The Role of Beliefs in Social Interventions

MASTER THESIS

MSc in Economics and Business Administration
Brand and Communications Management

MSc in Economics and Business Administration
Applied Economics and Finance

Authors:	James Tammesild (123681) Max Munique (125494)
Date of submission:	15.06.2020
Supervisor:	Luigi Butera
Number of pages:	59
Number of characters:	102 709

Abstract

We address the heterogeneity in the effectiveness of social interventions by focusing on one specific mechanism through which it can arise; a person's belief about how others behave. First, we conduct a social norm intervention to increase compliant book return behaviour at the Copenhagen Business School Library. Second, we elicit people's belief about others' compliant book return behaviour to determine if their belief type can predict the effectiveness of a social intervention in a Public Good Game. Third, we elicit people's willingness to pay for social recognition at contribution rank first to last. Our results are threefold. First, participants exposed to a social norm message return 12.5% more of their loaned items on-time. Second, people with a below median belief about others' compliant library behaviour, contribute on average 3.3% more to a public good when exposed to a social recognition intervention than when not. Third, people with a below median belief on average spend 29% of their willingness to pay budget to receive praise when ranked first, whereas participants with an above median belief on average spend 3% for the same preference. Our results help to explain the heterogeneity in the effectiveness of social interventions and suggest that a person's belief can predict the efficacy of a social recognition intervention.

Acknowledgments:

We are indebted to our supervisor Luigi Butera for his outstanding support throughout the project, as well as Anette Koll from CBS Library for her invaluable guidance in setting up our research partnership.

Table of Content

Introduction

Humans care about how they are perceived by others and social approval motivates behaviour. Over the last two decades, companies, governments and nonprofit organisations have increasingly used social interventions to increase the adoption of a desired behaviour. These social interventions often leverage accepted norms and use identification and comparison techniques to change human behaviours. Companies honor the ‘employee of the month’ (Loewenstein, Sunstein, and Golman, 2014; Bursztyn and Jensen, 2017), governments compare on-time tax payers with delinquents (Bø, Slemrod, and Thoresen, 2015; Perez-Truglia and Troiano, 2018) and nonprofit organisations publish donors' names (Karlan and McConnell, 2014). When compared to a traditional monetary incentive, social interventions are a cost-effective and powerful means to change behaviour.

Recent field experiments confirm that social recognition has a significant effect on behaviour across a variety of economically important domains. Examples include, increasing participation in an energy conservation program by allowing participants to identify their neighbours (Yoelia et al., 2003), increasing gym visits by publishing attendance (Butera et al., 2019) and increasing childhood immunization rates by providing different coloured bracelets (Karing, 2018).

Policy-makers have recently shown interest in using behaviourally motivated interventions to ‘nudge’ people toward a certain choice (Thaler and Sunstein 2008). Behavioural insight is being applied to public policy to create more efficient and effective outcomes (OECD, 2017). The research is in regard to individuals' positive or negative cooperation behaviour that gives rise to the ‘public good’ problem. The defining characteristic of a public good is that no person can be excluded from its consumption and therefore all parties are better off if the group members work together to share the cost (Fehr and Fischbacher 2004).

While there is evidence that these types of interventions can increase the adoption of a desired behavior, there is substantial heterogeneity in their effectiveness. For example, a social recognition intervention to motivate bureaucrats' accurate bookkeeping improved the performance of one Nigerian state by 18% but had no effect on another state (Gauri, et al., 2018). The differences in observables did not explain the result's heterogeneity.

In this thesis we investigate one specific mechanism through which such heterogeneity can arise: people's belief about how others behave. A person's reaction to a social recognition intervention may be influenced by their belief about how others behave. Tirole's social signaling theory states that people's utility depends on the expectations others have about their type, based on previous behaviour (Bénabou and Tirole, 2006; Bénabou and Tirole, 2011). When people's prosocial reputation is at stake, it is valuable to avoid being perceived as mainstream and to instead appear as a generous outlier (Bernheim, 1994). There is little empirical evidence to suggest that social signalling works. This paper addresses the question of whether people's belief about how others behave in one domain can predict the efficiency of a social intervention in a related area. The question has important implications for policy-makers when designing a social intervention in an area where information may be difficult to collect or if acquiring the information in itself invalidates the intervention.

We address these issues directly by conducting two field experiments. First, we collaborate with the Copenhagen Business School (CBS) Library to establish the effectiveness of increasing compliant loan behaviour through a social norm intervention. The treatment involved manipulating the email message students receive three days before their item is due to be returned. We complement the social norm intervention with a monetary intervention to measure the relative change in behaviour. Our first experiment establishes that a social norm message is on average effective at increasing compliant book return behaviour whereas a monetary intervention is not. To address the heterogeneity in previous social recognition research, we investigate whether a person's belief about others' compliant behaviour can predict the effectiveness of a social

intervention in a related domain by conducting a second field experiment with a new pool of students.

In the second experiment, we elicit people's belief about others' compliant book return behaviour and split the sample into groups of people with a below median belief and an above median belief. The information establishes whether a person's belief can predict the effectiveness of a social recognition intervention in a Public Good Game (PGG)[1]. Two PGGs measure people's contributions to a communal money pot in an anonymous and a visible game, after completion each player's name, contribution and rank is observable to other players. People are clustered into groups of five and individually ranked in order of contribution to the public good, the first rank is the highest in-group contribution. Finally, a revealed preference schedule measures people's social recognition utility by eliciting their (possibly negative) willingness to pay (WTP) for social recognition at each contribution rank. We calculate the monetary value of receiving praise for being ranked first and avoiding shame for being ranked fifth.

Our results are threefold. First, in the loans return experiment, participants exposed to a social norm message return 12.5% more of their loaned items on-time relative to a control group. In comparison, there is suggestive evidence that participants exposed to a monetary reward intervention return 5.5% fewer loaned items on-time relative to the same control group.

Second, in the PGG, people with a below median belief about others compliant library behaviour contribute on average 3.3% more to the communal pot when being exposed to a social recognition intervention than when not. In comparison, people with an above median belief contribute on average 0.3% more to the communal when being exposed to a social recognition intervention than when not.

[1] At the beginning of the PGG, players are endowed with tokens and decide to keep them for themselves or place the tokens into a communal pot, which is multiplied and evenly divided among players. Experimental economics uses PGGs to measure people's selfish or prosocial behavior toward a public good.

Third, people's belief type was found to have predictive power over the monetary value placed on receiving praise from being ranked first and avoiding shame from being ranked fifth. People with a below median belief on average spent 28.8% of their WTP budget to receive praise when ranked first, whereas participants with an above median belief on average spent 3.1% of their WTP budget for the same preference. In comparison, there is suggestive evidence that people with an above median belief on average spent 18.8% of their WTP budget to avoid shame when ranked fifth, whereas participants with a below median belief on average spent 10.6% of their WTP budget for the same preference.

For a number of reasons, caution should be taken when extrapolating meaning from our results. However, we hope that future research can focus on understanding the predictive nature of a person's belief on the efficacy of a social intervention.

The remainder of the paper is organised as follows. Section 1 is a literature review of the role of beliefs in decision making, monetary and social interventions, social signaling theory and. Section 2 outlines the methodology, data collection, data analysis and considerations. Section 3 presents the analysis and results. Section 4 outlines the discussion, limitations and conclusion with ideas for future research.

Literature Review

1. Decision Making and the Role of Beliefs

1.A. Expected Utility Theory

Before discussing Expected Utility Theory, the concepts of risk and uncertainty should be introduced, as well as earlier concepts of decision theory. A risky decision is one, whose outcome is unknown and according to Fox and Poldrack (2009) risk increases with variance in the probability distribution of possible outcomes, regardless of whether a probable loss is involved. In such decisions (under risk), the decision maker knows precisely the probability distribution of all possible outcomes, akin to flipping a 50/50 coin (Fox and Poldrack, 2009). Decisions under uncertainty however, provide the decision maker with no such information and they must estimate the probabilities of potential outcomes; usually with some degree of ambiguity (Fox and Poldrack, 2009). In the subsequent section, all situations assume decisions under risk.

In its purest form, decision theory dates back to Pascal and Fermet, who stated that a decision maker should always choose the option with the highest expected value (Fox and Poldrack, 2009):

$$E(V) = p * x \qquad (1)$$

Where x is the DKK value with probability p. It is further argued that expected value maximization assumes a risk neutral attitude, rendering decision makers indifferent between a gamble and its expected value. Outlining this as a major flaw of expected value maximization, Fox and Poldrack (2009) point towards the inability to exhibit risk aversion; e.g. why would a person prefer a guaranteed DKK 499 over a 50/50 chance to win DKK 1000.

Bernoulli (1738) addressed this issue, stating that people do not evaluate an outcome according to their objective value but rather by their utility or 'moral value'. He examined that people value x amount of money less when they are wealthy than when they are poor, indicating that marginal utility decreases as wealth increases. Hence, decision makers will choose whatever option offers the highest expected utility, ever dependent on the individual's situation:

$$E(U) = p * u(x) \quad (2)$$

The utility of attaining outcome x, $u(x)$ with probability p. When Von Neumann and Morgenstern (1947) introduced a set of axioms that, as reiterated by Fox and Poldrack (2009), are both necessary and sufficient for representing a decision maker's choice to maximise expected utility, Bernoulli's concept became a central part of economic theory. Pennacchi (2008) confirms that there exists an expected utility function only if the following axioms hold, where ≥ (preferred or equivalent) represents the binary preference relation over the set of all possible outcomes:

1. Completeness: Given the choice between A or B, either A ≥ B, B ≥ A, or both is true.

2. Transitivity: For any three options A, B or C, if A ≥ B and B ≥ C, then A ≥ C. The Completeness and Transitivity axioms establish that decision makers can order their preferences.

3. Continuity: For any three options, a mixture of the best and worst option is preferred to the intermediate option and vice versa. For A ≥ B ≥ C, there exists α, β ∈ (0,1) such that αA + (1 - α)C ≥ B, and B ≥ βA + (1-β)C. The Continuity axiom establishes a continuous trade off between probability and outcomes.

4. Independence (a.k.a. Substitution): When a person prefers one option over another, this inclination should not be affected by a mixture of both options with a common, third option. Hence, for any A, B and C, as well as any $\alpha \in (0,1)$, $A \geq B$ only if $\alpha A + (1-\alpha) C \geq \alpha B + (1-\alpha) C$. The Independence axiom establishes that utilities of outcomes are weighted by their respective probabilities.

Having established the foundations of Expected Utility Theory, it must be mentioned that there exist two violations of said theory, known as the *Allais paradox*. Kahneman and Tversky (1979) presented the following to illustrate the first violation, the Common Ratio effect:

Decision 1: Choose between (A) an 80% chance of DKK 4000 or (B) DKK 3000 for sure.
Decision 2: Choose between (C) a 20% chance of DKK 4000 or (D) a 25% chance of DKK 3000.

To comply with the Independence (Substitution) axiom, people should choose B over A only if they also prefer D over C as $C = 0.25A$ and $D = 0.25B$. Most respondents however, chose B over A and C over D, violating the Independence axiom. This also shows that when the probability of a possible gain moves towards smaller percentages, people prefer the option with a higher outcome.

The second violation, the Common Consequence effect, is illustrated by Kahneman and Tversky (1979) as follows:

Decision 3: Choose between (E) a 33% chance of DKK 2500, a 66% chance of DKK 2400, and a 1% chance of nothing or (F) DKK 2400 for sure.
Decision 4: Choose between (G) a 33% chance of DKK 2500 or (H) a 34% chance of DKK 2400.

In this example, E and G, and F and H are the same if not for the common consequence (Appendix 17). A common consequence added or subtracted to the two options should have no impact on the desirability of one option over the other. Yet, most people in Kahneman and Tversky's study prefer F to E and G to H. Both the Common Ratio effect and Common Consequence effect reiterate the understanding that people are more sensitive to differences in probability near impossibility and certainty (tail events) than in the intermediate range of probability (Fox and Poldrack, 2009).

