

A Model of Social Duties*

Tore Ellingsen[†] Erik Mohlin[‡]

January 26, 2022

Abstract

We develop a structural model of social duties. Following Cicero, we divide social duties into obligations to respect others' rights (duties of justice) and responsibilities to promote others' well-being (duties of care). Rights are either defined outside of the situation, as in the case of property rights, or derive from moral ideals together with the distribution of power within the situation. The resulting utility function can have kinks due to both obligations and responsibilities. The model accounts for several documented unselfish behaviors that sympathy motives cannot easily explain. Pre-registered elicitations of social and personal appropriateness ratings support the model's predictions.

JEL Codes: D91, Z13

Keywords: Social norms, Social morality, Entitlements, Dutifulness, Charity

*Earlier drafts have been entitled "Situations and Norms" and "Decency". Thanks to Matilde Casamonti for immaculate research assistance, and to Johannes Abeler, Sandro Ambuehl, Björn Bartling, Douglas Bernheim, Pol Campos-Mercade, Andrew Caplin, Vincent Crawford, Armin Falk, Ernst Fehr, Sebastian Fehrler, Gustaf Karreskog, Erik Gaard Kristiansen, Ulrike Malmendier, Karine Nyborg, Robert Östling, Zoltán Rácz, Alexandros Rigos, Felix Schafmeister, Christian Schultz, Peter Norman Sørensen, Mark Voornveld, Roberto Weber, and especially Klaus Schmidt, for helpful comments and discussions. Ellingsen gratefully acknowledges financial support from the Torsten and Ragnar Söderberg Foundation and from the the Norwegian Research Council (grant 250506). Mohlin gratefully acknowledges financial support from the Swedish Research Council (grant 2015-01751), and the Knut and Alice Wallenberg Foundation (Wallenberg Academy Fellowship 2016-0156).

[†]Address: Department of Economics, Stockholm School of Economics, Box 6501, S—11383 Stockholm, Sweden. Email: gte@hhs.se

[‡]Address: Department of Economics, Lund University, Tycho Brahes väg 1, 220 07 Lund, Sweden. Email: erik.mohlin@nek.lu.se

1 Introduction

Why do so many people give to charity? Why do they tip? Who do they vote? Why do they pay taxes that they might easily have avoided? Why do they help colleagues and friends even when they understand that the recipient will be unable to reciprocate? Why do they engage in social distancing in order to protect vulnerable strangers from virus infection? Why do they tell the truth when they could lie with impunity? More generally, why do they prioritize the social interest?¹

One reason is *sympathy*. People are genuinely kind, taking pleasure from others' joy and pain from their suffering. Another reason is *duty*. People feel that they ought to act justly and charitably even if they incur material losses from doing so. There is already a rich array of models of sympathy and antipathy.² Here, our objective is to build and evaluate a model of the duty motive that is comparable to these sympathy-based models.

Is it important to disentangle duty from sympathy? For some purposes, the distinction is immaterial. For example, if we merely want to predict the extent of charitable behavior when a person faces a fundraiser, it is irrelevant whether the person is happy or unhappy to be facing the choice. The two models would be observationally equivalent. However, if we want to predict whether people will try to avoid fundraisers, or if we seek to compute the welfare effects of fundraising, the distinction matters. Altruistic donors seek the charitable opportunity, dutiful donors avoid it. As demonstrated by the field experiments of DellaVigna, List, and Malmendier (2012), Andreoni, Rao, and Trachtman (2017), and Exley and Petrie (2018), charitable giving is often reluctant.

A second difference is that duties come in two distinct flavors. On the one hand are the "hard" duties to respect others' rights, also called *obligations* or *duties of justice*. On the other hand are the "soft" duties to promote others' well-being, also called *responsibilities* or *duties of care*. In our model, violations of obligations are more blameworthy than are violations of responsibilities. Hence, mildly dutiful people might perform their obligations but neglect their responsibilities. In technical terms, obligations create a kink in the decision maker's utility function. This kink created by the obligation comes in addition to any kink created by responsibilities, and it allows us to explain phenomena that other social preference models struggle to capture, such as positive but non-ideal giving in Dictator experiments and, more generally, the pervasive practice of exact norm compliance even in the absence of sanctions.

¹A possible response to this question is to deny the premise. Maybe people do promote their own material well-being in these cases too. They may be rationally afraid that a selfish act hurts them by causing social contagion (Kandori, 1991). They may even hold "magic beliefs" that if they fail to cooperate, bad consequences will immediately follow (Shafir and Tversky, 1992). While both these effects may matter, the evidence that we survey below indicates that other effects are frequently at play.

²For example, there are models of altruism and spitefulness (Edgeworth, 1881; Becker, 1974; Levine, 1998), fair-mindedness (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), and even a model that contains all of these motives (Charness and Rabin, 2002).

A third difference is that sympathy is a property of the individual while the group determines what our social duties are and to whom.³ In some societies, people have many social duties; in others, individual freedom is greater (e.g., Gelfand et al, 2011). In some societies, duties are mostly confined to the family, clan, or nation; in others, they are more universal (e.g., Enke, Rodríguez-Padilla, and Zimmermann 2020). Thus, the study of social duties has the potential to shed light on the large observed cross-culture differences in moral behavior (e.g., Cohn et al, 2019).

Since the nature of an obligation is to respect others' rights, our model of social duties requires a formulation of rights. Our formulation recognizes that rights can be either exogenous or endogenous to the situation. Exogenous rights are general rules that pertain directly to actions, which are expressly permitted or forbidden. Legal rights, often formulated as property rights, are typically exogenous. By contrast, endogenous rights arise on the basis of the actions and payoffs that define the situation. Specifically, we assume that situations give rise to (payoff) *entitlements* which in turn determine which actions are permitted and forbidden.

Concretely, payoff entitlements are weighted averages of morally ideal payoffs and maximin payoffs. Thus, they strike a balance between "right" and "might".⁴ The impact of the maximin payoffs implies that the model allows for *context* in addition to consequences. Thereby the model rationalizes the impact of unused options on people's moral choices documented by List (2007) and Bardsley (2008).⁵

Context can also enter entitlements through the definition of what is right. For example, if the parties have previously talked and agreed on an outcome, that outcome might have now become the moral ideal.⁶ Similarly, even if the moral ideal is given, the weights on might and right can reflect the history in which the situation is embedded. For example, moral entitlements to a sum of money might depend on how the money was earned. Our model is thus compatible with equity theory and the evidence associated with it (e.g., Adams, 1965; Leventhal, 1980; Konow, 1996; Cappelen et al, 2007, 2013a).⁷

We empirically measure perceptions of rights through the elicitation procedure devised by Krupka and Weber (2013) (KW). The KW-procedure elicits the social appropriateness of actions by inducing people to truthfully report what they consider to be the

³This is not to deny that people can experience personal duties which are unrepresentative of the society that they currently live in. For example, reading moral philosophy as a student might permanently affect the reader's values. Below we make a distinction between the individual's and the society's view of appropriate behavior, comparable to the common distinction between morals and ethics.

⁴Maximin payoffs are appealing since they are unique and easy to find. However, we would not be surprised if there are applications in which it would be better to represent "might" through some equilibrium of the material game (i.e., the game with material payoffs as utilities) instead.

⁵This central part of our model relates closely to the model of moral reference points recently developed by Cox et al (2019).

⁶See Krupka, Leider, and Jiang (2016). We do not elaborate on this important application here.

⁷However, since we do not develop a structural theory of how the weights on might and right arise, we make no contribution to equity theory itself.

most common assessment of the social appropriateness of the various available actions. We define an action to violate a social obligation—i.e., to encroach on another’s rights—if the action is classified as (somewhat or very) inappropriate.

For simplicity, we assume that people comply with social duties only because they have internalized them. We do not consider compliance that is caused by future rewards or punishments. Nor do we consider that people perform duties in order to gain social esteem, as in Bernheim (1994), or to protect others from being disappointed, as in Char-ness and Dufwenberg (2006). That is, compliance is driven entirely by fear of guilt from moral failure rather than by farsighted material concerns or concerns for what others think. In this respect, our model is a close cousin of DellaVigna, List, and Malmendier (2012).

To gauge the model’s explanatory power, we confront it with a variety of evidence that sympathy-based models fail to accommodate. To corroborate the model’s explanation for the evidence, we also make several new KW-elicitations, one for each application. The phenomena that we address are:

1. *Giving*. (i) Standard Dictator experiments yield gifts of zero and of half the available money, but also several other fractions between these two extremes (Engel, 2011). (ii) Giving depends on how the endowment was earned (e.g., Konow, 2000; Cherry, Frykblom, and Shogren, 2002; Kameda et al, 2002; Cappelen et al, 2007).
2. *Giving and Taking*. In Dictator experiments with taking options, the opportunity to take reduces the propensity to give even among people who do not avail themselves of the taking opportunity (List, 2007; Bardsley, 2008; Cappelen et al 2013b; Korenok, Millner, and Razzolini, 2014).
3. *Giving and Exit*. In Dictator experiments with unexpected exit options, the opportunity to exit is more attractive to subjects that gave more initially (Dana, Cain, and Dawes, 2006, Broberg, Ellingsen, and Johannesson, 2007, and Lazear, Malmendier and Weber, 2012).
4. *Willful Ignorance*. In Dictator experiments with uncertain externalities, many subjects prefer to be ignorant rather than to learn the externality prior to acting (Dana, Weber, and Kuang, 2007; Feiler, 2014; Grossmann, 2014; Grossman and van der Weele, 2017; Freddi, 2021; and Serra and Szech, 2021).
5. *Incentive Paradox*. Fines sometimes backfire; they encourage the punished behavior instead of discouraging it (Gneezy and Rustichini, 2000a).
6. *Lost Wallet*. People have a greater tendency to return found wallets to their owner when there is more money in them (Cohn et al, 2019).

Items 1-4 were also central to the analysis of KW, which we build on. The final two items show that the model has applications far beyond those that it was built to address.

The moral philosophy of social duties has deep roots. According to Marcus Tullius Cicero, whose book *On Duty* was written in year 44 BC, the social duties of justice and care constitute the glue that holds societies together.⁸ This idea has stood the test of time. Major works of David Hume (1751) and Adam Smith (1759) devote considerable effort to explaining how the duties of justice and care are essential to prosperity.⁹ For example, Adam Smith argues in some detail that duties of justice and care constitute more important drivers of prosocial behavior than does sympathy, and that justice is more important than care (Part II, Section II, Chapter III). Throughout, we follow Smith's practice of referring to duties of justice as obligations and to duties of care (which Smith called duties of beneficence) as responsibilities.¹⁰

However, in the 20th century as social sciences became more specialized, economists largely left the study of morality to philosophers, psychologists, legal scholars, and sociologists.¹¹ In social psychology, the question of social duties got a new lease of life when Piaget (1932) studied how duties become internalized in children. The modern literature on child development studies behavior and emotions related both to justice (e.g., Kohlberg, 1964) and care (e.g., Gilligan, 1982). Recent work by social psychologists that is especially relevant to our analysis is that of Janoff-Bulman, Sheikh, and Hepp (2009) and Janoff-Bulman and Carnes (2013), who refer to justice as *proscriptive morality* and to care as *prescriptive morality* and demonstrate how the two motivations differ. For a broader survey of the literature on social morality, see Haidt (2008). In political science, the concept of duty has also maintained a stronger presence than in economics. For example, the duty motive is widely accepted as a leading explanation for why people vote (e.g., Riker and Ordeshook, 1968; Blais and Galais, 2016).

Granovetter (1985) criticizes the failure of 20th century economists to recognize the social embeddedness of economic transactions.¹² As he points out, economists generally

⁸*On Duty* is widely regarded as one of the most important books Western intellectual history. It is based on an even earlier manuscript, now lost, by the stoic Greek philosopher Panaetius. The central statement on justice and care (charity) starts in Section 1.20. (Apart from pursuing justice and care, the major duties are to seek the truth and to act bravely—as well as to strike an appropriate balance between all the goals in one's life.)

⁹David Hume (1751) explicitly cites Cicero and can almost in its entirety be read as an expansion of the parts of *On Duties* that are concerned with the *social* duties. While Adam Smith (1759) represents a return to Cicero's broader perspective on morality, Smith also pays close attention to the social duties.

