €4.

Person B does not make any decision in this task.

(*Person A's text*)

You have been assigned to the role of Person A.

The other two participants have been assigned the role of Person B and Person C. At first, both of the other participants will take a decision in the role of Person C. Both will indicate how many deduction points they would assign to you (Person A) in case they were Person C. The other participants will learn only at the end of the experiment which role they were assigned to: one of them Person B and the other one Person C. The earnings for all participants will be

calculated based on this assignment of roles and the decisions taken.

The two other participants have to indicate how many deduction points they would assign to you for each of your possible decisions (€0, €2 , or €5), in case they were assigned to the role of Person C. Only the decision of Person C that corresponds to your actual decision will be implemented.

(*Person B's and Person C's text*)

One of the participants was assigned to the role of Person A. You and the remaining participant, who was not assigned to the role of Person A, will at first both take a decision in the role of Person C. You will both indicate how many deduction points you would assign to Person A in case you were Person C. You will both learn only at the end of the experiment which role you were assigned to: one of you Person B and the other one Person C. The earnings for all participants will be calculated based on this assignment of roles and the decisions taken.

You have to indicate how many deduction points you would assign for each of the possible decisions of Person A (€0, €2, or €5), in case you were assigned to the role of Person C. Only the decision of Person C that corresponds to the actual decision of Person A will be implemented.

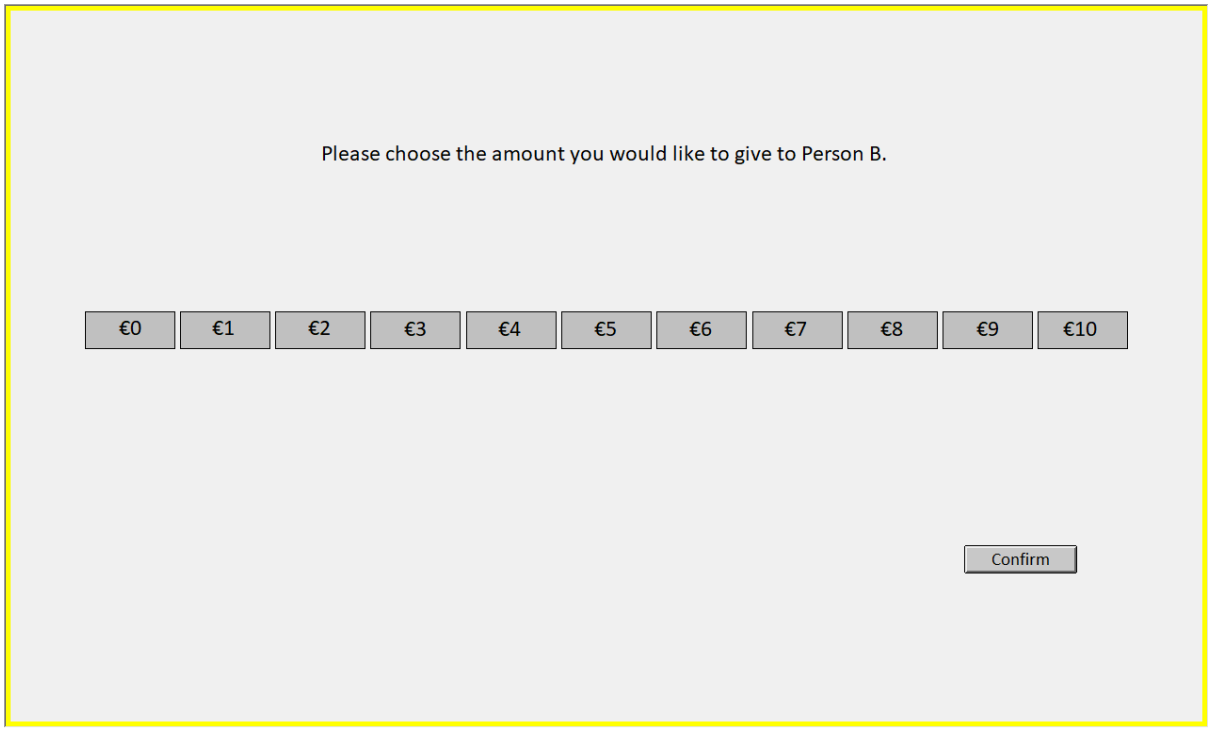
(*Text for all participants*)

Before you take your decision on the next page, please answer the following two questions.

1. How much would Person A, B, and C earn, if Person A gives €0 to Person B and Person C has assigned 1 deduction point to Person A for that case?
2. How much would Person A, B, and C earn, if Person A gives €5 to Person B and Person C has assigned 0 deduction point to Person A for that case?

Decision

(*An example of a decision screen in DG,* PRIVATE *treatment.*)



B.2.2 Reputation questionnaire

The following questions relate to the four tasks that you have just completed.

Please think about how you felt during the tasks and indicate to which extent the following statements apply. Please answer on a scale from "I completely disagree" to "I completely agree".

1. During the task I did not think about what other participants would say about me.
2. It's important that the other participants will accept me.
3. During the task, I thought about how the other participants would think about me.
4. It's important to me that the other participants have a positive evaluation about me.