1.B. The Role of Beliefs

Attitudes are enduring evaluations people have of other objects, ideas or people and can vary in the degree of extremity of positive, negative and neutral (Millon and Lerner, 2003). According to tripartite theory, attitudes are based on cognitive beliefs (eg. "this phone has eight hours of battery life"), feelings (eg. "owning this phone makes me happy") or behaviour (eg. "I have always used an Apple phone") (Breckler, 1984; Rosenberg, 1960; Zanna and Rempel, 2008). Attitudes are formed through cognitive beliefs, feelings and behavior and influence decision making behaviour.

The intention to perform an action precedes the performance of the behaviour (Ajzen, 2008). In line with this reasoning, psychologists agree that intention mediates attitude and behaviour (Bagozzi and Warshaw, 1990; Bentler and Speckart, 1979; Fishbein and Ajzen, 1975; Gollwitzer, 1993; Kuhl, 1985; Locke and Latham, 1990; Triandis, 1977). The causal sequence of belief-attitude-intention hierarchy is used in Ajzen's (1991) planned behavior theory (Appendix 8). The intention to perform a behaviour is influenced by three factors: a favourable or unfavourable evaluation of the behaviour (attitude toward the behaviour), the perceived social pressure to perform or not perform the behaviour (subjective norm) and the self-efficacy of the related behaviour (perceived behavioural control) (Ajzen, 2008). The more favourable the attitude and subjective norm and the greater the perceived behavioural control, the greater the intention to perform the behaviour. The attitude toward the behaviour is determined by the accessible belief about the consequence of the

behavior. The normative beliefs are the perceived behavioural expectations of referent individuals or groups which determine the subjective norm. The control beliefs are the perceived presence of factors that can facilitate or impede the performance of the behaviour. Ajzen's (2008) theory of planned behaviour assumes that intention proceeds the behavioural, normative and control beliefs people have about a behaviour.

In a PGG, Fischbacher and Gächter (2010) investigated how individuals' belief about others' contribution influenced their own contribution. They found that people's belief type had a direct and significant influence on individuals' own contributions. In a multiple-round PGG, people based their belief on their beliefs and observation of others' contributions from the previous period. For example, in period one, participants rely on their intuitive belief about others' contribution and in period two, participants update their belief based on their beliefs and observations of others' contribution in period one (Fischbacher and Gächter, 2010). If conditional contributors observe or believe others free ride, their contributions will decline (Fischbacher, Gächter, and Fehr, 2001; Burlando and Guala 2005; Kurzban and Houser, 2005; Bardsley and Moffatt, 2007; Kocher et al., 2008; Laurent Muller et al., 2008; Duffy and Ochs, 2009; and Herrmann and Thöni. 2009).

2. Monetary Interventions

2.A. The Economic Man

Classical economist Adam Smith championed *Homo Economicus*, whereby human decision making is guided by the instrument of rationality. People are inherently self-interested, expected utility maximisers who achieve y reward by doing x behavior (Elster, 1989.) Rational actors are incentivised by future reward, adapt to changing circumstances and continuously optimise their actions. The assumption is that intrinsic motivation is unaffected when a monetary incentive is introduced, allowing the monetary and non-monetary motivations to be added up (Janssen and Mendys-Kamphorst, 2004).

In microeconomics, *Homo Economicus* is used to model changes in behaviour from changes in price. Frey (2012) explains the price effect; on the demand side a price increase decreases the quantity demanded and on the supply side an increase in price creates an increase in supply.

2.B. Monetary Incentives

A significant amount of literature supports the founding premise of economics that people often change their behaviour in response to an incentive (Angrist et al., 2002; Charness and Gneezy, 2009; Friebel et al., 2017; Gneezy et al., 2011). Monetary incentives have a standard direct price effect that makes the incentivised behaviour more attractive (Gneezy, 2011).

2.C. Crowd-out Effect

However, price effect theory is not applicable to all circumstances and conditions. Sen (1977) and Lane (1991) argue that the theory ignores human behaviour that falls outside of explicit pricing. For example, it does not account for non-market changes in implicit preference which can lead to different behavioural responses to price. Frey (2012) has explored the complex interplay between a monetary incentive and intrinsic motivation. He argues that if individuals perceive the intervention as controlling by limiting their self-determination, self-esteem or self-expression, they will react by reducing their intrinsic motivation toward the activity (Frey, 2012). Therefore, the introduction of a monetary incentive can crowd-out an individual's intrinsic motivation causing a reversal of the price effect and a decrease in the incentivised behavior (Frey, 1994).

Numerous studies have documented the crowd-out effect on prosocial behaviour. Titmuss (1970) found that providing a monetary reward to volunteer blood donors led to less overall blood donation behaviour (Janssen and Mendys-Kamphorst, 2004). He urged economists to rethink the relationship between monetary and non-monetary motivations

because a monetary reward can destroy self-esteem and self-expression to contribute to a public good. Frey and Götte (1999) agree that volunteers complete less work if a monetary incentive is provided. These findings are in line with Gneezy and Rustichini (2000) research that less monetary donations were collected by schoolchildren when provided with monetary performance incentives. To award intrinsically motivated volunteers with monetary compensations can cause a crowd-out effect and reduce overall task performance (Deci, 1975; Deci and Ryan, 1985).

2.D. Social Image

Introducing a monetary incentive can limit intrinsically motivated individuals in receiving social approval, status and recognition. Dickinson (1989) explains that “in our society, people are often admired and praised when they engage in intrinsically controlled behavior (...). The very fact that the behavior is unrelated to any obvious extrinsic consequence is, in part, the basis for the approval” (p. 6). Frey and Jegen (2001) continue by explaining that “an intrinsically motivated person is deprived of the chance of displaying his or her own interest and involvement in an activity when someone else offers a reward” (p. 594). Providing a monetary incentive can prohibit people from expressing themselves and cancel out social status gains which may cause less prosocial behaviour.

Ariely, Bracha and Meier (2009) discuss the interplay between monetary incentives and self-image by using a hypothetical scenario: an individual considers buying an electric car which is more expensive than an equivalent combustion engine car. In a community that values environmentally friendly technologies, driving a car that is recognisable as electric is likely to improve the owner's image. The government decides to provide a large tax benefit to the purchase of electric cars which reduces the cost of the car, making it more attractive to purchase. However, the tax incentive also decreases the social image value gained from driving an electric car. Without the tax incentive, the purchase conveyed the individual's concern for the environment (positive image), with the tax incentive it does so

to a lesser degree. By providing an electric car tax benefit, the government may unintentionally damage the signaling value of buying an electric car. The hypothetical scenario explains the interplay between a monetary incentive and a social signal, which requires careful consideration when designing an incentive.

3. Social Interventions

3.A. The Marketplace

Decisions are made in either a social or monetary marketplace and the introduction of an incentive can transform the nature of the relationship from one domain to the other. Heyman and Ariely (2004) questioned students about assisting others to move furniture. By introducing a monetary incentive, the perception of the interaction was shifted from social to monetary, which reduced student's overall willingness to assist others move. The intrinsic social reward of helping others may have been diminished by the monetary incentive. The type of incentive can also shift the interaction between markets. In another Heyman and Ariely (2004) experiment, students paid in candy (non-monetary) outperformed those paid in cash (monetary), presumably because candy is a social reward that can provide acknowledgement.

3.B. The Social Self

Sociologist Emile Durkheim championed *Homo Sociologicus*, whereby human behaviour is guided by a social norm mechanism (Elster, 1989). Actors are insensitive to circumstance, not outcome oriented and do not consider being rational. Fehr and Fischbacher (2004) define social norms as "standards of behaviour that are based on widely shared beliefs [about] how individual group members ought to behave in a given situation" (p. 185). A norm arises when actions cause positive or negative side-effects for other group members (Fehr and Fischbacher, 2004). They are obeyed voluntarily when the behaviour aligns with

individuals' goals and can be violated when individuals have opposing goals to the normative behaviour (Fehr and Fischbacher, 2004).

3.C. Social Norms

Social norms can enforce cooperation in a public good situation (Arrow, 1970; Ullmann-Margalit, 1977; North, 1981; Andreoni, 1990; Holländer, 1990). Coleman (1990) defines social norms as a rule of behaviour that is enforced by social sanctions of approval or disapproval. Sanctions do not have to be direct or verbal because a social norm can be enforced through internal feelings of self respect and guilt (Lindbeck, 1997). Rege and Telle argue that the mere suspicion that someone disapproves of your behaviour can create a significant social cost (2004). For example, a social norm is that people should not use profanity in public. Therefore, individuals who curse can experience direct and indirect external social sanctions by those who observe the behaviour and internal sanctions of disapproval and feelings of guilt.

Emotions are understood to play a critical role in social norms influence on behaviour (Rilling, et. al., 2002; Fehr and Fischbacher, 2004). Social norms' influence is caused by the strong emotions they trigger at the prospect of being obeyed or violated (Elster, 1989). The anticipation of losing face through negative feelings of shame or guilt can be triggered when a norm is violated (Batson, 1998). Guilt denounces a specific action whereas shame condemns the entire self (Tangney 1995; Tangney and Dearing 2003). Shame is the more socially destructive emotion because it signifies the entire self is bad, whereas guilt is more socially repairable as people are likely to change their behaviour to cause less damage (Batson and Powell, 2003). Contrarily, when a norm is obeyed, the anticipation of the positive feeling of honor can be triggered. Social norms' influence arise from the emotions they trigger which can cause people to perform more good and less selfish deeds.

A social norm can be used to rationalise self-interest. Deutsch (1985) and Messick and Sentis (1983) correlate the relationship between those who abide by and those who benefit from a social norm. They found that people prefer distributive norms that benefit themselves. Norms can be manipulated for personal advantage and post-rationalised to the self and to others. Elster (1989) discovered that when social norms align with self-interest, they acquire special salience in the mind, making them easier to process and comprehend. If self-interest is aligned to a social norm, people are more likely to cooperate. It can be challenging to navigate a smorgasbord of social norms because they may lead to contradiction, compromise or self-sacrifice. Individuals are bound by the norms of society and are unable to indiscriminately select a norm for each situation without justification to the self and to others. If social norms are inconsistently applied, individuals will damage their self-image and reputation.

3.D. Social Recognition

Social recognition has long been understood as a powerful motivation and it refers to situations in which individuals take actions to influence others' beliefs about them. In Hirschman cites (1973) Hobbes explained that "men are continually in competition for honour and dignity" (p. 4). Unlike monetary incentives, social recognition incentives are less likely to crowd-out the intrinsic motivation of agents (Ryan and Deci 2000). Instead, social recognition can galvanise the intrinsic motivation by making the positive attributes more salient (Gauri et al., 2018). Social recognition is a positional good because everyone cannot be ranked first. It generates utility esteem gains to individuals ranked first and disutility shame losses to those ranked last (Butera et al. 2019).

A growing body of literature across business, government and nonprofit organisations demonstrates the power of social recognition to change behaviour. Examples include organisations honoring the employee of the month (Kosfeld and Neckermann, 2011), governments compare on-time tax payers with delinquents (Bø et al., 2015; Perez-Truglia and Troiano, 2018) and nonprofit organisations publish donors' names (Karlan and

McConnell, 2014). Social recognition is an effective behavioural change strategy that can strengthen the motivation to cooperate.

People value their reputation and are willing to pay the price of cooperation in order to maintain their image. In a real world setting, Yoeli et al. (2003) collaborated with a utility company to increase voluntary participation into an electricity blackout-prevention program. The authors compare the natural experiment to a real world PGG because participants forgo a personal cost of less electricity during peak periods, in order to benefit the collective by reducing the amount of electricity blackouts. Yoeli et al. (2003) manipulated whether participants could identify their neighbours that had signed up to the program. The observability of others' actions and identity in signing up to the program tripled cooperation and was four times as effective as the utility company's previous $25 monetary incentive. The research indicates that making people's actions visible to others is a powerful behavior change intervention that can increase cooperation in a real world setting.

Levitt and List (2007) find that the "absence of anonymity will be associated with an increased level of prosocial behaviour relative to settings in which individuals are more anonymous" (p. 161). Hoffman et at. (1994) manipulated the anonymity between experimenter and subject in a dictator game and found that when subjects were aware of being monitored by the researchers, their contributions increased from 16% to 46%. The authors conclude that observed behaviour "may be due not to a taste for fairness, but rather to a social concern for what others may think, and for being held in high regard by others" (p. 371). Davis and Holte (1993) analyse the results and found that participants did not display altruism and instead were "concerned about the opinions of outside observers" (p. 269). In line with these findings, Andreoni and Bernheim (2009) found that as scrutiny increased subjects were more likely to split the pie 50-50. The results illustrate the importance of anonymity between subjects and from researchers to subjects.