¹⁰Smith's modern reputation for downplaying moral concerns could not be less deserved. Perhaps maintaining the image of Smith as an economist rather than a sociologist is easier if one forgets that *The Theory of Moral Sentiments* is both his first book and his last. The sixth and substantially revised edition of *Moral Sentiments* was published in the year that Smith died, 1790. By comparison, *The Wealth of Nations* was a one-off; see the discussion in Sen (2009).

¹¹We make no attempt to trace the history of the duty motive in social science during the 19th century, except to note that the concept was central both to Durkheim (1900) and Weber (1905).

¹²For related discussions, see Coleman (1988) and March and Olsen (1989).

assume either that moral concerns do not matter or that moral concerns impose binding constraints on behavior. Our objective is to provide a model that satisfies Granovetter's desire that individuals should be neither undersocialized nor oversocialized.

We suspect that 20th century economists were reluctant to address questions of morality in part because morality tends to complicate the formalization of decision theory, which modern economics relies on. Economic decision theory generally takes for granted that decision-makers' preferences depend only on ultimate consequences—usually the distribution of material outcomes—rather than on the full context within which the situation takes place.¹³ While such consequentialism is compatible with the duty of care, it is incompatible with the duty of justice. Justice is inherently a function of context. If Person A has the power to divide 10 dollars between herself and Person B, the just division depends on each person's explicit or implicit claim to the money. These claims may in turn depend on the set of actions available to Person A—for example whether Person A can also take money from B—as well as on the events and shared narratives leading up to the situation. If Person B contributed decisively to make the money available, or has been declared the rightful owner of the money, it is unjust for Person A to take all of it.

Of course, there is already a large literature on justice principles, and how they affect behavior—and economic models play a significant role in it, as documented in Konow's (2003) wide-ranging survey. But the main focus of this literature is on ideal outcomes and behavior. Less attention is paid to *acceptable* behavior, or what we call obligations. While context-specific entitlements do figure prominently in parts of the field, Konow (p.1235) ultimately concludes: "Probably the most significant challenge [to this theory] is to incorporate the impact of context on justice evaluation, and much work remains in this regard." This is where we try to contribute.

Lately, the descriptive failure of consequentialist decision theory to explain unselfish behavior has inspired new formal models. Recent non-consequentialist theories that are especially relevant to social duties include Dillenberger and Sadowski (2012), Saito (2015), and Evren and Minardi (2017). However, these theories consider the case in which decision makers worry about what others will be thinking about them. By contrast, and like us, Cox et al (2019) consider internalized morality and develop a model of observable reference points. Their formulation of reference points is similar to and predates ours. We see the two contributions as complements, not only because the proposed moral reference points differ, but because the models have different objectives. Our primary objective is to build a simple and portable preference representation, i.e., a parametrized utility function that captures established moral intuitions and is amenable to quantitative analysis. Their primary objective is to provide an axiomatic extension of rational choice theory which incorporates observable moral reference points. That said, the analyses clearly in-

¹³At times, this assumption has been so deeply ingrained that decision theorists themselves forget that they are making it, as Hammond (1996, footnote 3) frankly admits.

tersect. In particular, both provide qualitative predictions that are consistent with rich evidence from Dictator experiments.

Arguably, models of social norms constitute the best available formal analysis of entitlements and context. These models often draw a line between the permitted and the forbidden rather than between the ideal and the non-ideal. However, formal models of social norms tend to be either non-parametric or ordinal, like Stigler and Becker (1977), Akerlof and Kranton (2000), and Krupka and Weber (2013), or tailored to specific applications, such as charitable giving or abuse of social insurance.¹⁴ Only by developing portable parametric models, where norms can be derived from features of the choice set or other observable context, can we systematically evaluate the strength of obligations and responsibilities across a wide variety of situations and cultures.

The part of our model which concerns responsibilities is quite closely related to Brekke, Kverndokk, and Nyborg (2003), which in turn builds on Nyborg (2000). A model of social norms that shares our emphasis on justice is López-Pérez (2008). However, his central contribution concerns the issue of reciprocity, especially the response to unjust behavior, which we do not address here.¹⁵

The paper is organized as follows. Section 2 presents the model. Section 3 discusses the relationship between the formal model and Krupka–Weber appropriateness ratings. Section 4 applies the model to address a number of empirical puzzles. Section 5 concludes.

2 A Simple Model of Social Duties

A single decision-maker takes a decision that impacts herself and one other person. We refer to the decision-maker as Decider and the other person as Other. For the most part, we take for granted that Decider has duties toward Other. That is, we do not model the determination of moral boundaries. A simple generalization is to introduce a binary variable to indicate whether Decider has a social duty towards Other or not.¹⁶

¹⁴While Akerlof and Kranton (2000) emphasize the social norms interpretation of their model much more than do Stigler and Becker (1977), Sobel (2005) shows that the two models are mathematically equivalent. For parametric models of dutiful charitable giving, see Andreoni and Bernheim (2009) and DellaVigna, List, and Malmendier (2012); for a model of free-riding on social insurance, see Lindbeck, Nyberg, and Weibull (1999). Kandel and Lazear (1992) is an early example of a model of social norms that is both non-parametric and tied to a specific application, namely, peer pressure in the workplace.

¹⁵A third complementary strand of research is Bicchieri (2005). She develops a broader framework for thinking about social norms, with a focus on how people adapt—often unconsciously—to the behavior of others. Kimbrough and Vostroknutov (2016) offer support for Bicchieri’s social norms hypothesis by showing that prosocial behavior is greater among people who are inclined to comply with rules in other domains.

¹⁶In the standard Dictator experiments that we study below, it is quite clear that the vast majority of participants consider that there is in fact a duty to give, in the sense that most participants consider an even split to be the most appropriate outcome. However, that fraction falls when the dictator role is deserved rather than randomly allocated, when about a third of the subjects consider it very appropriate to give

Decider has access to a set of actions $\mathcal{A}$, with typical element a .¹⁷ Other has no action to take. Actions entail material consequences for both persons. Let $\mathcal{M} \subset \mathbb{R}^2$ be the set of feasible consequences, and let $x : \mathcal{A} \rightarrow \mathcal{M}$ denote the outcome function. That is, the pair $(x_d(a), x_o(a))$ represent the material consequences to D and O from action a . Let $\mathcal{M}_i$ denote the set of feasible consequences for Person i . We refer to $a_d = \arg \max x_d(a)$ as Decider's *selfish* action.

As a minimal running example, suppose $\mathcal{A} = \{H, N\}$, $x(H) = (0, 3)$, and $x(N) = (1, 1)$. This Helping situation is illustrated in Figure 1.

H	0, 3
N	1, 1

Figure 1. Helping situation

We next describe the two kinds of social duties. The first is the obligation to respect others' interests (duties of justice). The second is the responsibility to promote the community's objectives (duties of care).

2.1 Obligations: Duties of justice

An action is either morally permitted, in which case Decider has a *right* to take it, or morally forbidden, in which case Decider does not have a right to take it.¹⁸ Decider has an *obligation* not to take forbidden actions.

The blameworthiness associated with forbidden actions depends on how much the action harms others. To measure harm, we must therefore first define Other's payoff entitlement.

When rights are general, i.e., exogenous to the situation, it is easy to define Other's entitlement.

Definition 1 *When rights are exogenous, Other's entitlement is*

$$e_o = \min_{a \in \mathcal{P}} x_o(a). \quad (1a)$$

nothing. The more general formulation would capture this heterogeneity through a person-specific binary variable.

¹⁷We do not allow randomization over the set of actions.

¹⁸In what follows, we suppress the word "morally." Despite the legal rather than moral connotations, we prefer the word "rights" to the word "norms", because the latter has so many established uses. The definition of norms that is closest to our definition of rights is due to Ostrom (2000, p143-144). Following Crawford and Ostrom (1995, p583), she defines the word (injunctive/prescriptive) *norm* precisely to denote the binary distinction between permitted and forbidden actions. As Krupka and Weber (2013, p.499) observe, such binary definitions of injunctive norms are common; yet, they themselves prefer to use a more fine-grained measure, identifying the extent of norm compliance with the degree of appropriateness. In our setting, their concept of norm compliance measure corresponds to (the inverse of) our concept of blameworthiness, which we define below.

In words, Other is entitled to a payoff that is no smaller than the smallest payoff associated with a permitted action. In the Helping situation, our running example, $e_o = 3$ if only H is permitted and $e_o = 1$ if both H and N are permitted. For a real-life example, note that property rights are typically general. A person is entitled to keep what she has procured or produced, and therefore the action of stealing such property is forbidden.¹⁹

When the community has no general conception of rights that applies to the available actions in the situation, people instead need to infer the rights from the nature of the situation. In this case, we posit that entitlements strike a balance between some moral ideal (to be determined below) and the power possessed by the parties.

Specifically, let us confine attention to “generic” cases where $\hat{\mathcal{A}}$ is a singleton set, i.e., there is a unique morally ideal outcome $(\hat{x}_d, \hat{x}_o) = x(\hat{a})$. (All our applications are generic in this sense.) People construct entitlements as a weighted average of this ideal allocation and Divider’s selfish allocation, call it $(\tilde{x}_d, \tilde{x}_o)$.²⁰ If feasible, Other’s entitlement as perceived by Decider is thus

$$e_o^*(d) = \mu \hat{x}_o + (1 - \mu) \tilde{x}_o. \quad (1b)$$

However, $e_o^*(d)$ may not be feasible. To illustrate, suppose H is the morally ideal action in the Helping situation. Then $e_o^*(d) = 3\mu + (1 - \mu) = 1 + 2\mu$. This payoff is only attainable for Other if μ is either 0 or 1.

If $e_o^*(d)$ is not feasible, Other’s perceived entitlement is instead given by the feasible consequence closest to the point $e_o^*(d)$. If two feasible consequences are equally close, the tie is broken in favor of Decider. That is, when unique, the entitlement is

$$e_o(d) = \arg \min_{x_o \in \mathcal{M}_o} |x_o - e_o^*(d)|. \quad (1c)$$

When the solution is not unique, pick the solution with the highest x_d . In the Helping situation, we thus have $e_o(d) = 3$ if $\mu > 1/2$ and $e_o(d) = 1$ if $\mu \leq 1/2$.

In general, the weight μ may also depend on circumstances. For example, it will be larger when Decider’s powerful position is due to luck than when it is due to effort or skill (see Cappelen et al, 2007, and the references therein).

Having thus defined the entitlements, we define rights as follows.

Definition 2 *When rights are endogenous, an action a is perceived as permitted by Decider if $x_o(a) \geq e_o(d)$.*

¹⁹Cicero (44 BC, Book 1, Paragraph 21) and especially Hume (1751, Section III, Part II) devote considerable attention to the origin and wisdom of private property rights, and these are not exclusively Western ideas. Respecting prior possession is a universal moral value, as documented by Curry, Mullins, and Whitehouse (2019). For intriguing discussions of prior possession principles in the “lawless” context of the Californian gold rush, see Umbeck (1977) and Zerbe and Anderson (2001).

²⁰Formally, the selfish allocation satisfies $\tilde{x}_d = \max_a x_d(a)$; if $\arg \max_a x_d(a)$ is not unique, pick the a with the largest x_o .

That is, Decider considers that it is forbidden to take an action that gives Other less than Other's entitlement, but permitted to take any other action.

The *harm* to Other (as viewed by Decider) is defined as the payoff deficit relative to the entitlement,

$$h_o(a, d) = \max\{0, e_o(d) - x_o(a)\}. \quad (2)$$

In the Helping situation, if Other is entitled to get 3 and Decider plays N , the harm is $3 - 1 = 2$. If instead Other is entitled to get 1 and Decider plays H , the harm is $\max\{0, 1 - 3\} = 0$.

2.2 Responsibilities: Duties of care

Duties of care are defined in relation to the community's objectives. Suppose the community's objective is to pursue the maximal material outcome. In this case, communal value is

$$c(x) = \sum_{i \in \{d, o\}} x_i. \quad (3a)$$

If in addition the community values equality, communal value might be written

$$c(x) = \sum_{i \in \{d, o\}} x_i - \alpha \sum_{i \in \{d, o\}} \left| x_d - \sum_{j \in \{d, o\}} x_o / 2 \right|. \quad (3b)$$

Let $\hat{a}$ be a maximizer of $c(x(a))$, let $\hat{\mathcal{A}}$ denote the set of such maximizers, and let

$$s(a) = c(x(\hat{a})) - c(x(a)) \quad (4)$$

be called the *shortage* produced by action a .²¹

In the Helping situation, the shortage-minimizing action depends on α . If $\alpha \leq 1/3$, then $s(H) = 0$ and $s(N) = 1 - 3\alpha$. If $\alpha > 1/3$, then $s(H) = 3\alpha - 1$ and $s(N) = 0$.

From now on, we assume that the moral ideal referred to above corresponds to the shortage-minimizing action. We say that there is a duty of care to pursue the shortage-minimizing, except when it conflicts with a duty of justice. Equivalently:

Definition 3 *If a shortage-minimizing action is permitted, it represents a responsibility.*

Clearly, responsibilities may well coincide with endogenous obligations, in which case the action is doubly desirable. In such cases, people are thanked for fulfilling their responsibility, while if they fail to do so they are chastised for violating their obligation.

²¹In general, the weight α might depend on the context, and people might even have different views on it, but in the applications that we shall be considering here, it will only matter whether α is positive or not. For example, in the Dictator experiments we consider, all outcomes are efficient, and hence the equal split uniquely maximizes communal value for all positive α .

But why can there not be a responsibility that violates an obligation? For example, if H is an obligation, why cannot Decider justify N on the basis that $\alpha > 1/3$ and hence N is shortage-minimizing? In rejecting this possibility, we follow Cicero and Adam Smith, who both give strict priority to obligations. Adam Smith (1759, Section II, Chapter II, first paragraph) formulates it as follows:

There can be no proper motive for hurting our neighbour, there can be no incitement to do evil to another, which mankind will go along with, except just indignation for evil which that other has done to us. To disturb his happiness merely because it stands in the way of our own, to take from him what is of real use to him merely because it may be of equal or more use to us, or to indulge, in this manner, at the expense of other people, the natural preference which every man has for his own happiness above that of other people, is what no impartial spectator can go along with.

Cicero's (44 BC) similar views are most clearly articulated in Book III, Paragraphs 29-30.²²

2.3 Blameworthiness and behavior

The blameworthiness of an action is given by its contribution to harm and shortage. We take the view that duties of justice are always important, whereas duties of care are more context-dependent. Cicero (Book I, Paragraphs 40-59) provides an extensive discussion of how our duties of care depend on circumstances. Accordingly, we write blameworthiness as

$$b_d(a) = h_o(a) + \gamma s(a), \quad (5)$$

where γ is a parameter that captures the importance of care. Instead of Cicero's fine gradations, we assume that there are situations in which care is salient and situations in which it is not—either because care conflicts with justice or for other reasons. Clearly $\gamma = 0$ in the latter case. In the former case, we often set $\gamma = 1$ (even if it might be more realistic to set $\gamma < 1$, as responsibilities may be less important than obligations). Comparing the blameworthiness function b to the “social pressure cost” formulation of

²²One exception from this general priority of obligations over responsibilities occurs if it is clear that Decider does not pursue the shortage-minimizing action for selfish reasons, and that fulfilling the obligation is associated with a great shortage. As a dramatic but relevant example, Cicero (Book III, Paragraph 19) mentions the killing of another person—maybe even an old friend—who has become a tyrant.²³ While there is a general obligation not to kill others, there is also a responsibility to promote the proper functioning of society.

Cicero, Book I, Paragraphs 31-32, also provides one exception to the rule that selfish concerns cannot justify the violation of a duty. In the context of a promise, suppose something happens that suddenly changes the circumstances, making the cost of fulfilling the promise much larger than it was when the promise was made. Then, violating the promise might be justified—as the original obligation would not have been entered into had the new circumstance been foreseen.

We do not try to capture any of these exceptions formally, and they not arise in any of the applications that we consider.

DellaVigna, List, and Malmendier (2012), the main difference is that we impose more structure through h and s ; the kink at $h(a) = 0$ is particularly important.

We assume that Decider's utility function is

$$u_d(a) = x_d(a) - \delta b_d(a). \quad (6)$$

We refer to the individual-level parameter δ as Decider's *dutifulness* and to $\delta b_d(a)$ as Decider's *guilt*. Such an additive formulation is convenient and also preferred by DellaVigna, List, and Malmendier (2012). Note that any difference in the average value of δ across societies can be interpreted as differences in the weight that societies attach to duty.

When $\delta = 0$, Decider is *homo oeconomicus*, a selfish materialist whose behavior is unaffected by obligations and responsibilities. As δ grows large, Decider becomes *homo sociologicus*, whose choices are essentially determined by the society. Like Granovetter (1985), we are interested in the intermediate case.

3 Measuring Obligations and Responsibilities

The distinction between obligation and responsibilities is a key feature of the model. Obtaining a credible empirical measure of this distinction is therefore essential for building confidence in the model's mechanism.

3.1 The suitability of the Krupka–Weber measure

Krupka and Weber (2008,2013) define a measure that is perfect for our purposes. Their procedure elicits the social appropriateness of actions by inducing people—either participants or spectators—to truthfully report what they consider to be the group's most common assessment of the social appropriateness of the various available actions. The KW-scale runs from “very socially inappropriate” to “very socially appropriate”. We define an action to violate an obligation if the action is classified as inappropriate. We think that the definition is consistent with Krupka and Weber's (2008,Appendix) instruction to participants:

By socially appropriate, we mean behavior that most people agree is the “correct” or “ethical” thing to do. Another way to think about what we mean is that if Individual A were to select a socially inappropriate choice, then someone else might be angry at Individual A for doing so.

These instructions closely echo Adam Smith (1759, Part II, Section II, Chapter 1), when he distinguishes the duty of beneficence (responsibilities) from duties of justice (obligations), and explains how (i) beneficent actions are worthy of praise, but their absence

is not worthy of punishment, and (ii) unjust actions are worthy of punishment. Hence, obligations and responsibilities define a scale of duties where there is a well-defined origin, and actions of a positive nature are distinctly different from actions of a negative nature.²⁴

3.2 Our elicitations and tests

Our experiments were conducted in the online labor market Prolific during October and November 2021.²⁵ We had 150-200 subjects in each treatment. Before each experiment, we pre-registered our hypotheses with Open Science Foundation. Subjects were all based in the United States. Depending on the expected duration of the experiment, each subject received a participation fee of either 2 or 3 GBP, with the hourly pay falling in the interval 22-27 GBP (amounting to 29-36 USD at the time).

Each subject is asked to rate the appropriateness of behavior in several situations, but the experiments were designed both to maintain subjects' attention throughout the session and to net out any remaining order effects by varying the order of the situations that a subject is exposed to.

Subjects are either asked to guess the most common social appropriateness rating or to provide their own personal rating of appropriateness. In the experiments that elicit social (as opposed to personal) appropriateness ratings, participants earn either 2.5 or 4 GBP extra if they match the most common appropriateness rating when one of their ratings is randomly drawn for payment.

Our instructions closely follow the format of Krupka and Weber (2008, Appendix), except that we usually include the neutral option "neither appropriate nor inappropriate" to their categories "very inappropriate", "somewhat inappropriate", "somewhat appropriate", and "very appropriate".

When we want to test whether there is an obligation to perform a certain action, we test whether the proportion that rates it inappropriate is greater than 1/2 or greater than the fraction rating it appropriate (a one-sample test of proportions).

Many of our statistical tests compare the rating distributions for two different survey questions, applying instead a two-sample test of proportions. Sometimes, we are interested in the frequency of a particular response, in which case we apply a Mann-Whitney test.

Like Krupka and Weber, we also assign numerical value -1 to the lowest category and +1 to the highest, and place the other categories equidistantly in this interval. We can then apply t-tests to compare average appropriateness ratings across survey questions.

²⁴Moreover, in judging the morality of an action, Adam Smith clearly holds that intentions are important; reward is not due when benefits are unexpected, just as punishment is not due when the hurt is unexpected.

²⁵For a comparison of various online labor markets, see Peer et al (2021).

We mostly relegate regression tables to the Appendix, but report p-values in parentheses whenever we make statements about statistical significance in the main text.

4 Applications: Addressing Six Puzzles

We next confront the model with some well-known puzzles. In each case, we first briefly describe the puzzle. We then translate the situation that generates the puzzle into the language of our model and derive the model's predictions. Finally, we display the relevant new data that we have collected and discuss to what extent the new data support the model's interpretation of the original puzzle.

Our first four applications overlap closely with the applications considered by Krupka and Weber (2013). They are variations of the Dictator experiment. The difference is that we look at them from the perspective of a structural theory.

The last two applications concern two rather different field experiments, which both challenge existing theories. One considers the impact of fines at a daycare center and the other considers the choice to return a lost wallet to the rightful owner.

4.1 Standard Dictator Experiments

In a standard Dictator experiment, the experimenter has given Decider an endowment which she can share however she likes with Other (usually called the recipient). For concreteness, let the endowment be 10 dollars, let a denote how many dollars Decider decides to keep, and let Other get $10 - a$ dollars.

According to Equation (1b), Other's entitlement (as perceived by Decider) is

$$e_o(d) = \mu \cdot 5 + (1 - \mu) \cdot (0) = 5\mu. \quad (7)$$

Suppose moreover that communal value is relevant (i.e., $\gamma > 0$) and given by (3b). Thus, Decider's utility is

$$\begin{aligned} u_d &= a - \delta (h_o(a) + \gamma s(a)) \\ &= a - \delta (\max\{0, e_o(i) - (10 - a)\} + \alpha \gamma |a - (10 - a)|) \\ &= a - \delta (\max\{0, 5\mu - 10 + a\} + \alpha \gamma |2a - 10|), \end{aligned}$$

where the first equality uses (6) and (5), the second equality uses (2) and (3b), and the last equality uses (7). Maximizing this piece-wise linear objective function with respect to a subject to the constraint that $a \leq 10$ yields the following result.

Observation 1 *In a standard Dictator experiment, Divider's choice is*

$$a = \begin{cases} 10 & \text{if } \delta < \frac{1}{1+2\alpha\gamma}; \\ 10 - 5\mu & \text{if } \delta \in \left[\frac{1}{1+2\alpha\gamma}, \frac{1}{2\alpha\gamma} \right]; \\ 5 & \text{if } \delta > \frac{1}{2\alpha\gamma}. \end{cases} \quad (8)$$

The intuition is straightforward and illustrated in Figure 2. If Decider is sufficiently du-

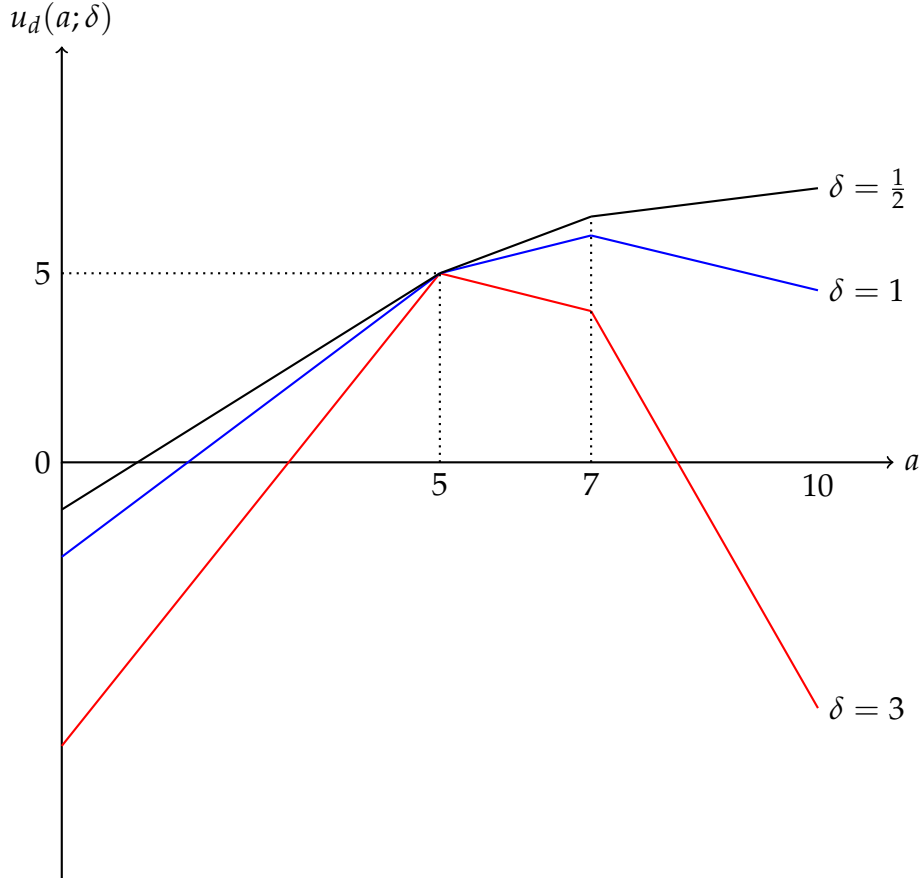


Figure 2. Decider's utility in the Dictator situation ($\alpha = 1/4, \mu = 3/5, \gamma = 1$)

tiful (e.g., $\delta = 3$ in Figure 2), she will maximize communal value and share equally. If her dutifulness is sufficiently low (e.g., $\delta = 1/2$), she will neglect both her obligation and her responsibility and keep everything. For an intermediate range of dutifulness (including $\delta = 1$), Decider neglects her responsibility but fulfills her obligation and gives Other exactly his entitlement.