The nature and extent of some naturally occuring environments may enable scrutiny between participants and increase the importance of prosocial behaviour (Levitt and List, 2007). For example, churches' open as opposed to concealed collection boxes increase scrutiny on contribution behaviour (Soetevent, 2005). The open collection box allows worshippers to view the total amount collected before contributing, snoop on and be snooped on by their neighbours and to receive scrutiny from the collector. The collection box is an example of a naturally occurring environment that enables scrutiny between people to influence prosocial behaviour.

Rege and Telle (2001) conducted a PGG by instructing participants to write their contribution in private and on a whiteboard in front of the others. The authors found that revealing individuals' behaviour to strangers increased voluntary contributions by 33.8%. An effective strategy to increase contributions is to make people's identity and contributions observable to others because it enables social recognition.

3.E. Crowd-in Effect

Crowding effects can cause positive intrinsic motivation gains if individuals perceive an intervention as supporting their self-determination, self-esteem or expression (Frey, 2012). Interventions that successfully target individuals' internal value system can shift individuals' motivation from the self to others and strengthen their intrinsic motivation to behave prosocially.

3.F. Self-image

People value the opinion others have of them and their self-image. Individuals gain social approval of their behaviour when they signal traits defined by the community as good (Ariely, Bracha and Meier, 2009). Economist Adam Smith (1776) described moral decisions as assessing one's own conduct through the eyes of an 'impartial spectator' who acts as an 'ideal mate within the breast'. In line with Smith (1776), contemporary

psychologists agree that people's behavior is influenced by their desire to maintain conformity between their emotions, thoughts, actions and identity they desire to uphold (Bénabou and Tirole, 2005). Batson (1998) describes the power of self as "the ability to pat oneself on the back and feeling good about being a kind, caring person" (p. 245). Self-image can guide people's behaviour to act in line with society's norms in order to avoid cognitive dissonance.

3.G. Social Signaling

Social singling theory can be used to better understand the motives behind prosocial behaviour. The theory encompasses three endogenous motives; intrinsic, extrinsic and reputation and argues that people seek to evaluate others' values by observing their behaviour (Benabou and Tirole, 2006). The choice architecture can influence the motives, change the meaning attached to the behaviour and the reason to engage in it. An extrinsic monetary reward can damage the reputation motive and cast doubt over the meaning and reason for behaving prosocially.

Social signaling theory suggests that people's utility depends on the expectations others have about their type which is based on previous behaviour (Benabou and Tirole, 2006; Benabou and Tirole, 2011). They behave in order to signal that their taste is close to the mainstream, which causes cooperation to numerous social norms (Bernheim, 1994). However, when people's prosocial reputation is at stake, it is valuable to avoid being perceived as mainstream and to instead appear as a generous outlier (Bernheim, 1994).

An incentive can cause unintended negative consequences, for example it can signal to other group members that the normal behaviour is for people to behave asocial (Gneezy, 2011). In line with this, Benabou and Tirole (2003) found that high powered incentives may convey bad news about the task or the agents ability.

Methodology

1. Research Problem and Hypotheses

The efficacy of social interventions: Can a social norm intervention increase compliant book return behaviour at the CBS Library? And do people's beliefs about others' cooperative behaviour also predict the effectiveness of a social recognition intervention in a PGG?

To answer the research problem we test the following hypotheses:

***H1**: Among participants who are exposed to a monetary gain intervention (vs. control group) more return their library items on-time.*

Students who return their loaned library items on-time go into a draw to win DKK 1,000. We test the effect of a monetary incentive on compliant book return behaviour when compared to a control with information of an experiment in-progress. Economic theory suggests that a monetary incentive has a standard direct price effect that makes the incentivised behaviour more attractive.

***H2**: Among participants who are exposed to a social norm intervention (vs. control group) more return their library items on-time.*

The social norm intervention informed students of the library's normative return behaviour and outlined the negative externalities caused when items are returned late. We test the effect of a social norm intervention on compliant book return behaviour when compared to a control with information of an experiment in-progress. Theory suggests that social norm messages influence is from the strong emotions that are triggered at the prospect of being obeyed or violated.

H3*: Participants with a below median belief about others' compliant loan behaviour have a greater increase in contribution to a PGG when exposed to a social recognition intervention than participants with an above median belief.*

Students' belief about others compliant loan behaviour were elicited and groups were formed by their belief type (below median and above median). We test if people's belief type has an effect on their contribution to a public good when exposed to a social recognition intervention. Social signaling theory suggests that when people's prosocial reputation is at stake, it is valuable to stand out as a generous outlier. Therefore, people with a below median belief about others' prosocial behaviour believe that they will stand out more from the crowd when being recognized in comparison to people with an above median belief.

H4.A*: Participants with a below median belief about others' compliant loan behaviour place a higher monetary value on receiving praise when ranked first in the Personal Recognition Program than participants with an above median belief.*

H4.B*: Participants with an above median belief about others' compliant loan behaviour place a higher monetary value on avoiding shame when ranked last in the Personal Recognition Program than participants with a below median belief.*

Participants' WTP for each contribution rank was elicited to determine the monetary value placed on esteem gains from being ranked first and shame losses from being ranked fifth. We test if people's belief type from the library influences their WTP for social recognition during a PGG. In line with social signaling theory, people with a below median belief are expected to place a higher value on receiving praise from being ranked first when compared to people with an above median belief. And people with an above median belief are expected to place a higher value on avoiding shame when ranked fifth when compared to people with a below median belief.

2. Data Collection

2.A. Experiment One

A between-subject field experiment was conducted in order to determine if participants return a greater amount of loaned CBS Library items when they are presented with an altruism appeal or a monetary reward. The purpose of the experiment is to measure the change in behaviour caused by respectively a social and monetary intervention. Participants were drawn from a convenience sample of CBS students who loaned a library item during the experiment period. Participants were automatically enrolled into the experiment and were informed that they could opt-out of the experiment by replying to our email.

The CBS Library notifies borrowers through an email three days before their item is due to be returned (Appendix 18). We manipulated the email text for the Social Norm, Monetary, Control One and Control Two treatments which ran consecutively for a one week duration (Appendix 19, 20, 21). The email was sent from the CBS Library email account and the participants were made aware of the experiment and its purpose to improve the library's service.

Once the email had been sent, a participant that returned a CBS Library item before the due date and during the Monetary Treatment period was randomly selected to win DKK 1,000. To comply with GDPR law, the CBS Library provided us with a print out of all student book returns during the experiment period. Personal information was replaced with a unique number so that we could select a random winner using an Excel equation and match the number with the participant’s contact details. The winner was asked to provide their MobilePay details in order to receive the monetary reward, thereafter we transferred DKK 1,000 and concluded the experiment.

Control One

The Control One was designed to measure the amount of CBS Library items returned late using the library's existing email text, which ran from the 30/10/2019 - 6/11/2019 and enrolled 231 subjects. The email included information about how loaned items could be renewed and how failure to return items on-time would result in a monetary fine (Appendix 18). It was also stated that the members' account would be blocked from loaning library items if their debt exceeded DKK 200. If the items were returned on-time or renewed, there was no monetary fine. The email text did not specify the monetary fine amount; however, the progressive fine structure can be viewed on the CBS Library's website (Appendix 1).

Control Two

During the Control Two, participants received the same email as Control One with the additional information of an experiment in-progress in the subject line and first paragraph, which ran from the 5/12/2019 - 12/12/2019 and enrolled 194 subjects (Appendix 4). As per Control One, participants were informed that if their library items were returned late, the regular fine structure applied. The purpose of Control Two was to isolate the treatment effect from the information of an experiment in-progress, as well as to create a second control which was adjacent to both treatment runs in order to minimize any interference from seasonality due to shifts in demands at the library. Due to these reasons, we decided not to use Control One in the analysis and only test against the baseline of Control Two.

Monetary Treatment

During the Monetary Treatment, participants received the same email as Control One with the additional information about a monetary reward lottery, which ran from the 27/11/2019 - 04/12/2019 and enrolled 619 subjects. The email subject line and opening paragraph informed participants that they would enter a draw to win DKK 1,000 if they returned their loaned library items on-time (Appendix 19). Participants were also informed that if their

loaned library items were returned late, the regular fine would apply. Participants were unaware of the probability of winning the lottery.

Social Norm Treatment

During the Social Norm Treatment, participants received the same email as Control One with a social norm message, which ran from the 14/12/2019 - 20/12/2019 and enrolled 223 subjects. The email subject line and opening paragraph urged participants to ‘do the right thing’ by the CBS community and to return their loaned library items on-time (Appendix 20: Social Norm Treatment Email Text). Participants were made aware of the negative consequences caused when loaned library items are returned late - it limits what other library members can access and creates additional administration work for the librarians. Participants were informed that if their library items were returned late, the regular fine structure applied.

2.B. Experiment Two

To address the heterogeneity in previous social recognition research, we conduct a second experiment with a new pool of students to investigate whether a person’s belief about others' compliant book return behaviour at the library can predict the effectiveness of a social intervention in a PGG.

Self-Report Survey

For Experiment Two, we designed a self-report survey and enrolled CBS students on-site at CBS Library through convenience sampling. The purpose was to understand participants self-reported loan history and their beliefs about other students’ return behaviour. The survey provided a baseline of information and insight into how participants perceive their own and others’ contribution to a public good. The user experience was

tested with ten participants and improvements to the survey were made before the experiment began.

Over the course of nine days, we enrolled participants into the survey by approaching them directly after they had loaned a library item at Solbjerg Plads, making use of a predetermined script of engagement. Enrolment concluded on 11/03/2020 due to the closure of CBS campus in response to the COVID-19 restrictions. 251 CBS students completed the survey.

Before participants completed the survey using our laptops or smartphones, they received an explanation of the survey's purpose verbally and written in the survey (Appendix 25). Participants were made aware of the experiment's goal to improve the CBS Library's service and were assured that no personal information would be shared with any third parties.

Participants began the survey by providing their CBS email address, so that they could be contacted to play the PGG. Thereafter, the survey elicited students' beliefs about the prosocial behavior of their fellow students; more specifically, they provided a percentage number (in integers) of how many students they believed returned their loaned library items on-time. To conclude the survey, participants were asked about their CBS Library loan history; if they had returned an item late, and if so, why was it returned late and by how many days on average. The survey was conducted in Google Forms, which does not measure the response completion time. However, based on our own time measurements the average survey took two minutes to complete. Participants were encouraged to ask the researchers questions about completing the survey, however no questions were asked. We are confident that valid assessment of the intended construct was achieved.

Public Good Game

A series of web-based one-shot PGGs employing the strategy method were conducted as developed by Fischbacher et al. (2001).

Participants played a Private Game, where participants' names and contribution decisions remained anonymous, and a Public Game, where players' names and contribution decisions were observable to other participants, introducing social recognition as an intervention to change behavior. To estimate a monetary measure of participants' (dis)utility from social recognition, we elicited WTP for social recognition. A within-subject design experiment was used to overcome potential statistical power difficulties due to a low completion rate.

Before starting the data collection, we conducted pilot testing to collect evidence of response process validity to assess whether participants understood the construct as defined by the researchers. Five subjects, representative of the target population, participated in face-to-face cognitive interviews by reviewing the survey items in detail and explaining their thought processes in selecting an answer. Five additional representative subjects received the survey via email, as intended for the target population, and were interviewed ex ante. We reviewed the test results and made adjustments to the survey to make the construct easier to understand. Once the survey was published, the team performed a reliability analysis with the first ten entries by calculating Cronbach's Alpha to test the internal consistency of scale items for the contribution tables and WTP schedule (Cronbach, 1951). The calculation is performed with the below function:

$$\alpha = \frac{k}{(k-1)} * (1 - \frac{\Sigma\sigma_i^2}{\sigma_t^2}) \qquad (4)$$

Where k is the number of items, $\Sigma\sigma_i^2$ is the sum of the item variances and σ_t^2 is the variance of the total scores. We did not expect issues with internal consistency due to the application of the strategy method, where any one question is essentially a modified version of the previous question. In fact, Cronbach's Alpha was 0.98 for the contribution tables as well as for the WTP schedule.