We think that this is a plausible account of available data, which displays two modes at 5 and 10, and where keeping 6, 7, and 8 is also not uncommon. Our interpretation is that these intermediate values correspond to the deciders' conception of the recipients' entitlement, i.e., to a non-degenerate distribution of μ .

Action	VSI	SI	N	SA	VSA
Keep \$ 10	0.62	0.20	0.08	0.07	0.03
Keep \$ 9	0.43	0.28	0.11	0.16	0.02
Keep \$ 8	0.27	0.40	0.10	0.20	0.02
Keep \$ 7	0.11	0.41	0.18	0.26	0.05
Keep \$ 6	0.03	0.22	0.23	0.43	0.10
Keep \$ 5	0.01	0.01	0.12	0.25	0.62
Keep \$ 4	0.04	0.07	0.13	0.42	0.35
Keep \$ 3	0.05	0.14	0.20	0.27	0.34
Keep \$ 2	0.08	0.20	0.16	0.18	0.38
Keep \$ 1	0.14	0.14	0.16	0.18	0.38
Keep \$ 0	0.20	0.10	0.16	0.14	0.41

Table 1. Standard Dictator Experiment: Social Appropriateness

This interpretation is supported by the elicited appropriateness rankings of Krupka and Weber (2013, Table 1). To ensure comparability across our different treatments, we replicate their elicitation here; see Table 1.

The even split is the only action that a majority considers to be very socially appropriate, with many of these also finding it very socially appropriate to give more than half. Let us now infer what the implied entitlements are. With the exception of a few subjects who think that an even split is inappropriate—here rounded up to 2 percent—all the subjects agree that the recipient is not entitled to more than \$5. But the agreement ends there. About a quarter of the subjects (0.03+0.22) consider that it is very or somewhat inappropriate to keep \$6. In other words, a fraction $0.25-0.02=0.23$ think that the recipient is entitled to exactly \$5. By the same argument, a fraction $(0.11+0.41-0.23-0.02)=0.26$ think that the entitlement is \$ 4, a fraction $(0.27+0.40-0.26-0.23-0.02)=0.14$ think that the entitlement is \$3, a fraction $(0.43+0.28-0.26-0.23-0.14-0.02)=0.06$ think that the entitlement is \$2, a fraction $(0.62+0.20-0.26-0.23-0.14-0.06-0.02)=0.11$ think that the entitlement is \$1, and the remaining 0.18 think that the recipient is not entitled to anything.²⁶

From these numbers and our linearity assumption, we obtain an upper bound on the fraction of gifts that lie strictly between 0 and 5, namely $0.26+0.14+0.06+0.11=0.57$. By contrast, there are no meaningful bounds on the fractions of gifts that are 0 or 5, since the distribution of the dutifulness parameter δ can drive either of these fractions arbitrarily high or low.

It may seem surprising that people can have such different views about entitlements. Is it really plausible that a fifth of the subjects think that the recipient is entitled to half and almost as many think that the recipient is entitled to nothing? A separate source

²⁶Note that we arrive at the latter number either as $(1-0.23-0.26-0.14-0.06-0.11-0.02) = 0.18$ or more directly as the fraction of subjects considering “Keep \$ 10” either “neutral”, “somewhat appropriate” or “very appropriate”—i.e., $(0.08+0.07+0.03)=0.18$.

of evidence on this question comes from Ellingsen and Johannesson (2008), who let the recipient write a message to the decider after observing the decider's decision. The messages' emotions range from profuse gratitude to great anger. While gratitude is more common after large gifts and anger is more common after small gifts, the overlap is striking. Some recipients express gratitude for small gifts, apparently thinking that they are not entitled to them.²⁷

Perhaps such disagreement is exactly what we should expect. The Dictator experiment is an unfamiliar setting. How strong is the divider's property right? Should the divider consider the endowment to be entirely common or should she consider it the same way as her other money?

It is this unfamiliarity of the Dictator situation that makes it difficult to generalize findings about donation levels to settings outside of the laboratory. However, even if people hold heterogeneous views about the level of entitlement, Dictator experiments can still be useful—even for studying the general determinants of entitlements. For example, Konow (2000) and Cappelen et al (2007) use Dictator experiments to demonstrate that people are more reluctant to share earned endowments than lucky endowments. The authors believe that this change in behavior is caused by a change in perceived entitlements, and hence the perceived duty to share.²⁸ In order to test their hypothesis, we here elicit social and personal appropriateness ratings under the different endowment regimes. Figure 3 reports our findings. Observe in particular how it is considered much less inappropriate to keep almost everything when the endowment is “earned”.²⁹ The average rating

Thus, standard Dictator experiments are instructive despite the heterogeneous perceptions of entitlements.³⁰

The next applications further illustrate the power of Dictator experiments to gain insights into the drivers of social behavior.

²⁷The messages are available at shorturl.at/uAW08. Here are excerpts from two contrasting messages from recipients who both got SEK 20 out of the endowment SEK 120. Recipient 1: “Thanks for the money! You made a generous choice by giving me twenty.” Recipient 2: “You greedy bastard. Normally you give at least 1/3. You have to be a boy!”

²⁸By contrast, Cherry, Frykblom, and Shogren (2002), offer a rather different explanation for this finding that is based on rationality and attention.

²⁹Of course, this movement in average ratings does not imply any reduction in heterogeneity. In fact, we find the opposite. People disagree even more regarding the appropriateness of low donations when the endowment is earned.

³⁰A related issue concerns the sensitivity of behavior to framing manipulations; some authors have portrayed Dictator experiments as being potentially highly sensitive to minute changes in wording. Addressing this worry, Dreber et al (2013) find to the contrary that pure framing effects are minuscule in Dictator experiments (but see the “Bully” treatment of Krupka and Weber, 2013).

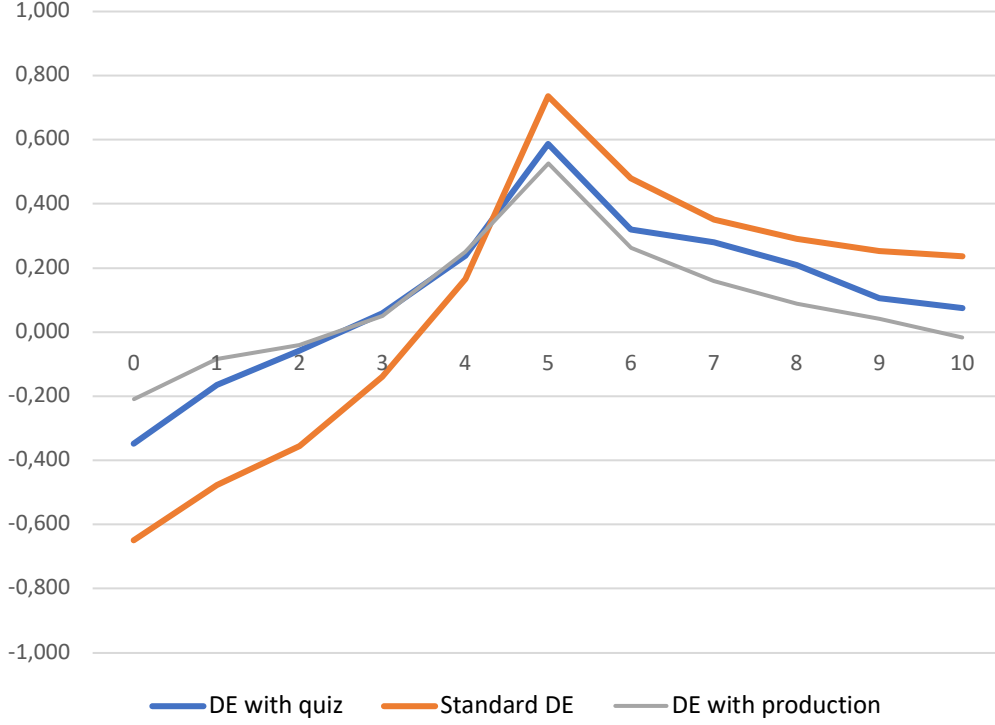


Figure 3. Various Dictator Experiments: Social Appropriateness Ratings

4.2 Giving and Taking Experiment

List (2007) and Bardsley (2008) conduct a Dictator experiment in which the availability of an option to take (steal) influences subjects' willingness to give. The most striking finding is that the taking option does not merely shift behavior away from keeping the full initial endowment towards taking—positive gifts also become less common and smaller. This is not predicted by any of the sympathy-based models of social preferences.

For easy comparison, we transform the 5 dollar Dictator experiment considered by List to a 10 dollar experiment and consider two treatment variations in addition to the baseline. In one treatment, the divider can take up to two dollars from the recipient, and in another she can take up to 10 dollars. Call them Take 2 and Take 10 respectively.

In our framework, the taking option can affect entitlements in two ways. First, any participant who continues to infer the norm from Equation (1b) will tend to consider that the recipient is entitled to less. Specifically, recall that $e_j = 5\mu$ in the baseline situation. Once it becomes possible for Decider to take 10 dollars, Other's new implicit entitlement is

$$e_o^T = \mu \cdot 5 + (1 - \mu) \cdot (-10) = 15\mu - 10.$$

Second, the taking option may alternatively suggest to the participants the more general obligation not to steal. This exogenous obligation might replace the implicit obligation the participants have previously been inferring. Thus, participants who pre-

viously have experienced an *obligation* to give might now only experience a *responsibility* to do so. In this case, we instead have $e_o = 0$.

Solving the utility-maximization problem as before (Equation (8)), but with the two alternative entitlements, yields the following solution—where we set $\gamma = 1$ to save notation.

Observation 2 *In the Dictator experiment with an additional option to take 10, Divider’s choice is*

$$a = \begin{cases} 20 & \text{if } \delta < \frac{1}{1+2\alpha}; \\ 20 - 15\mu & \text{if } e_o = e_o^T \text{ and } \delta \in \left[\frac{1}{1+2\alpha}, \frac{1}{2\alpha} \right]; \\ 10 & \text{if } e_o(i) = 0 \text{ and } \delta \in \left[\frac{1}{1+2\alpha}, \frac{1}{2\alpha} \right]; \\ 5 & \text{if } \delta > \frac{1}{2\alpha}. \end{cases} \quad (9)$$

Comparing to (8), we see that the dictator types that kept all their money in the standard treatment now keep all their money and take all they can from the recipient. However, there are new types that keep all their own money, namely those who now consider that the obligation is merely to not steal. The reduction in positive giving comes partly from these latter types and partly from those who revise the recipient’s implicit entitlement from 5μ to $15\mu - 10$. For those that hold $\mu = 1$ there is no revision, but others revise downwards. Given the previously inferred distribution of μ , some of the latter will now be net takers.

Both interpretations are consistent with the evidence on appropriateness ratings reported by Krupka and Weber (2013, Figure 5), but their aggregate analysis cannot discriminate between the mechanisms. It reveals that not giving is generally rated as “very socially inappropriate” when the taking option is not available, but that the average rating is only “somewhat socially inappropriate” when the taking option is available. Here, we replicate Krupka and Weber’s elicitations. When studying our disaggregated data, as displayed in Figure 4, we see that it yields support for each of our two mechanisms. In support of the moving-indirect-entitlement hypothesis, the whole rating distribution moves to the left as taking options increase. In support of the new-direct-entitlement-at-0 hypothesis, the ratings move steeply from inappropriate to appropriate around the gift of 0 when taking options are available.

4.3 Exit Experiment

In an experiment devised by Dana, Cain, and Dawes (2006), subjects are initially informed that they are in a Dictator situation. But after having made the allocation choice,

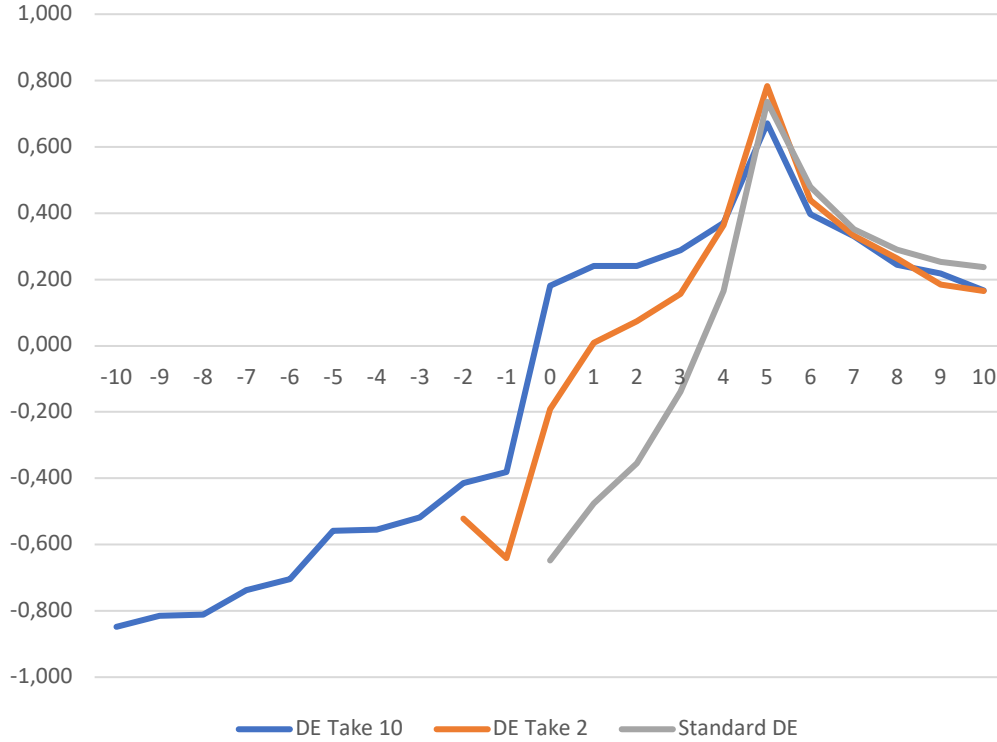


Figure 4. Dictator Experiments: Social Appropriateness Ratings

dividers (Decider) are told that recipients (Other) are not yet aware of the experiment. Decider is given the option to exit for a price of 1. In the case of exit, Decider thus keeps 9, and Other will never be informed. About a third of the subjects choose to exit, a finding that was replicated and elaborated—with even greater exit rates—by Broberg, Ellingsen, and Johannesson (2007). There is also a tendency that generous sharers are more likely to exit, a finding that is further corroborated in closely related work by Lazear, Malmendier, and Weber (2012).

As noted by Dana, Cain, and Dawes (2006), exiting is inconsistent with all social preference models that are based purely on sympathy, since the allocation $(10, 0)$ dominates $(9, 0)$. However, according to our duty-based model, exiting is potentially easy to rationalize. For example, suppose that there is no duty to refrain from exiting. In that case, subjects will compare the utility of getting 9 dollars to the utility given by the best choice according to Equation (8). Straightforward computations, where we again have set $\gamma = 1$, reveal that the outcome is as follows.

Observation 3 *In the Dictator experiment with an exit option, if there is no blame from exiting,*

the divider's final choice is

$$a = \begin{cases} 10 & \text{if } \delta < \min \left\{ \frac{1}{1+2\alpha}, \frac{1}{5(\mu+2\alpha)} \right\}; \\ 10 - 5\mu & \text{if } \delta \in \left[\frac{1}{1+2\alpha}, \frac{1}{2\alpha} \right] \text{ and } \mu \leq \frac{1-10\alpha\delta}{5-10\alpha\delta}; \\ \text{Exit} & \text{otherwise.} \end{cases} \quad (10)$$

Comparing (10) to (8), we see that when there is no duty to remain, exiting with 9 dollars is attractive for all subjects that were originally giving one dollar or more. Even some of those who were originally keeping everything are willing to sacrifice a dollar in order to exit. They can exit without feeling any guilt, but will feel guilty if they stay.

What if the divider perceives a duty to remain in the original situation? In this case, we need to know the guilt associated with exiting, which depends on whether exiting violates an obligation or a responsibility. Since these two cases are quite similar, we focus on the former.

If the divider perceives that exiting violates an obligation, she must take into account the harm that exit causes, namely the loss for the recipient due to exit, $10 - a$. Thus, using (8), we have that the divider prefers exit if and only if

$$9 - \delta(10 - a) \geq a - \delta(\max\{0, 5\mu - 10 + a\} + \alpha(2a - 10)).$$

For those that originally kept half of the endowment, $a = 5$, the condition simplifies to

$$\delta \leq \frac{4}{5}.$$

We know from (8) that anyone who initially chose $a = 5$ has dutifulness $\delta > 1/(2\alpha)$. Thus, under the reasonable assumption that $\alpha < 5/8$ no such divider would take the exit option.³¹ Similarly, for those that initially chose $a = 10$ exit is preferred if and only if

$$\delta \geq \frac{1}{5\mu + 10\alpha}.$$

Here, we know from (8) that anyone who initially chose $a = 10$ has dutifulness $\delta < 1/(1 + 2\alpha)$. Thus, a sufficient condition for there to exist some type δ that originally kept everything and prefers exit when offered is that $\mu > 1/5$. Thus, when there is an obligation to exit, it is no longer true that those that gave most initially are also most keen to exercise the exit option.

In our experiment, we elicit KW-ratings for the original experimental setting of Dana,

³¹Whenever $\alpha > 1/2$, the community would prefer to forgo any gain to better-off person unless the worse off person also gets something.

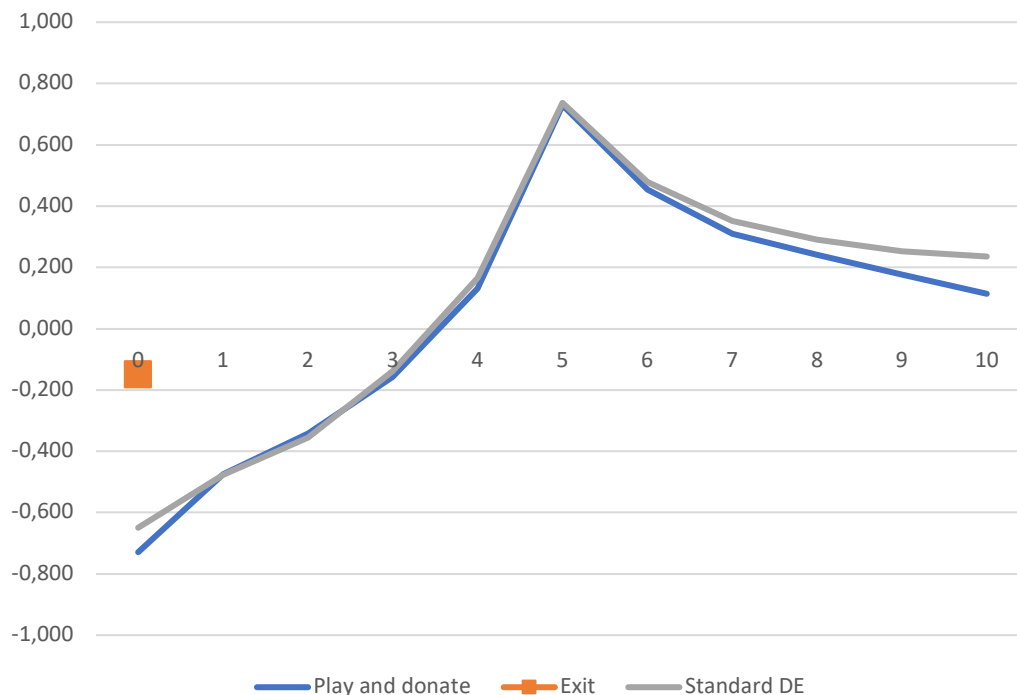


Figure 5. Dictator Experiment and Exit: Social Appropriateness Ratings

Cain, and Dawes (2006).³² Figure 5 reports our finding.

As the curves demonstrate, compared to the standard Dictator experiment, the appropriateness of the various donations is virtually the same. Moreover, just as in Krupka and Weber (2013), the exit option is rated to be almost neutral on average. But yet again people have very different notions of what is right. Roughly half the participants consider exit to be either very or somewhat inappropriate, about 20 percent consider it neutral, and the remaining 30 percent consider exit to be either somewhat or very appropriate.³³ These elicitations provide a clear reason why everyone do not exit; half of the participants think that they have an obligation to refrain from exiting, just as they have an obligation to give a non-negative amount. On the other hand, the remaining half think that they do not have an obligation to stick to their initial decision, and these should therefore behave as described above, with all the generous sharers exiting.

Overall, the distribution of appropriateness ratings is consistent with the observed heterogeneous behavioral pattern. There are enough people judging the exit option appropriate to justify the sizable exit rate, but also enough people considering exit inappropriate to justify that as many or more refrain from exiting.

³²Eliciting social appropriateness ratings in a closely related setting, Krupka and Weber (2013, Figure 3) find that choosing exit is almost neutral on average, whereas sticking with a selfish option is very socially inappropriate.

³³Of these numbers, it is hardest to interpret the neutral rating. In the context of anonymous exit, this option could mean either that people feel guilty for not taking the most socially appropriate action or that they consider that no duty applies.

4.4 Information Avoidance Experiment

Dana, Weber, and Kuang (2007) provide another important objection to the notion that charity is mostly based on sympathy. They demonstrate that many people prefer not to know whether there is a reason to be charitable or not.³⁴ In their experiment, Decider chooses between two actions, A and B . The actions generate a known material payoff to Decider, whereas the payoff to Other depends on the state. More precisely, the payoffs are:

- State 1 (non-aligned) payoffs: $A = (6, 1), B = (5, 5)$.
- State 2 (aligned) payoffs: $A = (6, 5), B = (5, 1)$.

In the Baseline treatment, the state is known to Decider, and most subjects choose action B in State 1 and A in State 2. The more interesting treatment is the Hidden Information treatment. There, Decider is not informed about the state, but told that both states are equally likely. Decider is given the choice between privately revealing the state before choosing the action, or to take the action without knowing the state. That is, Decider now takes two actions. The first action is whether to reveal or not. The second action is either A or B . Altogether, Decider has six strategies. Let us denote them $(RAA, RAB, RBA, RBB, NA, NB)$, where RAA denotes “reveal, then play A in both states” and NB denotes “not reveal, then play B ” and so on.

In the Hidden Information treatment almost half of the subjects choose NA . That is, they remain ignorant and pick the selfish action A . The remainder mostly play RBA . The behavior NA is incompatible with sympathy-based preferences, according to which Decider would always want to play RBA . By contrast, as we shall now demonstrate, the duty-based model does admit NA if people are considered to have a right to be uninformed.

In order to conduct the analysis, we must make an additional assumption regarding the blameworthiness of actions that generate uncertain harm and shortage. We assume that the blameworthiness is proportional to expected harm and shortage.

We must also make an assumption regarding the duty to reveal information. Having a right to be uninformed is not the same as lacking responsibility to be informed. We therefore solve the model under the assumption that NA does not violate an obligation, but does violate a responsibility. With probability $1/2$, NA generates a shortage amounting to $(5 + 5) - (6 + 1) = 3$. (For simplicity, we here stick to formulation (3a) rather than (3b).) If Decider decides to reveal, we assume that any choice that hurts Other violates an obligation. Since these outcomes also fail to maximize communal value, they

³⁴See also, among others, Bartling, Engl, and Weber (2014), Feiler (2014), Grossmann (2014), Grossman and van der Weele (2017), Freddi (2021), and Serra and Szech (2021) for further evidence about strategic moral ignorance both in laboratory settings and in practice.

violate a responsibility. For example, the total expected blameworthiness associated with the strategy *RAA* is $(4+3)/2$, where 4 represents harm, 3 represents shortage, and $1/2$ is the probability that the state is 1 (non-aligned). Table 2 summarizes the material pay-offs, Other's harm, the community's shortage, and Decider's utility for all six strategies. Clearly, there are only two undominated choices, *NA* and *RBA*.