The responses were collected using a Qualtrics survey (Appendix 29), which was distributed to the participants who had completed the phase one self-report survey. Participants were informed that it would take approximately 10 minutes to complete the survey and were asked to read the instructions carefully. Participants were able to start the survey and return to it on another occasion, enabling them to email us a question and complete the survey thereafter. We expected the web-based method to create higher friction for participants, compared to asking questions on-site, so to ensure thoughtful decisions, we did not impose a time limit and offered the chance to gain a relatively high monetary payoff of up to DKK 350 per subject. We selected terminology was assumed to be easily understood by the CBS student population. Before participants provided their name and consent, they were presented with the PGG instructions and were asked to answer all questions carefully, as each had an equal chance to determine the final payoff.

Participants were randomly assigned to groups of 5 and were endowed with 10 tokens, each worth DKK 10. In each scenario, they could either invest their endowment into a private account that counts only to their own payoff, or contribute $c \in [0, 10]$ tokens to a communal pot which pays a Marginal Per Capita Return (MPCR) of 0.5 tokens to all group members. All contributions to the communal pot were summed, multiplied by 2.5 and re-distributed equally to the 5 group members. The MPCR from investing in the public good was perfectly observable by all participants, so there was no Knightian uncertainty in this game. Participant's total income of any scenario s was the sum of the private account and the redistributed communal pot, which can be calculated by:

$$\pi_{si} = 10 - c_i + 0.5 * \sum_{j=1}^{5} c_j \tag{5}$$

Where c_i is player i's contribution to the public good and c_j is player j's contribution to the public good. Furthermore, participants were informed that, as several scenarios would be played out, any total income was independent of the other and no tokens would be accrued over scenarios played. The final payoff would be determined by one, randomly

selected contribution. Examples were provided to assure the payoff structure was clear and participants were informed that only one group would be randomly selected for payment.

After the PGG instructions, participants were informed that they would play both an anonymous game as well as a visible game where players' names and contribution rank would be observable to other participants in the experiment via a temporary web-link. Each participant's contribution rank was determined by the contribution to the public good, relative to that of the other four group members; where first rank equaled the highest in-group contribution to the communal pot and the last rank equaled the lowest in-group contribution. This ranking was only calculated for the one, payout-specific contribution, as outlined in a below paragraph.

The Private Game and the Public Game were displayed in random order for each player, so that the anchoring effect by referencing the previously played game would be balanced out over the sample. Though individuals have different (dis)utility levels from social recognition, the actual payoff as described in equation (X) did not differ between the anonymous and observable games. Subjects were unaware that they would be asked to elicit a one-shot unconditional contribution and a conditional contribution table for both anonymous and observable games before entering the first screen of the PGG. For both the Private and Public Game, subjects first provided their contribution $c \in [0, 10]$ to the public good without the knowledge of other group members' contributions. After players made their unconditional contribution decisions, they filled out a contribution table using the strategy method; subjects provided a contribution to the public good for each average group contribution $\in [0, 10]$, i.e. a vector of contributions. As this was a one-shot game, any contribution table provided by participants that is increasing in the average contribution of others, cannot be due to reputation formation or any kind of repeated game consideration; instead, it can be taken as an unambiguous measure of the participant's willingness to be conditionally cooperative (Fischbacher et al., 2000).

To estimate a money-metric measure of consumer surplus from social recognition, WTP for social recognition was elicited, both for the preference to be visible to others and for the preference to be anonymous. As applied by Butera et al. (2019), the team developed a revealed preference schedule for estimating the social recognition utility function by eliciting participants' (possibly negative) WTP for social recognition, conditional on different possible realized future within-group rankings. This procedure was introduced to participants as the Personal Recognition Program (PRP) (Appendix 36) and only the group that was randomly selected for payment would enter the PRP. To make this method incentive compatible, participants were endowed with an additional five tokens, independent of any previous contribution decisions, which was their 'WTP budget'. Assumed that CBS students are unlikely to have any reference point for WTP for social recognition, the team chose a 'WTP budget' without a middle value on the scale (0, 1, 2, 3, 4, 5) to avoid participants using the average as a reference point.

To generate random assignment, we guaranteed that the PRP responses would be used to determine observability with a 10% chance, whereas with a 90% chance, observability would be determined by a random draw between the Public Game and the Private Game (Butera et al., 2019). To measure social recognition utility from a single-response sequential method, we applied the Becker–DeGroot–Marschak (BDM) method; an incentive compatible procedure used to measure WTP (Becker, DeGroot and Marschak, 1964). Participants made a bid for their preference to be anonymous or observable for each contribution rank in the PGG, which was compared to a value $\in [0, 5]$ randomly generated by Excel. If the WTP bid was greater than the random number, the elicited preference would be implemented. If the bid was less than the random number, observability was determined by random assignment between Public and Private Game.

Upon completing the data collection, a sample size of 65 players was achieved and 13 groups of 5 players were randomly assigned using a combination of Excel functions. To determine if names and ranks would be made observable to others, we randomly selected the Public Game via Excel's RAND() function. Then, each participant's unconditional one-shot contribution for the Public Game was used to calculate each group's average

contribution, upon which each group member's conditional contribution was selected. This contribution determined each participant's potential payoff relative to their group members as well as the within-group rank to be published. This way we were able to incorporate both unconditional and conditional games to calculate the payoff, remaining true to the statement that any answer may be payoff-relevant. Then, Excel functions were used to randomly select one group to receive the monetary payoff and to enter into the Personal Recognition Program. To determine whether respondents' WTP decisions would be implemented, the team matched each player's rank with the corresponding WTP and applied the BDM-method to determine whether the elicited WTP bid was high enough to guarantee that participants' preference for observability would be implemented. Indeed, three out of five participants provided bids that trumped the randomly generated value, respectively. However, for all these players, the RAND() function selected values above 10, meaning that the PRP decisions carried no decision weight in determining observability (10% / 90% rule).

Over the course of approximately two weeks, we updated the email list of participants of the self-report survey and sent out three reminders to maximise enrolment. Within seven days of concluding the experiment the team thanked all participants via email and shared a web-link (Appendix 40) where each player's name, final contribution and rank was visible to all participants of the PGG. The five winners were contacted separately in order to retrieve their MobilePay number and were subsequently paid.

2.C. Design Considerations

Covid-19

During Experiment Two, the CBS Library was closed due to the spread of the Covid-19 virus. The Danish government ordered the immediate closure of all library's and we were uncertain as to when the facility would reopen. Fortunately, the self-reported survey data had been collected, however the experiment could no longer be completed at the CBS

Library. We decided to shift the experiment to an online web based platform in order to investigate if people's belief about others behaviour at the library could predict the efficacy of a social intervention in a related area.

From the Field to the Real World: Contextual Factors

When designing the experiments, we analysed common pitfalls experienced by scholars in the field of social sciences. The purpose of this section is to explain some of the critical laboratory and field experiment limitations and how we designed the experiments in order to reduce adverse impact on the results.

It is helpful to first understand the systematic difference between laboratory and field experiment environments. In the laboratory, researchers benefit from having great control over confounding factors and are able to observe phenomenons that would otherwise be unobservable in the field (Levitt and List, 2007). Laboratory experiment drawbacks include external validity and the environment can often influence subjects' behaviour.
In contrast, field experiments often overcome external validity, however it can be challenging for researchers to find an adequate experiment setting and researchers require intimate knowledge of how the environment operates (Levitt and List, 2007).

Levitt and List (2007) caution researchers' attempts to generalise laboratory experiment results to the real world and identified five areas in which a laboratory setting is systematically different from a natural environment. We analysed Levitt and Lists (2007) recommendations in an attempt to accurately generalise laboratory and field experiment results to the real world.

The Presence of Moral and Ethical Considerations

Levitt and List (2007) develop a utility function to calculate factors that influence decision making in order to compare laboratory and field results. A utility maximising individual

makes a decision i which is affected by moral M and wealth W actions a. The greater the monetary value v at stake, the greater the impact on the individuals wealth.

When individuals perceive a decision as immoral, antisocial or incongruent with their identity, a cost is imposed on the decision maker (Gazzaniga, 2005; Akerlof and Kranton, 2000, 2005). The nonpecuniary moral outcome is impacted by the external cost imposed on others as v, the social norm or law associated with the action as n and the nature and extent of scrutiny by others as s.

$$U_i(a,\ v,\ n,\ s) = M_i(a,\ v,\ n,\ s)\ + W_i(a,\ v) \qquad (6)$$

During the Social Norm Treatment, we communicated the cost imposed on others when an individual's loaned library item is returned late. These negative externalities included less items for others to borrow and additional administration work created for the librarians. To understand people's belief about others library behaviour, participants were asked how many out of one hundred of their peers they thought returned their loaned items on-time. In Experiment Two, during the PGG a social recognition intervention imposes scrutiny on people's contribution by other players. Participants played a public and private game to determine if revealing their identity and contribution influenced the amount participants contributed. Levitt and List (2007) moral components were incorporated into the experiment in order to analyse the trade-off between a moral dilemma and a financial incentive within the field of public good decision making.

The Nature and Extent of Scrutiny of One's Actions by Others

In a laboratory setting, subjects can be hyper-aware of their behaviour being monitored, scrutinised and recorded by others (Levitt and List, 2007). The scholars found that the "absence of anonymity will be associated with an increased level of prosocial behaviour relative to settings in which individuals are more anonymous" (p. 160). Therefore, the nature and extent of scrutiny of one's actions between subjects and from

research-to-subject was carefully considered during the design process and when possible, anonymity was maintained.

The nature and extent of between-subject scrutiny of one's actions can influence behaviour to be more prosocial. During Experiment One, participants were sent an individual email reminder to ensure anonimity between subjects. When participants returned their loaned items to the library, their behaviour could be observed by others at the library. However, the extent of the scrutiny from other subjects and library goers was low because it was difficult to observe if a loaned library item was returned on-time or late. The item return message was displayed on the library's check-out terminal which may be read by other people waiting in line. During the Experiment Two survey enrollment, if people approached the researchers in small groups and agreed to complete the survey, between-subject scrutiny was reduced by dispersing participants into a private area of the library to complete the survey. For the PGG, between-subject anonymity was maintained by sending the enrollment email to participants in BCC (blind carbon copy) which enables an email to be sent to multiple email addresses without the recipients being able to identify one another. Also, the experiment was conducted online which meant that subjects did not have to be in a laboratory together.

Subjects also feel scrutiny of their actions from the researchers conducting the experiment. To eliminate researcher-subject scrutiny in Experiment One, we remain anonymous and were referred to as the 'research team'. However, during Experiment Two, our identity was revealed to the subjects during the face-to-face survey enrollment process. Ideally, the survey would have displayed on the library's check-out terminal, however this was not technically feasible. Therefore, we trade-off our identity in order to enroll survey participants. Some university students recognised us, however no special treatment was awarded to these participants and we deemed these entries as valid. These subjects may have felt additional scrutiny by the researchers to act prosocially. However, this only occurred on five occasions. During Experiment Two, subjects could identify the researchers from the address the email was sent. We intended to use an anonymous

Gmail account, however attempts went to spam. In the email text, we referred to ourselves as the ‘research team’. We acknowledge that participants could identify us during Experiment Two which may have led to researcher-to-subject scrutiny causing individuals to act more prosocially.

The Context in Which the Decision is Embedded

Researchers often lack complete control over the context in which subjects make decisions within an experiment and subtle manipulations of context can change participants' behaviour (Harrison and List, 2004; Roth, 1995; Hertwig and Ortmann, 2001; Bohnet and Cooter, 2005). Levitt and List (2007) distinguish contextual factors within and outside of our control and found that changes to significant contextual factors can predict changes in behaviour.