<i>a</i>	Non-aligned			Aligned			u_d
	x_d, x_o	h_o	s	x_d, x_o	h_o	s	
<i>RAA</i>	6,1	4	3	6,5	0	0	$6 - \frac{1}{2}(4+3)\delta$
<i>RAB</i>	6,1	4	3	5,1	4	5	$5.5 - \frac{1}{2}(4+3+4+5)\delta$
<i>RBA</i>	5,5	0	0	6,5	0	0	5.5
<i>RBB</i>	5,5	0	0	5,1	4	5	$5 - \frac{1}{2}(4+5)\delta$
<i>NA</i>	6,1	0	3	6,5	0	0	$6 - \frac{1}{2}3\delta$
<i>NB</i>	5,5	0	0	5,1	0	5	$5 - \frac{1}{2}5\delta$

Table 2. Information avoidance experiment

Observation 4 *In the information avoidance situation, people with $\delta < 1/3$ play NA, while people with $\delta > 1/3$ play RBA.*

As claimed, the model's prediction is thus consistent with the behavioral evidence.

The key assumptions for generating our result are that it is not an obligation to reveal while *RBA* is the most responsible choice. As before, a way to evaluate our assumptions is to elicit appropriateness ratings. While the published version of Krupka and Weber (2013) does not address the information avoidance experiment, the earlier draft Krupka and Weber (2008) does. The elicitation reveals that it is considered inappropriate to take the selfish action when the state is known to yield a low payoff for the opponent. However, it is not generally inappropriate to take the selfish action when the impact of the opponent is unknown; on average this action is considered neutral. In other words, there is not a general obligation to be informed, just as we assumed. Is there a social responsibility to be informed? Again the answer is affirmative. According to Krupka and Weber (2008), *RBA* obtains a much higher social appropriateness score than the neutral score obtained by *NA*.

We have also made our own elicitations. Figure 6 reports the averages. We replicate all the key findings from Krupka and Weber (2008): On average, there is no obligation to reveal; non-revelation is somewhat appropriate. It is highly appropriate to reveal, find that the state is non-aligned, and play the unselfish action. By contrast, it is inappropriate to reveal, find that the state is non-aligned, and play the selfish action. In other words, the average ratings justify our above assumptions. The disaggregated data show that there is a minority of about 20 percent that considers it somewhat (13 percent) or very

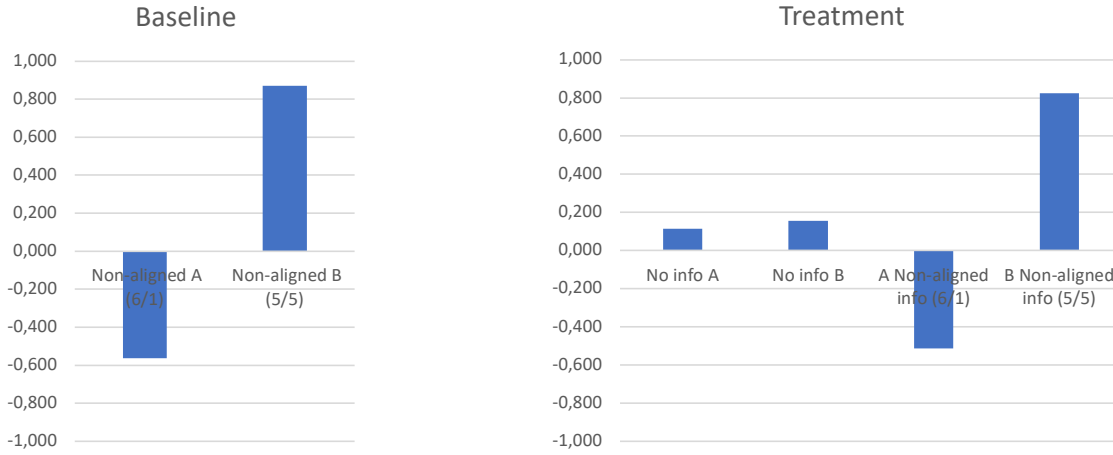


Figure 6. Binary Dictator Experiment and Ignorance: Social Appropriateness Ratings

(4 percent) inappropriate not to reveal the state; remarkably, these fractions are almost identical irrespective of whether the the action choice following non-revelation is *A* (best for Decider) or *B* (best for Other).³⁵

We now turn to two rather different applications.

4.5 Incentive Paradox: The Daycare Experiment

A famous field-experiment by Gneezy and Rustichini (2000) documents that monetary incentives can backfire. The study considers the impact of imposing a penalty on parents who pick up their children late from daycare. If they are more than fifteen minutes late, parents need to pay a penalty corresponding to about eight US dollars in today's currency. Instead of encouraging greater punctuality, the penalty prompted more parents to pick up late. As we shall now see, this outcome is consistent with the model under the mild assumption—which we corroborate shortly—that picking up late is considered a violation of an obligation.

Think of the initial situation between a parent and the daycare center workers. Let β denote the material benefit to the workers from a punctual parent and let κ be the parent's material cost of being punctual. Let *S* denote "sacrifice" (being punctual) and let *N* denote "no sacrifice" (being late), and suppose the latter is considered forbidden. In other words, the set of permitted actions is $\mathcal{P} = \{S\}$. To simplify, set $\gamma = 0$. That is, society only cares about the narrow goal of protecting the daycare center's right to punctuality, not about efficiency or distribution.

The situation is summarized by Table 3. The penultimate column shows that the action *N* is associated with a harm of β . The final column display's the parent's utility. Let $\delta(0)$ denote the threshold level of dutifulness, above which $u_d(S) > u_d(N)$. From

³⁵See Appendix, Table)

a	x_d, x_o	h_o	u_d
S	$-\kappa, \beta$	0	$-\kappa$
N	$0, 0$	β	$-\delta\beta$

Table 3. Daycare experiment, no penalty

a	x_d, x_o	h_o	u_d
S	$-\kappa, \beta$	0	$-\kappa$
N	$-\pi, \pi$	$\beta - \pi$	$-\pi - \delta(\beta - \pi)$

Table 4. Daycare experiment, penalty

Table 3, we see that

$$\delta(0) = \kappa/\beta.$$

Suppose now that the daycare center introduces a penalty $\pi < \beta$ for being late. Since π does not fully compensate the workers' loss, being late is still considered forbidden. Thus, the new situation is captured by Table 4. Note that the penalty has two opposing effects on the parent's utility. On the one hand, it represents a loss of money, but on the other hand it represents a reduction in guilt. The new dutifulness threshold is

$$\delta(\pi) = (\kappa - \pi)/(\beta - \pi).$$

Observe that the threshold coincides with $\delta(0)$ when $\pi = 0$. We say that the penalty π backfires if $\delta(\pi) > \delta(0)$, i.e., if parents must be more dutiful in order to be punctual. Observation 5 states the result of this comparison.

Observation 5 *A penalty $\pi < \beta$ can make a parent less punctual if and only if $\kappa > \beta$.*

In other words, a parent whose cost of being punctual exceeds the workers' benefit from punctuality will become less punctual when there is a fine. The intuition runs as follows. A parent who is punctual despite $\kappa > \beta$ must have a high degree of dutifulness, and hence would feel considerable guilt when imposing harm on the workers by not being punctual. Since the fine serves to compensate the workers, it alleviates the dutiful parent's guilt to such an extent that the guilt reduction outweighs the material cost of paying the fine. The fine offers an avenue for *restitution*.

In order for there to be an aggregate decrease in punctuality, parents with $\kappa > \beta$ must thus constitute a large enough fraction of those parents who are late to begin with.

Note finally that if picking up late were considered to be merely a matter of responsibility, our model's implications are different. In that case, a fine could not backfire,

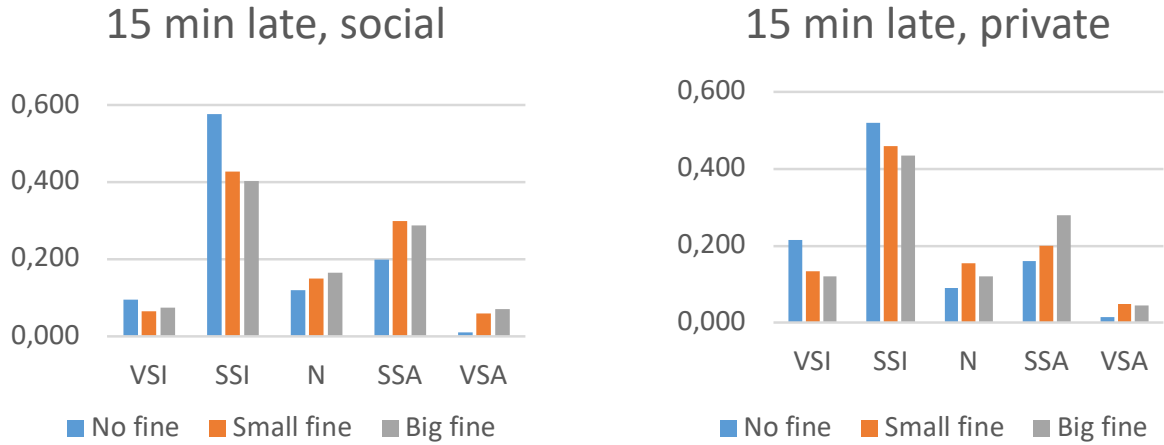


Figure 7. Daycare Experiment: Social and Personal Appropriateness Ratings

because fines do not affect the communal value c as defined by Equation (3a).³⁶ Hence, the fine would not affect the social value v or the parents' blameworthiness.

Gneezy and Rustichini (2000) mention several theories that could shed light on their findings, and social norms is one of them. However, the theory of social norms that they consider is quite different from the restitution theory. It essentially says that the presence of the fine switches off the obligation to be punctual. In terms of our vocabulary, the fine eliminates the daycare center's right to punctuality—making action N permitted (moving N inside the set $\mathcal{P}$). Let us call this the norm-switching theory.³⁷

To learn more about why the fine backfires, we conducted an experiment to elicit social and personal appropriateness ratings, for picking up either 15 or 30 minutes late, and facing either no fine, a small fine of 1 dollar, or a big fine of 10 dollars (see Appendix a for Instructions). Figure 7 displays our key findings.

As predicted both by the restitution hypothesis and the norm-switching hypothesis, picking up kids on time is considered an obligation. Specifically, social ratings and personal ratings both yield more than 90 percent agreement (which is significantly different from 1/2 with practical certainty) that it is somewhat or very inappropriate to come late. Moreover, it is considered more inappropriate when there is no fine than when fines are positive. For example, the average social appropriateness of being 15 minutes late is around -0.28 without a fine, but -0.07 with a small fine, and -0.06 with a large fine. All pairwise differences in behavior between no fine and positive fines are statistically significant at the $p < 0.0001$ level, whereas not all differences between behavior under

³⁶We do not think that Equation (3b) applies here, since the surplus is not of a windfall nature.

³⁷A puzzling feature of the norm-switching theory is that fines are conventionally viewed as signaling inappropriateness rather than appropriateness. For example, the literature on the expressive function of law is based on the view that (in comparison to no law or punishment) a law supported by mild punishment creates deterrence by reinforcing the impression that the action is wrong; see, e.g., Sunstein (1996), Kahan (1997), and Funk (2007).

a	x_d, x_o	h_o	u_d
S	$-\eta, 0$	0	$-\eta$
N	$\kappa, -(\beta + \kappa)$	$0, \beta + \kappa$	$\kappa - \delta(\beta + \kappa)$

Table 5. Lost wallet experiment

small and large fines are statistically significant. The first observation supports both the restitution and the norm-switching hypothesis, whereas the second fails to show decisive support for the restitution hypothesis over the norm-switching hypothesis.

4.6 Lost Wallet Experiment

In a large-scale field experiment conducted by Cohn et al (2019), the researchers turned in “lost” wallets at banks, hotels, post offices, and other public and private service providers around the world. The wallets took the form of transparent envelopes, which included a padlock key (in one treatment the key was removed), a grocery list in the local language, some business cards, and either no money or an amount of money corresponding to 13.5 US dollars, adjusted for local purchasing power. In a smaller number of cases, the money in the envelope amounted to 94 US dollars.