Contextual factors that were within the control of researchers include the experiment instructions and the framing of information (Levitt and List, 2007). During Experiment Two, the terminology used in the experiment instructions and questions may have influenced the subjects response. For example, Andreoni (1995) found that asking players to “contribute” vs. “allocate funds” and framing the PGG as positive vs. negative changed participants' contributions. We decided to frame the public good as a ‘communal pot’ and to refer to participants as ‘players’ that belong to a ‘group.’ ‘Tokens’ were used to represent Danish Kroner and the ‘Personal Recognition Program’ was used for social recognition. The neutral language was selected to limit the framing effect on biasing participants' behaviour.

Contextual factors that were outside of our control include the experiment location and individuals’ internalised norms and past experiences. Henrich et al. (2005) conducted small scale dictator and PGGs across 15 different communities and found that the context in which the subjects were brought to the game influenced the outcome. We conducted the experiments using a sample of CBS students in Copenhagen, Denmark. The subjects may have been influenced by their internalised norms that arise from socialising in a

Scandinavian society. It is common for CBS students to have a job during their studies where the average income tax is 40%-45% (Study Copenhagen, n.d.). The participants' past experience of paying tax to a public good may have anchored their contribution. The contextual factors that subjects bring to an experiment can influence people's behaviour.

Self Selection

If participants in the laboratory experiment systematically differ from those in the real world setting, attempts to generalize laboratory results to the real world can be compromised (Levitt and List, 2007). Our research relies on a convenience sample of university students whose education may distinguish them from society. For example, students are more likely to be "scientific do-gooders" interested in science and willing to cooperate with researchers and seek social approval (Rosenthal and Rosnow, 1969). Therefore, in Experiment One, we isolated the treatment effect from the researchers social approval by utilising a control with and without information of an experiment underway. Furthermore, during Experiment Two, the social recognition incentive may be less effective in increasing contributions with a non-student sample. Our choice of sample may have caused exaggerated prosocial behaviour relative to a naturally occurring setting.

During Experiment One, participants were required to opt-out of the experiment, however in Experiment Two, participants were asked to opt-in to the experiment. The different selection processes can hinder attempts to generalise results from the laboratory to the field.

In the marketplace, participants are likely to have traits that help them excel in the domain (Levitt and List, 2007). For example, people with strong social preference are likely to self-select out of the stock trading marketplace and instead hire agents with different preferences to manage their finances. The same self-selection bias may be present with students at a library.

The Stakes of the Game

The amount of money at stake during the experiment is important because it can influence people's decision to cooperate or act in self interest. For laboratory games with elements of morality and money, Levitt and List (2007) argue that concerns for the financial payoff increases when the stakes are high and it is intuitive for people to prioritise the financial outcome when the stakes are high and the ethical choice when the stakes are low.

However, Fehr, Fischbacher and Tougareva (2002) conducted competitive labour market experiments and found that fairness did not diminish as the stakes rose. For the high stakes treatment, participants earned on average two to three times their monthly income and the earnings was reduced for the normal stakes treatment by a factor of ten. In social preference experiments, the amount of money at stake can affect people's decision to prioritise the financial or moral outcome.

To determine the stakes of the financial incentives, we assumed that on average the student population earned DKK 130 per hour after tax. During the Experiment One Monetary Treatment, participants who returned their loaned library item on time could win a DKK 1,000 lottery (unaware of the odds). We assumed that the amount was sufficient and a realistic financial incentive to persuade participants to return their loaned item on-time. During Experiment Two, five participants could win up to DKK 350 depending on the decision they made during the PGG. The amounts were assumed to be a sufficient financial incentive to enrol participants and to elicit realistic contribution behaviour. The amount of money at stake during the experiment can influence participants to prioritise the ethical choice when the stakes are low and the monetary payoff when the stakes are high.

2.D. Statistical Models Used in Analysis

Ordinary Least Squares Regression

An Ordinary Least Squares (OLS) regression estimates the unknown parameters in a linear regression model. The regression is fitting a line through the sample points such that the sum of squared residuals is as small as possible. By applying this model one can relate one or more explanatory variables to the dependent variable. For example, how contributions to a PGG vary when looking at participants with a below median belief or an above median belief. A typical OLS equation looks as follows:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (7)$$

Where Y_i is the dependent variable, β_0 is the intercept and β_1 is the slope estimate, the sample covariance between X_i and Y_i divided by the sample variance of X_i. X_i is the explanatory variable and u_i is the residual, an estimate of the error term.

Tobit Model

Censored regression models are linear regressions where the dependent variable is censored above or below a certain threshold. The most commonly used censored regression is the Tobit model. We apply this model e.g. when comparing the difference between public and private contributions to a PGG, since the difference can only range between -10 and 10. The equation for the Tobit model is equal to that of an OLS, where only the parameters change. How is beyond the scope of this project.

Logistic Regression

A logistic regression predicts the probability of an outcome that can only have two values, such as on-time or late book returns, by predicting a binary dependent outcome from a set of independent variables. The model looks as follows:

$$logit\,(\pi_i) = log\,\frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (8)$$

Where π_i is $P(Y_i = 1 \mid X_i)$, the probability of Y_i being equal to 1. All else being equal to a regular linear regression.

One-Sample T-test

A One-Sample T-test compares the mean of a sample population to a predetermined value. In its standard form the One-Sample t-test determines whether the sample mean is statistically different from zero. The equation for the t-statistic is:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \quad (9)$$

Where μ is the proposed constant for the sample mean, $\bar{x}$ is the sample mean, n is the sample size and s is the sample standard deviation.

Wilxocon-Mann-Whitney Test

The Wilcoxon-Mann-Whitney (WMW) test is a non-parametric alternative to the Two-Sample T-test that does not make any assumption about normality for the sample populations. If the sample size is moderately large, the Two-Sample T-test is robust to non-normality due to the central limit theorem. If that is not the case, we revert to the WMW which tests whether the medians of the two samples are statistically different. The equation for the U-statistic is:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad (10)$$

Where n_1 and n_2 are the sample sizes from population 1 and 2, respectively and R_1 is Σ ranks from sample 1. This test was only applied for alternative hypotheses in Appendix 22.

Kendall's Tau Test

Kendall's Tau is a non-parametric measure of correlation between columns of ranked data. The Tau correlation coefficient returns values between 0 and 1, where 0 shows no relationship and 1 shows perfect correlation. The equation for Kendall's Tau is:

$$T_B = \frac{C - D}{C + D} \tag{11}$$

Where C is the number of concordant pairs - where the pair of row items is ranked in the same order in both columns - and D is the number of discordant pairs - where the pair of row items is ranked in the opposite direction in the two columns.

Shapiro-Wilk Test

The Shapiro-Wilk test is a method to determine whether a random sample comes from a normal distribution. The test statistic is calculated as follows:

$$W = \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{12}$$

Where x_i are the ordered random sample values and a_i are constants generated from the covariances, variances and means of the sample from a normally distributed sample. This test was only applied for alternative hypotheses in Appendix 22.

3. Analysis and Results[2]

The analysis for this thesis was performed in R Studio.

Analysis terminology

Experiment One

Control Two	Control treatment, informing subjects of a study in progress
Monetary	Monetary gain treatment, subjects enter the draw to win DKK 1,000 if items are returned on-time
Social norm	Social norm treatment, asking subjects to 'do right' by the library community
On-time / late	This binomial takes on the value 1 if an item was returned late and 0 if an item was returned on-time

Experiment Two

Private	Actions and identity are anonymous
Public	Actions and identity are observable to other participants after completion
Belief type	People with an above median belief think on average that 84.3% of CBS students return loaned items on-time, whereas people with a below median belief think on average that 57.4% return on-time

[2] The analysis for this thesis was performed in R Studio. All outputs are available upon request.

Hypothesis 1

Among participants who are exposed to a monetary gain intervention (vs. control group) more return their library items on-time.[3]

An OLS regression generated suggestive evidence[4] that participants exposed to a monetary gain intervention return on average 5.5% less items on-time than participants exposed to the control, with on average 43.1% and 37.6% late returns, respectively (Figure 1).[5]

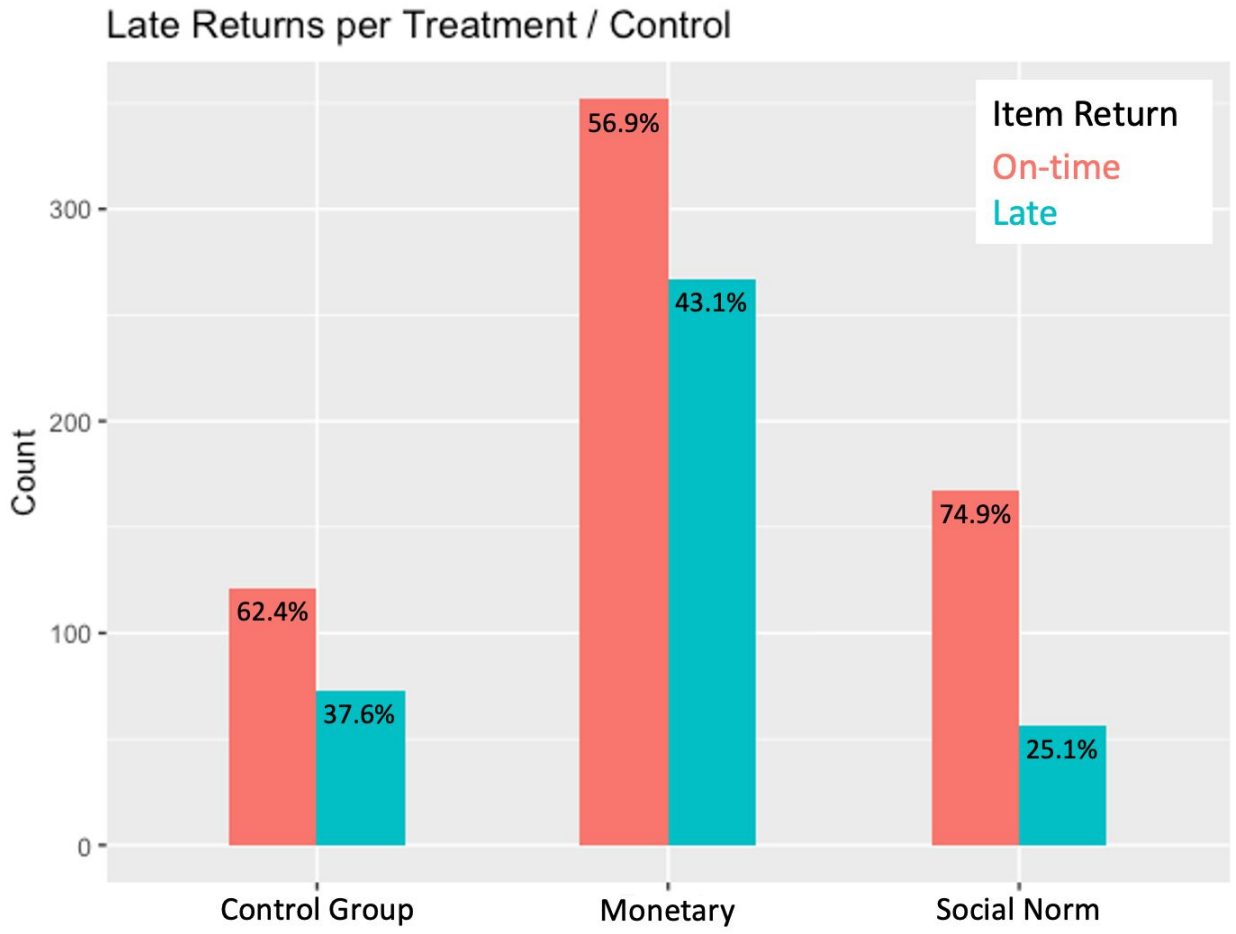

Figure 1: A bar chart shows the count and percentage amount of loaned items returned late (turquoise) and on-time (red) for the control group, the monetary treatment and the social norm treatment.

[3] Refer to Appendix 22 for optional hypothesis about the amount of days an item is returned late
[4] Results that show an effect size, are however not statistically significant
[5] Robustness check in Appendix 23

To further test the above relationship, a non-parametric test was applied since the 'On-time / Late' variable follows a binomial distribution. A Kendall's Tau test established that returning an item late during the Monetary Treatment is not associated with returning late during the control with a Kendall rank correlation coefficient of 0.23 and P-value < 1%.