The business cards included the owner’s email address, allowing the authors to measure the fraction of wallets that were honestly handled at the various places. A major finding is that a larger fraction of wallets is returned when they contain money. This finding was not predicted by experts, and it contradicts available models of unselfish behavior based on sympathy.

In the terminology of our model, the experiment can be described as follows. Decider finds Other’s lost wallet. The wallet contains κ units of money and a padlock key worth β to Other (the envelope, the few business cards, and the grocery shopping list presumably has little value to the owner). Returning the wallet costs η for Decider. This cost is the expected hassle of writing an email, engage in correspondence with the owner, and arranging for the return of the wallet. Suppose there is an obligation to return the wallet. That is, $\mathcal{P} = \{S\}$. To simplify, we again set $\gamma = 0$. Thus, actions, material payoffs, harm, and utilities are as described in Table 5.

Comparing the utility from each of the two actions, we see that Decider returns Other’s wallet (plays S) if $-\eta > \kappa - \delta(\beta + \kappa)$, i.e., if

$$\delta > \frac{\eta + \kappa}{(\beta + \kappa)}.$$

The first thing to note is that the condition is easier to satisfy when β is larger. Thus, the model is consistent with the finding that more envelopes are returned when they contain

a padlock key than when they do not. It is also easily checked that the right-hand side is decreasing in κ when $\eta > \beta$. In other words, it requires less dutifulness to make the choice to return the wallet when this inequality is fulfilled.

Observation 6 (i) *The wallet is more likely to be returned if it contains a key.* (ii) *If $\eta > \beta$, wallets with more money are more likely to be returned.*

The intuition for the first result is obvious: The harm from keeping is greater when the envelope contains a key, and there is no benefit from keeping the key. The intuition for the second result is more subtle, since there is a benefit from keeping the extra money: If $\eta > \beta$, it takes substantial dutifulness δ to consider returning the envelope. Thus, for the people who are close to indifferent between keeping or returning the wallet, more money makes the guilt from keeping all of it grow faster than the profit from keeping the wallet.³⁸ We find it likely that many people consider their hassle cost to exceed the owner's value of the padlock key (padlocks often come with spare keys; alternatively, a padlock can be replaced at modest cost), and hence that the inequality is satisfied.³⁹

Is it true that people consider that it is more inappropriate to keep a wallet with money in it than to keep the same wallet when it only contains a padlock key and other low-value items? To investigate this and other questions, our experiment elicited social and personal appropriateness ratings for the lost-wallet situation.

As expected, there is virtual unanimity that it is inappropriate to keep the wallet, and in all conditions a large majority considers keeping to be *very* inappropriate. A drawback of this strong condemnation is that the floor effect makes it hard to test whether it is more inappropriate to keep a wallet with more money. In order to reduce the floor effect, we used seven appropriateness categories instead of five—with the category “inappropriate” inserted between “very inappropriate” and “somewhat inappropriate.” These results are displayed in Figure 8, where “small money” denotes wallets with 13.5 dollars and “big money” denotes wallets with 94 dollars.

Comparing the social appropriateness of keeping a wallet with small money compared to one with no money, the average rating is about 0.05 lower ($p=0.013$). The same comparison between big money and small money yields a difference of about 0.04 ($p=0.009$). However, the personal rating differences are only about half as large, and not significant ($p=0.145$, $p=0.096$).

Let us briefly remark on the vast cross-country differences discovered by Cohn et al (2019) and Tannenbaum et al (2020). The authors document that for wallets containing money the propensity to contact the wallet's owner is around 80 percent in the Nordic

³⁸Of course, it is possible to return the wallet while keeping a fraction of the money. That very rarely occurs in the experiment, and the model suggests a reason: This action is dominated, since it fails to save η and it involves positive guilt.

³⁹A testable implication of the model is that the presence of money would not have the same effect on the return-ratio if the wallet contained an item that is highly valuable to the owner but not to the finder.

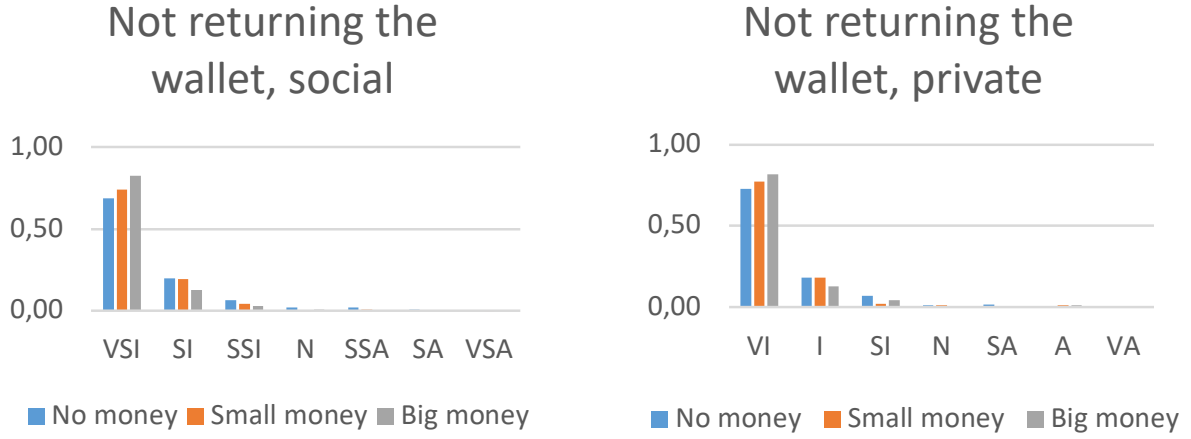


Figure 8. Lost Wallet Experiment: Social and Personal Appropriateness Ratings

countries, but only about 60 percent in the US and 20 percent in China. In our model, such differences can be explained by cross-country variation in the average level of the parameter δ . At first sight, this explanation may seem counter-intuitive, since the Nordic countries are among the least collectivistic countries in the world.⁴⁰ Thus, duty should matter less there, not more. However, a simple resolution to this puzzle is that high individualism goes together with high universalism. People in the Nordic countries experience moral duties toward everyone, rather than merely toward family, close relations, and other in-group members. Accordingly, Tannenbaum et al. (2020) find that a larger fraction of wallets are returned from countries with highly “generalized” or “universalistic” morality. (This is where the indicator-function mentioned at the beginning of Section 2 becomes important.) In short, we sympathise with the long-standing notion that the returning of lost wallets might provide a good measure of “social capital”,⁴¹ and we surmise that such behavior is most accurately ascribed to universalistic dutifulness.

5 Final Remarks

Levitt and List (2007) consider the role that experimentation plays in the natural and social sciences. They argue that social scientists face at least five exclusive challenges: 1) the presence of moral and ethical considerations; 2) the nature and extent of scrutiny of one’s actions by others; 3) the context in which the decision is embedded; 4) self-selection of the individuals making the decisions; and 5) the stakes of the game. The presence of such considerations does not imply that social scientists should give up experimentation, but that there is an important and challenging role for theory in interpreting and

⁴⁰See Inglehart (2018, Chapter 3) for a discussion of the world-wide distribution of the family of values that are defined as individualism, autonomy, and self-expression.

⁴¹See, e.g., Knack and Keefer, 1997.

extrapolating from the results.

Our model of social duties addresses items 1, 3 and 4. It is encouraging to us that it offers a parsimonious explanation for several experimental regularities that other approaches fail to rationalize. Others' entitlements constitute the central concept that is lacking in many previous models of unselfish decision-making. Indeed, the model comes with a built-in measure of entitlements—the point of origin of the Krupka–Weber scale.

Perhaps taking entitlements into account might overcome the weak external validity of laboratory experiments on the domain of unselfish behavior, as documented by Galizzi and Navarro-Martinez (2019) among others? A way to design more externally valid laboratory experiments might be: First measure entitlements for the intended *field* application. Then construct a laboratory experiment in a way that reflects these entitlements. If this approach sufficiently improves external validity, one might even use the laboratory to evaluate policy-proposals: Fine tune the design until the distribution of appropriateness ratings in the lab align with those from the field elicitation. In the final step, consider how potential policy-reforms affect behavior in the lab.

Our analysis has many limitations. Three stand out. First, we make no attempt to account for the effect of stakes (Levitt and List's item 5). Second, we do not take into account that people might have self-serving beliefs about entitlements. Third, and perhaps most importantly, we only apply the model to understand simple situations in which only one person makes a decision. An obvious next step is to consider interactions between several decision makers. However, interactions typically admit reciprocity, which itself is a large topic. Therefore, the analysis of interactive situations awaits a separate paper.

Another avenue for future research is to evaluate the quantitative aspects of the model. To what extent are the person-specific parameters robust across situations? What is the fraction of unselfish behavior that can be ascribed to the duty motive in comparison with other motives? To answer these questions, we need data for the same subjects across a range of different situations.⁴²

Our purpose here is to explain behavior. However, to the extent that dutifulness is malleable through upbringing, education, and other public policies, the model might also be relevant for normative analysis. On one hand, raising dutifulness reduces the tension between common goals and individual desires. On the other hand, duties are constraining. The constraining nature of duties is the reason why liberal thinkers and politicians spent much of the 19th century trying to reduce social pressures on the individual.

Suppose one takes a welfarist approach to the normative question. The social goal is to maximize a (weighted) sum of utilities. Does our utility function strike an acceptable balance between the benefits and costs of raising people's dutifulness? The benefits of increased dutifulness seem relatively uncontroversial. They consist of improvement of

⁴²For a good example of how such an analysis can be performed, see Bruhin, Fehr, and Schunk (2019).

others' material outcomes. But what about the costs? Is it reasonable to include the full cost of compliance with duties as well as the guilt associated with non-compliance in the overall welfare measure? This is the approach taken by DellaVigna, List, and Malmendier (2012) and Andreoni, Rao, and Trachtman (2017). But the issue is controversial. At one end of the spectrum, Nietzsche (1887) would argue that these costs are understated, as the imposition of guilt hampers the individual's natural freedom ("will to power"). At the other end of the spectrum, Hume (1751, Section IX, Part II, final paragraph) would argue that the costs are exaggerated. In his view, the fulfilment of social duties brings a "peaceful reflection on one's own conduct" which is incomparably more worth than "the feverish, empty amusements of luxury and expense". That is, in Hume's view, lack of dutifulness represents either a failure of comprehension or excessive discounting. The welfare analysis of moral education is thus another topic that leaves much room for future research.

References

- Adams S.J. (1965) Inequity in Social Exchange, *Advances in Experimental Social Psychology* 2, 267-299.
- Akerlof, G. and Kranton, R. (2000). Economics and Identity, *The Quarterly Journal of Economics* 115(3): 715-753.
- Andreoni, J. and Bernheim, B.D. (2009). Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects, *Econometrica* 77(5), 1607-1636.
- Andreoni, J., Rao, J.M., and Trachtman, H. (2017). Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving *Journal of Political Economy* 125(3), 625-653.
- Bartling, B., Engl, F., and Weber, R.A. (2014). Does Willful Ignorance Deflect Punishment? - An Experimental Study, *European Economic Review* 70, 512-524.
- Becker, G.S. (1974). A Theory of Social Interactions, *Journal of Political Economy* 82, 1063-1093.
- Bernheim, B.D. (1994). A Theory of Conformity, *Journal of Political Economy* 102(5), 841-877.
- Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge, MA: Cambridge University Press.
- Blais, A. and Galais, C. (2016). Measuring the Civic Duty to Vote: A Proposal, *Electoral Studies* 41(1), 60-69.