Hypothesis 2

Among participants who are exposed to a social norm intervention (vs. control group) more return their library items on-time.

An OLS regression determined at 1% significance that participants exposed to a social norm intervention return on average 12.5% more items on-time than participants exposed to the control, with on average 25.1% and 37.6% late returns, respectively (Figure 1).[6]

Since the 'On-time / Late' variable follows a binomial distribution, a non-parametric test was applied to further test the above relationship. A Kendall's Tau test suggested that returning an item late during the Social Norm Treatment is not associated with a late return during the control with a Kendall rank correlation coefficient of 0.04, however a P-value of 60%. Hence, we failed to reject the null that there is no correlation between the two variables.

Hypothesis 3

Participants with a below median belief about others' compliant loan behaviour have a greater increase in contribution to a PGG when exposed to a social recognition intervention than participants with an above median belief.

[6] Robustness check in Appendix 23

Before discussing this hypothesis it should be mentioned that participants with a below median belief about their peers' prosocial behaviour on average stated that 57.4% of CBS students return their loaned items on-time. Participants with an above median belief reported on average that 84.3% return their items on-time, which is slightly higher than the true average of 82.2%, measured by library staff across 2018 and 2019.

H3 was tested by calculating the difference in contribution between the anonymous game (private) minus the observable game (public) and comparing the estimates between participants with a below median belief and an above median belief. This was done for all contributions combined (average) as well as for each scenario in the contribution table (0-10) individually (Figure 2).

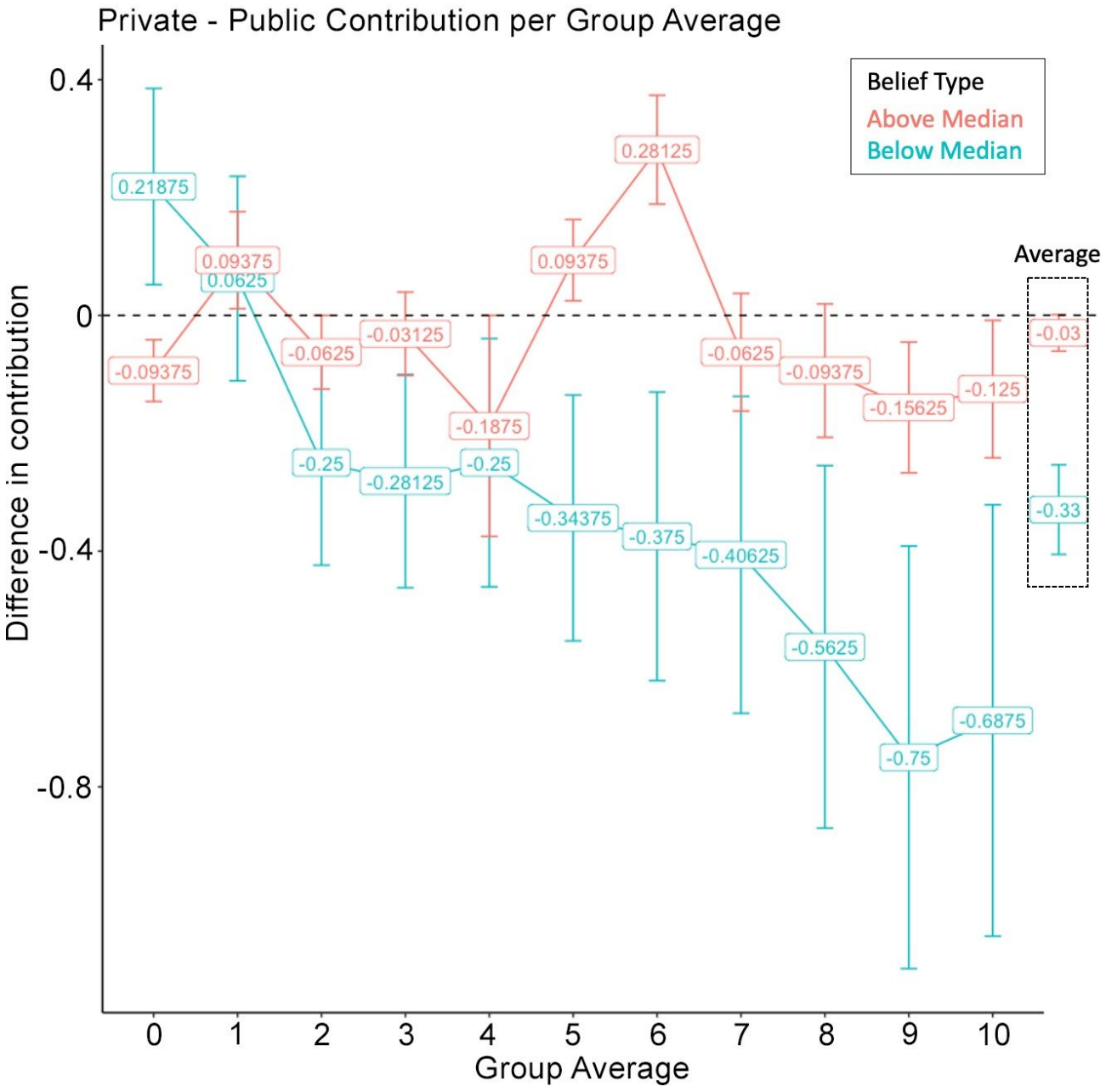

Figure 2: A line chart shows data points, each representing the difference in mean contributions (and standard error bars) to the PGG with and without a social recognition intervention (private minus public). A negative value shows that the average contribution was larger when participants were exposed to a social recognition intervention than when contribution decisions and names were private. The figure shows means and standard error bars for each scenario in the contribution table (0-10) as well as the average across all scenarios; for participants with an above median belief (red) or a below median belief (turquoise).

As the difference in contribution could only take on values between - 10 and 10 (private contribution of 0 minus public contribution of 10 and vice versa), we ran a censored regression using the Tobit model.

Comparing the average across all scenarios in the contribution table, subjects with a below median belief provided on average 0.33 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.03 tokens more. This difference in sensitivity to social recognition of 0.30 tokens (3%) is statistically significant at < 1%.

To complement this finding, we ran a One-Sample T-test for the mean difference between the public and private game with H_0 $\mu = 0$. Below median belief participants' estimate of -0.33 is statistically different from zero (P-value < 1%) and above median belief participants' estimate of -0.03 is not (P-value 32%).

On the condition of an average group contribution of 0 tokens, subjects with a below median belief provided on average 0.22 tokens less when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.09 tokens more. This difference in sensitivity to social recognition of 0.31 tokens (3.1%) is statistically significant at < 10%, which is argued among scholars to serve as suggestive evidence only (> 5%).

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of 0.22 is not statistically different from zero (P-value 20%) and above median belief participants' estimate of -0.09 is statistically different from zero (P-value 8%).

On the condition of an average group contribution of 1 token, subjects with a below median belief provided on average 0.06 tokens less when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.09

tokens less. This difference in sensitivity to social recognition of 0.03 tokens (0.3%) is not statistically significant and is used as suggestive evidence.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of 0.06 is not statistically different from zero (P-value 72%) as is above median belief participants' estimate of 0.09 (P-value 26%).

On the condition of an average group contribution of 2 tokens, subjects with a below median belief provided on average 0.25 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.06 tokens more. This difference in sensitivity to social recognition of 0.19 tokens (1.9%) is not statistically significant and is used as suggestive evidence.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of -0.25 is not statistically different from zero (P-value 16%) as is above median belief participants' estimate of -0.06 (P-value 33%).

On the condition of an average group contribution of 3 tokens, subjects with a below median belief provided on average 0.28 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.03 tokens more. This difference in sensitivity to social recognition of 0.25 tokens (2.5%) is not statistically significant and is used as suggestive evidence.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of -0.28 is not statistically different from zero (P-value 13%) as is above median belief participants' estimate of -0.03 (P-value 66%).

On the condition of an average group contribution of 4 tokens, subjects with a below median belief provided on average 0.25 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.19

tokens more. This difference in sensitivity to social recognition of 0.06 tokens (0.6%) is not statistically significant and is used as suggestive evidence.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of -0.25 is not statistically different from zero (P-value 24%) as is above median belief participants' estimate of -0.19 (P-value 33%).

On the condition of an average group contribution of 5 tokens, subjects with a below median belief provided on average 0.34 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.09 tokens less. This difference in sensitivity to social recognition of 0.43 tokens (4.3%) is statistically significant at < 5%.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of -0.34 is not statistically different from zero (P-value 11%) as is above median belief participants' estimate of 0.09 (P-value 18%).

On the condition of an average group contribution of 6 tokens, subjects with a below median belief provided on average 0.38 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.28 tokens less. This difference in sensitivity to social recognition of 0.66 tokens (6.6%) is statistically significant at < 5%.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of -0.38 is not statistically different from zero (P-value 14%) and above median belief participants' estimate of 0.28 is statistically different from zero (P-value < 1%).

On the condition of an average group contribution of 7 tokens, subjects with a below median belief provided on average 0.41 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.06

tokens more. This difference in sensitivity to social recognition of 0.35 tokens (3.5%) is not statistically significant and is used as suggestive evidence.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of -0.41 is not statistically different from zero (P-value 14%) as is above median belief participants' estimate of -0.06 is statistically different from zero (P-value 54%).

On the condition of an average group contribution of 8 tokens, subjects with a below median belief provided on average 0.56 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.09 tokens more. This difference in sensitivity to social recognition of 0.47 tokens (4.7%) is not statistically significant and is used as suggestive evidence.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of -0.56 is statistically different from zero (P-value 8%) and above median belief participants' estimate of -0.09 is not (P-value 41%).

On the condition of an average group contribution of 9 tokens, subjects with a below median belief provided on average 0.75 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.16 tokens more. This difference in sensitivity to social recognition of 0.59 tokens (5.9%) is not statistically significant and is used as suggestive evidence.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of -0.75 is statistically different from zero (P-value 4%) and above median belief participants' estimate of -0.16 is not(P-value 17%).

On the condition of an average group contribution of 10 tokens, subjects with a below median belief provided on average 0.69 tokens more when being exposed to social recognition, whereas subjects with an above median belief committed on average 0.13

tokens more. This difference in sensitivity to social recognition of 0.56 tokens (5.6%) is not statistically significant and is used as suggestive evidence.

One-Sample T-test with H_0 $\mu = 0$: Below median belief participants' estimate of -0.69 is statistically different from zero (P-value 7%) and above median belief participants' estimate of -0.13 is not (P-value 29%).

Although the Tobit model suggests no statistical significance for many of the above relationships, the standard error (SE) bars in Figure 2 suggest that the mean is statistically different for differences in contribution conditional on group averages of 0, 5, 6, 8, 9 and 10.

Hypothesis 4.A

Participants with a below median belief about others' compliant loan behaviour place a higher monetary value on receiving praise when ranked first in the Personal Recognition Program than participants with an above median belief.

This hypothesis was tested by creating a vector with both the negative WTP values - those to avoid shame - and the positive WTP values - those to receive praise - when being ranked first. The vector displayed every participant's preference and money-metric utility to be recognized or to be anonymous, which could be compared across belief types.

As the WTP could only take on values between 0 and 5, a Tobit model was used to regress the relationship between below median and above median subjects' WTP. Forcing the data into this model distorted the output to a degree where no sensible pattern was detectable however, so we ran an OLS instead.

The linear regression determined that subjects with an above median belief on average had a WTP of 0.156 tokens (3.1%)[7] when ranked first; which indicates that the WTPs to be recognised outweighed the WTPs to be anonymous. Subjects with a below median belief had on average a WTP of 1.438 tokens (28.8%) when ranked first. This difference of 1.281 tokens (25.7%) is statistically significant at < 5% (Figure 3).

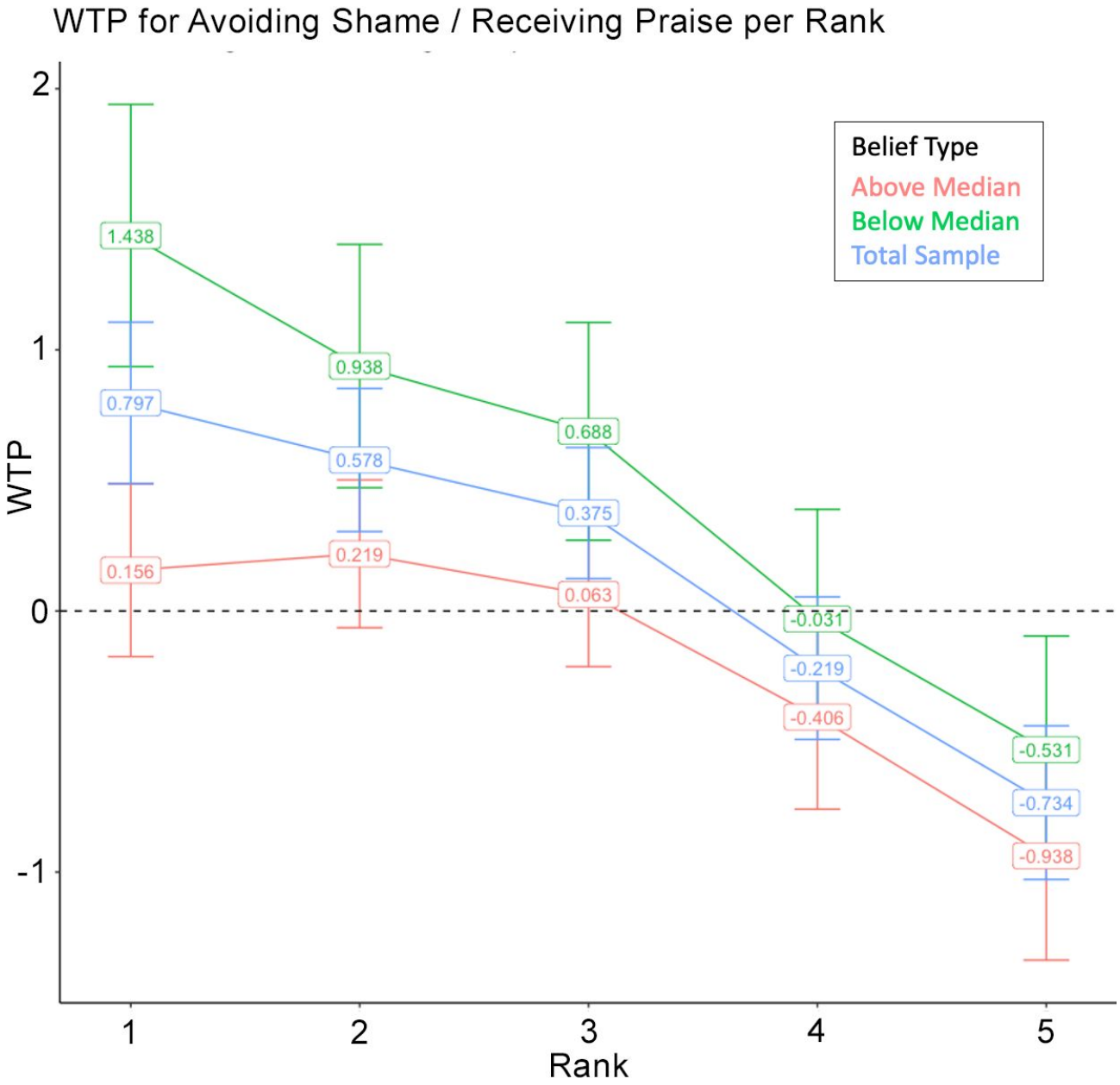

Figure 3: A line chart shows the mean willingness to pay and standard error bars for each respective within-group rank, for participants with an above median belief (red), below median belief (green) as well as

[7] Percentage of WTP budget (5 tokens) endowed to subjects; applies to all percentages shown here

the entire sample (blue). The mean is constructed of negative willingness to pay values which represent the preference to be anonymous at that rank and positive willingness to pay values which represent the preference to be recognized at that rank.

To support this finding we ran One Sample T-tests which determined that below median belief subjects' mean WTP of 1.438 is statistically different from zero at < 1% and above median belief subjects' mean WTP of 0.156 is not statistically different from zero.

Hypothesis 4.B

Participants with an above median belief about others' compliant loan behaviour place a higher monetary value on avoiding shame when ranked last in the Personal Recognition Program than participants with a below median belief.

As for H4.A, an OLS regression was used to determine that subjects with a below median belief on average had a WTP of -0.531 tokens (10.6%) when ranked fifth; which indicates that the WTPs to be anonymous outweighed the WTPs to be recognised. Subjects with an above median belief had on average a WTP of -0.938 tokens (18.8%) when ranked fifth. This difference of 0.406 tokens (8.2%) is not statistically significant and used as suggestive evidence.

Again, to support this finding we ran One Sample T-tests which indicated that above median belief subjects' mean WTP of -0.938 is statistically different from zero at < 5% and subjects with below median beliefs' mean WTP of -0.531 is not statistically different from zero and used as suggestive evidence.

Discussion

The findings of our research suggest that a social norm intervention can increase compliant book return behaviour at the CBS Library. And a person's belief about others' cooperative behaviour can predict the effectiveness of a social recognition intervention in a PGG.

Result One

Participants exposed to a social norm message return 12.5% more of their loaned items on-time relative to a control group (P-value < 1%). In comparison, there is suggestive evidence that participants exposed to a monetary reward intervention return 5.5% fewer loaned items on-time relative to the same control group.

Previous research has documented how a monetary incentive can crowd-out intrinsic motivation and cause less prosocial behavior (Janssen and Mendys-Kamphorst, 2004; Frey and Götte, 1999; Gneezy and Rustichini, 2000; Deci, 1975; Deci and Richard Ryan, 1985). The previous experiments used volunteers as subjects, our results demonstrate a crowd-out effect on prosocial behaviour toward a public good. Literature also suggests that a monetary incentive can cause unintended negative consequences; for example by signaling to the other group members that the normal behaviour is to behave asocial (Gneezy et al., 2011), by conveying bad news about the agent's ability to act prosocial (Benabou and Tirole, 2003), by shifting the perception of the interaction from social to monetary (Heyman and Ariely, 2004) or by damaging the value of the signal that one is good (Ariely et al., 2009).

Our research suggests that policy-makers could benefit from a social norm intervention because it is more effective and cost-efficient in changing behavior when compared to a monetary incentive. Furthermore, crowd theory suggests that interventions that are

perceived as supportive crowd-in intrinsic motivation and interventions that are judged as controlling crowd-out intrinsic motivation (Frey's, 2012). To achieve the crowd-in effect, our research suggests that policy-makers could benefit from surveying participants to ensure that the intervention is perceived as supportive.

Result Two

In a PGG, people with a below median belief about others' compliant library behaviour contribute on average 3.3% more to a communal pot when being exposed to a social recognition intervention (P-value < 1%). In comparison, people with an above median belief contribute on average only 0.3% more to the communal pot when being exposed to a social recognition intervention (suggestive evidence). The result confirms that people's response to a social recognition intervention varies depending on their belief about others' cooperative behaviour.

A growing body of literature across a broad range of domains demonstrates that social recognition can increase prosocial behaviour (Kosfeld and Neckermann, 2011; Bø et al., 2015; Perez-Truglia and Troiano, 2018; Karlan and McConnell, 2014; Yoelia et al., 2003; Butera et al., 2019; Karing, 2018). It is less likely to crowd-out intrinsic motivation (Ryan and Deci, 2000) and can galvanise intrinsic motivation by making positive attributes more salient (Gauri et al., 2018).

Our result varies from Rege and Telle (2001) who conducted a PGG with an in-person social recognition intervention. The in-person environment may explain why the result was different to our experiment conducted online. The absence of anonymity in-person can increase the degree of prosocial behaviour because there is greater scrutiny of one's actions compared to an online environment (Levitt and List, 2007).

Our results are in line with Tirole's social signaling theory because a person's social recognition utility depends on the expectations others have about their type (Benabou and

Tirole, 2006; Benabou and Tirole, 2011). People with a below median belief are more responsive to a social recognition intervention because if they change their behaviour to be more prosocial, they will stand out from how they believe others behave and therefore gain greater utility from the recognition. In comparison, people with an above median belief are less responsive to a social recognition intervention because if they change their behaviour to be more prosocial, they will blend in with how they believe others behave and therefore gain little recognition.

Our results provide support for a shift in public policy from monetary incentives toward greater reliance on esteem based incentives (Brennan and Pettit, 2004). The implications for policy-makers are that a social recognition intervention is effective at increasing contribution toward a public good for people who believe others misbehave. Furthermore, policy-makers could benefit from being aware of the context within which the intervention occurs because applying social recognition online may be less effective than in-person.

Result Three

People's belief type has predictive power over the monetary value placed on receiving praise from being ranked first and avoiding shame from being ranked fifth. People with a below median belief on average spent 28.8% of their WTP budget to receive praise when ranked first (P-value < 1%), whereas participants with an above median belief on average spent 3.1% of their WTP budget for the same preference (suggestive evidence). In comparison, people with an above median belief on average spent 18.8% of their WTP budget to avoid shame when ranked fifth (P-value < 5%), whereas participants with a below median belief on average spent 10.6% of their WTP budget for the same preference (suggestive evidence).

In line with result two and social signaling theory, a person's belief has a predictive power over the monetary value placed on receiving praise and avoiding shame. People with a below median belief place a higher monetary value on gaining praise from being ranked

first when compared to people with an above median belief. In comparison, people with an above median belief place a higher monetary value on avoiding shame from being ranked fifth when compared to people with a below median belief.

Literature supports the finding that social recognition generates utility esteem gains to individuals when ranked first and disutility shame losses to those ranked last (Butera et al. 2019). Furthermore, social recognition can trigger strong emotions of praise or shame and the suspicion that someone disapproves of your behaviour can create a significant social cost (Rege and Telle, 2001).

When designing a social intervention in an area where information may be difficult to collect, our research suggests that policy-makers can benefit from using information from a related domain to predict the effectiveness of a social intervention. For example, if tasked to elicit the risk aversion of a large population, it may not be feasible to use experiments with monetary incentives at scale. A cheaper alternative method is required to collect the attitudes that correlate with risk aversion. Dohmen et al.'s (2011) method elicits people's willingness to take risks 'in general' which is a valid predictor of risk taking behaviour in a specific domain. Our research suggests that policy-makers can address areas in which information is difficult to collect by using information from a related domain.

Under certain circumstances, acquiring information for a social intervention may in itself invalidate the intervention. For example, Manski's reflection problem arises when a researcher observes a group's behaviour to determine whether the average group behaviour influences the individuals that make up the group (Manski, 1993). Our research suggests that policy-makers can avoid invalidating an intervention by using information from a related domain.

Limitations

There are a number of reasons that caution should be taken when extrapolating meaning from our results because the experiment design is not immune from bias which may lead to misguided conclusions.

Experiment One

First, the CBS Library denied us permission to conduct a naturally occurring experiment and insisted that participants be aware of an experiment in-progress. To address the limitation, we isolate the treatment effects by using a control that includes information of an experiment in-progress.

Second, our preference was for the treatments and controls to be conducted simultaneously during the same time period. However, this was not technically feasible because of the CBS Library email system. Similarly, due to a technical issue outside of our control, a 21 day break occurred between the end of Control One and the start of the Monetary Treatment. Therefore, the results would be susceptible to seasonal events, such as students' return behaviour changing during the exam period.

Third, we intended for the treatments and controls to run for a two week period to avoid bias caused by a small sample size. However, when the librarians implemented the experiment in the library's mailing system, technical issues caused a six week delay. One of the researchers required some of the experiment results for a CBS Behavioural Finance exam and therefore the decision was made to reduce the duration of the treatments and controls to a one week period.

Four, our sample is limited to CBS students who use the library and non-student library members were excluded from participating in the experiment. Therefore, the demographic

of our sample is homogeneous and restricted to typically young well educated adults. Our sample selection may frustrate out of sample generalisations.

Five, the CBS Library complies with European GDPR law and does not store students demographic or loan history data. Therefore, we could not utilise the information in our analysis.

Six, all participants may not have read and understood the emails received from the CBS Library and therefore may not have been exposed to the treatment effects. To minimise this caveat, we included the treatment message in the subject line of the email. We were also unable to attain the email open rate analytics from CBS Library due to technical limitations and therefore could not determine the amount of participants that opened the treatment emails.