- Bolton G.E. and Ockenfels A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition, *American Economic Review* 90: 166-193.
- Brekke, K.A., Kverndokk, S., and Nyborg, K. (2003). An Economic Model of Moral Motivation, *Journal of Public Economics* 9-10, 1967-1983.
- Broberg, T., Ellingsen, T., and Johannesson, M. (2007). Is Generosity Involuntary? *Economics Letters* 94, 32-37.
- Bruhin, A., Fehr, E., and Schunk, D. (2019). The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences, *Journal of the European Economic Association* 17(4), 1025-1069.
- Cappelen, A.W., Hole, A.D., Sørensen, E.Ø., and Tungodden, B. (2007). The Pluralism of Fairness Ideals: An Experimental Approach, *American Economic Review* 97(3), 818-827.
- Cappelen, A.W., Konow, J., Sørensen, E.Ø., and Tungodden, B. (2013a) Just luck: An Experimental Study of Risk Taking and Fairness, *American Economic Review* 103(3), 1398-1413.
- Cappelen, A.W., Nielsen, U.H., Sørensen, E.Ø., Tungodden, B., and Tyran, J.-R. (2013b). Give and Take in Dictator Games, *Economics Letters* 118(2), 280-283.
- Charness, G. and Dufwenberg, M. (2006). Promises and Partnership, *Econometrica* 74, 1579-1601.
- Charness G. and Rabin M. (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics* 117: 817-869.
- Cicero, M. Tullius (44 BC) *De Officiis (On Duty)*. English Translation: Walter Miller, Cambridge MA: Harvard University Press, 1913.
- Cohn, A., Maréchal, M.A., Tannenbaum, D., and Zünd, C.L. (2019). Civic Honesty around the Globe, *Science* 365 (6448), 70-73.
- Cooter, R. (1998). Expressive Law and Economics, *Journal of Legal Studies* 27 (Supplement 2), 585-607.
- Cox, J.C., List, J.A., Price, M., Sadiraj, V., and Samek, A. (2019). Moral Costs and Rational Choice: Theory and Experimental Evidence, manuscript, Georgia State University.
- Crawford, S.E.S. and Ostrom, E. (1995). A Grammar of Institutions, *American Political Science Review* 89(3), 582-600.

- Curry, O.S., Mullins, D.A., and Whitehouse, H. (2019): Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies, *Current Anthropology* 60(1), 47-69.
- Dana, J., Weber, R.A., and Kuang, J.X. (2007). Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness. *Economic Theory* 33: 67-80.
- Dana, J, Cain, D.M., and Dawes, R.M. (2006). What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games, *Organizational Behavior and Human Decision Processes* 100(2), 193-201.
- DellaVigna, S., List, J. A., and Malmendier U. (2012). Testing for Altruism and Social Pressure in Charitable Giving, *Quarterly Journal of Economics* 127(1), 1-56.
- Dillenberger, D. and Sadowski, P. (2012). Ashamed to Be Selfish, *Theoretical Economics* 7(1), 99-124.
- Dreber, A., Ellingsen, T., Johannesson, M., and Rand, D. (2013). Do People Care about Social Context? Framing Effects in Dictator Games, *Experimental Economics* 16, 349-371.
- Durkheim, E. (1958/1900). *Professional Ethics and Civil Morals*, Glencoe IL: Free Press. (Note: The manuscript was completed around 1900, but was first published in French in 1950.)
- Ellingsen T. and Johannesson M. (2008). Anticipated Verbal Feedback Induces Prosocial Behavior, *Evolution and Human Behavior* 29(2), 100-105.
- Enke, B., Rodríguez-Padilla, R., and Zimmermann, F. (2020). Moral Universalism and the Structure of Ideology, NBER Working Paper 27511.
- Evren, Ö, and Minardi, S. (2017). Warm-glow Giving and Freedom to be Selfish, *Economic Journal* 127(603), 1381-1409.
- Exley, C.L. and Petrie, R. The Impact of a Surprise Donation Ask, *Journal of Public Economics* 158, 152-167.
- Fehr, E. and Schmidt, K.M. (1999). A Theory of Fairness, Competition and Cooperation, *The Quarterly Journal of Economics* 114, 817-868.
- Feiler L. (2014). Patterns of Information Avoidance in Binary Choice Dictator Games, *Journal of Economic Psychology* 45, 253-267.
- Freddi, E. (2021). Do People Avoid Morally Relevant Information? Evidence from the Refugee Crisis, *Review of Economics and Statistics* 103(4), 605-620.

- Funk, P. (2007). Is There an Expressive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines, *American Law and Economics Review* 9(1), 135-159.
- Galizzi, M. and Navarro-Martinez, D. (2019). On the External Validity of Social Preference Games: A Systematic Lab-Field Study, *Management Science* 65(3), 976-1002.
- Gelfand, M.J. et al (2011). Differences Between Tight and Loose Cultures: A 33-Nation Study, *Science* 332 (27 May), 1100-1104.
- Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*, Cambridge MA: Harvard University Press.
- Gneezy, U. and Rustichini, A. (2000). A Fine is a Price, *Journal of Legal Studies*, 29, 1-17.
- Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness, *American Journal of Sociology* 91(3), 481-510.
- Grossman, Z. (2014). Strategic Ignorance and the Robustness of Social Preferences, *Management Science* 60(11), 2659-2665.
- Grossman, Z. and Van der Weele, J. (2017). Self-Image and Strategic Ignorance in Moral Dilemmas, *Journal of European Economic Association* 15(1), 171-217.
- Haidt, J. (2008). Morality, *Perspectives on Psychological Science* 3(1), 65-72.
- Henrich, J. et al (Eds.) (2004). *Foundations of Human Sociality*, Oxford: Oxford University Press.
- Hume, D. (1751). *An Enquiry Concerning the Principles of Morals*, London: A Millar.
- Inglehart, R. (2018). *Cultural Evolution: People's Motivations are Changing, and Reshaping the World*, Oxford: Oxford University Press.
- Janoff-Bulman, R., Sheikh, S., Hepp, S. (2009). Proscriptive versus Prescriptive Morality: Two Faces of Moral Regulation, *Journal of Personality and Social Psychology* 96(3), 521-537.
- Janoff-Bulman, R. and Carnes, N.C. (2013). Surveying the Moral Landscape: Moral Motives and Group-Based Moralities, *Personality and Social Psychology Review* 17(3), 219-236.
- Kahan, D.M. (1997). Social Influence, Social Meaning, and Deterrence, *Vanderbilt Law Review* 83, 349-395.
- Kahan, D.M. (2013). Ideology, Motivated Reasoning, and Cognitive Reflection, *Judgment and Decision Making* 8(4), 407-424.

- Kameda, T., Takezawa, M., Tindale, R.S., and Smith, C. (2002). Social Sharing and Risk Reduction: Exploring a Computational Algorithm for the Psychology of Windfall Gains, *Evolution and Human Behavior* 23, 11-33.
- Kandel, E. and Lazear, E.P. (1992). Peer Pressure and Partnerships, *Journal of Political Economy* 100(4), 801-817.
- Kandori, M. (1992). Social Norms and Community Enforcement, *Review of Economic Studies* 59(1), 63-80.
- Kimbrough, E.O. and Vostroknutov, A. (2016). Norms Make Preferences Social, *Journal of the European Economic Association* 14(3): 608-638.
- Knack, S. and Keefer, P. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation, *Quarterly Journal of Economics* 112(4), 1251-1288.
- Konow, J. (1996). A Positive Theory of Economic Fairness, *Journal of Economic Behavior and Organization* 31(1), 13-35.
- Konow, J. (2003). Which Is the Fairest One of All? A Positive Analysis of Justice Theories, *Journal of Economic Literature* 41(4), 1188-1239.
- Korenok, O., Millner, E.L., and Razzolini, L. (2014). Taking, Giving, and Impure Altruism in Dictator Games, *Experimental Economics* 17(3), 488-500.
- Krupka, E.L., Leider, S., and Jiang, M. (2016). A Meeting of the Minds: Informal Agreements and Social Norms. *Management Science* 63(6), 1708-1729.
- Krupka, E.L. and Weber, R.A. (2008). Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? Institute for the Study of Labor, IZA DP No. 3860.
- Krupka, E.L. and Weber, R.A. (2013). Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of European Economic Association* 11(3), 495-524.
- Lazear, E.P., Malmendier, U., and Weber, R.A. (2012). Sorting, Prices, and Social Preferences, *American Economic Journal: Applied Economics* 4(1), 136-163.
- Leventhal, G.S. (1980). What Should Be Done with Equity Theory? In Gergen K.J., Greenberg, M.S., and Willis, R.H. (eds.) *Social Exchange*: New York: Springer, 27-55.
- Levitt, S. and List, J.A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? *Journal of Economic Perspectives* 21, 153-174.

- Lindbeck, A., Nyberg, S., and Weibull, J. (1999). Social Norms and Economic Incentives in the Welfare State, *Quarterly Journal of Economics* 114 (1), 1-35.
- List, J.A. (2007). On the Interpretation of Giving in Dictator Games, *Journal of Political Economy* 115(3), 482-493.
- López-Pérez, R. (2008). Aversion to Norm-breaking: A Model, *Games and Economic Behavior* 64, 237-267.
- Malmendier, U., te Velde, V., and Weber, R.A. (2014). Rethinking Reciprocity, *Annual Review of Economics* 6, 849-874.
- March, J. and Olsen, J.P. (1989). *Rediscovering Institutions*, New York: Free Press.
- Nietzsche, F. (2003/1887). *On the Genealogy of Morals: A Polemic*, translated by H.B. Samuel, New York: Courier Dover Publications.
- Nyborg, K. (2000). Homo Economicus and Homo Politicus: Interpretation and Aggregation of Environmental Values, *Journal of Economic Behavior and Organization* 42, 305-322.
- Ostrom, E. (2000). Collective Action and the Evolution of Social Norms, *Journal of Economic Perspectives* 14(3), 137-58.
- Peer, E., Rothschild, D., Gordon, A., Evernden, Z., and Damer, E. (2021). Data Quality of Platforms and Panels for Online Behavioral Research, *Behavior Research Methods* published online at <https://doi.org/10.3758/s13428-021-01694-3>.
- Riker, W.H. and Ordeshook, P.C. (1968). A Theory of the Calculus of Voting, *American Political Science Review* 62(1), 25-42.
- Saito, K. (2015). Impure Altruism and Impure Selfishness, *Journal of Economic Theory* 158(A), 336-70.
- Sen, A. (2009). Introduction to the 250th Anniversary Edition of *The Theory of Moral Sentiments*, London: Penguin Books.
- Shafir, E. and Tversky, A. (1992). Thinking through Uncertainty: Nonconsequential Reasoning and Choice, *Cognitive Psychology* 24(4), 449-474.
- Smith, A. (1759). *A Theory of Moral Sentiments*, printed for Andrew Millar, in the Strand; and Alexander Kincaid and J. Bell, in Edinburgh.
- Sobel, J. (2005). Interdependent Preferences and Reciprocity, *Journal of Economic Literature* 43(2), 392-436.

- Sunstein, C. (1996). On the Expressive Function of Law, *University of Pennsylvania Law Review* 144, 2021-2053.
- Tannenbaum, D., Cohn, A., Zünd, C.L., and Maréchal, M.A. (2020). What Do Cross-country Surveys Tell Us About Social Capital? University of Zürich Working Paper No. 352.
- Umbeck, J (1977). The California Gold Rush; A Study of Emerging Property Rights, *Explorations in Economic History* 14(3), 197-226.
- Weber, M. (1930/1905). *The Protestant Ethic and the Spirit of Capitalism*, London: Allen & Unwin.

6 Appendix

	Donation										
	0	1	2	3	4	5	6	7	8	9	10
VSI	0,615	0,432	0,277	0,108	0,034	0,007	0,041	0,054	0,081	0,142	0,196
SSI	0,203	0,284	0,399	0,412	0,216	0,007	0,068	0,135	0,196	0,142	0,095
N	0,081	0,108	0,101	0,176	0,230	0,115	0,128	0,203	0,162	0,162	0,162
SSA	0,068	0,155	0,203	0,257	0,426	0,250	0,419	0,270	0,182	0,176	0,135
VSA	0,034	0,020	0,020	0,047	0,095	0,622	0,345	0,338	0,378	0,378	0,412

Table 6. Social Appropriateness Ratings in Standard Dictator Experiment

	Donation											
	Exit	0	1	2	3	4	5	6	7	8	9	10
VSI	0,236	0,671	0,373	0,205	0,081	0,031	0,006	0,019	0,050	0,093	0,168	0,230
SSI	0,261	0,211	0,379	0,484	0,422	0,211	0,012	0,081	0,143	0,168	0,155	0,118
N	0,199	0,056	0,106	0,118	0,261	0,304	0,124	0,199	0,230	0,217	0,174	0,199
SSA	0,174	0,031	0,112	0,174	0,205	0,373	0,230	0,379	0,292	0,205	0,161	0,099
VSA	0,130	0,031	0,031	0,019	0,031	0,081	0,627	0,323	0,286	0,317	0,342	0,354

Table 7. Social Appropriateness Ratings in Dictator Experiment with Exit