Seven, participants that returned their items in advance of the three day reminder were not exposed to the treatment effect, however their return was recorded in the intervention period. We tried to address this limitation by excluding the participants who return their items before receiving the treatment. However, the information was not provided by the library and attempts to collect the data ex ante were prohibited because the library was closed due to Covid-19.

Eight, we were unable to guarantee a between-subject experiment design because a participant may have loaned multiple items across different time periods and therefore been part of more than one sample population. The previously mentioned GDPR regulation prohibited us from identifying these participants. However, we assume this event to be unlikely and treat the samples as independent.

Nine, we were provided a hardcopy of each loan item that was marked as either late or on-time. The library had agreed to provide a spreadsheet with additional book return data, however due to technical limitations the information could not be exported. Instead, the

library staff manually processed the data and we believe that some people who were marked as late may have actually renewed their contract. We addressed the limitation by having greater control over the data collection process in Experiment Two.

Ten, the experiment results provided by the library did not display the actual return dates and we could therefore not analyse the amount of days items were returned late. To address this limitation, we manually collected the return dates of ~ 50 late return observations per sample. However, we were prohibited from collecting the data for the entire sample because the CBS Library was closed due to covid-19.

Experiment Two

First, the experiment could not be conducted in a way that participants had to opt-out. The opt-in requirement may have caused a pre-selection bias because the sample may not be representative of the target population.

Second, the web based survey environment caused greater anonymity and less scrutiny of one's actions in comparison to an on-site setting. The experiment context may have caused the social recognition incentive to be less effective at changing behaviour.

Third, the PGG was effortful and took on average 10 minutes to complete, which resulted in a small sample size and limited our ability to achieve statistical significance.

Fourth, we intended to endow subjects with 20 tokens as per Fischbacher et al., (2001) experiment design to gain more data points for the analysis. However, this considerably increased completion time which we deemed unacceptable.

Fifth, the single-shot PGG occured in a context-free theoretical universe with little real world parallels (Sutherland, 2019). In a repeated-play setting with real word context, participants may respond differently to a social intervention.

Conclusion and Future Research

The purpose of our thesis is to establish the validity of social interventions and to investigate the heterogeneity in their effectiveness to change behaviour. We build on social signaling theory by focussing on one specific mechanism through which such heterogeneity can arise: a person's belief about how others behave. To address the research gap, we conduct two field experiments. In Experiment One, we establish that a social norm intervention can increase prosocial behaviour at the CBS Library. In contrast, a monetary intervention reduces compliant book return behaviour. In Experiment Two, we elicit people's belief about others' book return behaviour to determine if a person's belief can predict the effectiveness of a social recognition intervention in a PGG. We determine that beliefs indeed impact the sensitivity to a social recognition intervention, where subjects with a below median belief on average react more intensely to social recognition. To substantiate the impact of beliefs on behaviour, we elicit participants' WTP for social recognition conditional on different possible within-group ranks. We find that individuals with a below median belief on average place a higher value on receiving praise when ranked first than those with an above median belief. In contrast, our results suggest that participants with an above median belief on average have a higher WTP to avoid shame when ranked last, relative to those with a below median belief. Our findings have implications for policy-makers who design interventions in an area where information may be difficult to collect or if acquiring the information in itself invalidates the intervention. Our results suggest that information collected in one domain has predictive power over the behaviour observed in a related domain.

We have identified three areas for future research. First, we are interested in how people's response to a social intervention varies between an online and in-person setting. Second, future research can untangle the numerous contextual factors that may influence a persons' behaviour by replicating our study design in a natural setting. Third, we are interested in how a person's belief about others' behaviour influences the effectiveness of a social norm intervention. By incorporating these considerations, we hope that future

research can develop social interventions that address the ‘public good’ problem in order to improve outcomes for society.

Appendix

Refer to separate PDF.

Bibliography

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211.

Ajzen, I. (2008). Consumer attitudes and behavior. *Handbook of consumer psychology*, 1, 525-548.

Akerlof, G. A., and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715-753.

Akerlof, G. A., and Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic perspectives*, *19*(1), 9-32.

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

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

Andreoni, J. (1995). Cooperation in public-goods experiments: kindness or confusion? *The American Economic Review*, 891-904.

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.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review,* 92(5), 1535-1558.

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

Arrow, K. (1970). Political and economic evaluation of social effects and externalities. *The Analysis of Public Output*, pp. 1-30. NBER.

Bagozzi, R. P., and Warshaw, P. R. (1990). Trying to consume. *Journal of Consumer Research*, *17*(2), 127-140.

Batson, C. D., and Powell, A. A. (2003). Altruism and prosocial behavior. *Handbook of Psychology*, 463-484.

Batson, D. (1998). Altruism and Prosocial Behavior, in D. Gilbert, S. Fiske, and G. Lindzey eds., *Handbook of Social Psychology vol. II. McGraw Hill*, pp. 282-316.

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226-232.

Benabou, R., and Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, *70*(3), 489-520.

Bentler, P. M., and Speckart, G. (1979). Models of attitude–behavior relations. *Psychological Review*, *86*(5), 452.

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy*, 102(5), 841-877.

Bernoulli, D. (1954/1738). Exposition of a new theory on the measurement of risk [translation by L. Sommer of D. Bernoulli, 1738, Specimen theoriae novae de mensura sortis, Papers of the Imperial Academy of Science of Saint Peterburg 5, 175–192]. *Econometrica* 22(1), 23–36.

Bohnet, I., and Cooter, R. D. (2005). Expressive Law: Framing or Equilibrium Selection? John F. *Kennedy School of Government Faculty Research Working Paper*, RWP03-046.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, *47*(6), 1191.

Brennan, G., and Pettit, P. (2004). *The economy of esteem: An essay on civil and political society*. *OUP Oxford*.

Bursztyn, L., and Jensen, R. (2015). How does peer pressure affect educational investments? *The Quarterly Journal of Economics*, *130*(3), 1329-1367.

Butera, L., Metcalfe, R., Morrison, W., and Taubinsky, D. (2019). The deadweight loss of social recognition (No. w25637). *National Bureau of Economic Research*.

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

Bénabou, R., and Tirole, J. (2011). Laws and norms (No. w17579). *National Bureau of Economic Research*.

Bø, E. E., Slemrod, J., and Thoresen, T. O. (2015). Taxes on the internet: Deterrence effects of public disclosure. American Economic Journal: Economic Policy, 7(1), 36-62.

CBS Library, Copenhagen Business School. 2019. *Regulations and fees*. Available at: https://www.cbs.dk/en/library/practical-information/regulations-and-fees. [23-10-2019].

Charness, G., and Gneezy, U. (2009). Incentives to exercise. *Econometrica*, *77*(3), 909-931.

Choi, J. K., and Ahn, T. K. (2013). Strategic reward and altruistic punishment support cooperation in a public goods game experiment. *Journal of Economic Psychology*, *35*, 17-30.

Coleman, J. S. (1990). Foundations of Social Theory. *Cambridge, MA: Belknap Press of Harvard University Press*.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*. 16, 297-334.

Davis, D. D., and Holt, C. A. (1993). Experimental economics. *Princeton University Press.*

Deci, E. L., and Ryan, R. M. (1985). Intrıinsic motivation and self-determination in human behaviour. *NY: Plenum*.

Deci, E. L. (1975). Intrinsic motivation. *New York, NY, US.*

Deutsch, M. (1985). Distributive justice: A social-psychological perspective.

Dickinson, A., (1989). The detrimental effects of extrinsic reinforcement on 'intrinsic motivation'. *The Behavior Analyst,* 12, 1–15.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522-550.

Ellickson, R.C. (2001). The evolution of social norms: a perspective from the legal academy. In Social Norms (Hechter, M. and Opp, K.D., eds), pp. 35-75, *Russell Sage Foundation*.

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

Fehr, E., and Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4), 185-190.

Fehr, E., Fischbacher, U., and Tougareva, E. (2002). Do high stakes and competition undermine fairness? *Evidence from Russia*.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397-404.

Fishbein, M., and Ajzen, I. (1975). Intention and Behavior: An Introduction to Theory and Research. 1975. *Addition-Wesley, Boston, MA*.

Fiske, A. P. (1992). The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological Review*, *99*(4), 689.

Fox, C. R., and Poldrack, R. A. (2009). Prospect theory and the brain. *Neuroeconomics* (pp. 145-173). Academic Press.

Frey, B. S. (1994). How intrinsic motivation is crowded out and in. *Rationality and Society*, *6*(3), 334-352.

Frey, B. S. (2012). Crowding out and crowding in of intrinsic preferences. *Reflexive Governance for Global Public Goods*, *75*, 78.

Frey, B. S., and Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, *15*(5), 589-611.

Frey, B. S., and Goette, L. (1999). Does pay motivate volunteers? *Working paper/Institute for Empirical Research in Economics*, *7*.

Friebel, G., Heinz, M., Krueger, M., and Zubanov, N. (2017). Team incentives and performance: Evidence from a retail chain. *American Economic Review*, *107*(8), 2168-2203.

Gauri, V., Jamison, J. C., Mazar, N., Ozier, O., Raha, S., and Saleh, K. (2018). Motivating bureaucrats through social recognition: evidence from simultaneous field experiments. *The World Bank*.

Gazzaniga, M. S. (2005). The Ethical Brain. *Dana Press.*

Gneezy U. and Rustichini A. (2000). A Fine Is a Price. *The Journal of Legal Studies*, 29(1), 1-17.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and Why Incentives (Don't) Work to Modify Behaviour. *Journal of Economic Perspectives*, *25*(4), 191-210.

Gollwitzer, P. M. (1993). Goal achievement: The role of intentions. *European Review of Social Psychology*, *4*(1), 141-185.

Hardin, G. (1968). The tragedy of the commons. The population problem has no technical solution; it requires a fundamental extension in morality. *Science* 162(3859): 1243–1248.

Harrison, G. W., and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009-1055.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., ... and Henrich, N. S. (2005). “Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(6), 795-815.

Hertwig, R., and Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*(3), 383-403.

Heyman, J., and Ariely, D. (2004). Effort for payment: A tale of two markets. *Psychological Science, 15(11), 787-793.*

Hirschman, A. O. (1973). An alternative explanation of contemporary harriednes. *The Quarterly Journal of Economics*, *87*(4), 634-637.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic behavior*, *7*(3), 346-380.

Holländer, H. (1990). A social exchange approach to voluntary cooperation. *The American Economic Review*, 1157-1167.

Horne, C. (2001). Sociological perspectives on the emergence of norms. In Social Norms (Hechter, M. and Opp, K.D., eds), pp. 3–34, *Russell Sage Foundation*.

Jacquet, J., Hauert, C., Traulsen, A., Milinski, M. (2011). Shame and honour drive co-operation. *Biology Letters* 7(6):899–901.

Janssen, M. C., and Mendys-Kamphorst, E. (2004). The price of a price: on the crowding out and in of social norms. *Journal of Economic Behavior & Organization*, *55*(3), 377-395.

Kahneman, D. and Tversky A (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291.

Karing, A. (2018). Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone. *University of California, Berkeley*.

Karlan, D., and McConnell, M. A. (2014). Hey look at me: The effect of giving circles on giving. *Journal of Economic Behavior and Organization*, 106, 402-412.

Kosfeld, M., and Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, *3*(3), 86-99.

Kuhl, J. (1985). Volitional mediators of cognition-behavior consistency: Self-regulatory processes and action versus state orientation. *Action Control* pp. 101-128. Springer, Berlin, Heidelberg.

Lane, R. E. (1991). *The market experience*. *Cambridge University Press*.

Levitt, S. D., and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, *21*(2), 153-174.

Lindbeck, A. (1997). Incentives and social norms in household behavior. *The American Economic Review*, *87*(2), 370-377.

Locke, E. A., and Latham, G. P. (1990). A theory of goal setting & task performance. *Prentice-Hall, Inc*.

Loewenstein, G., Sunstein, C. R., and Golman, R. (2014). Disclosure: Psychology changes everything. *Annu. Rev. Econ.*, *6*(1), 391-419.

Machina, M. J. (1992). Choice under uncertainty: Problems solved and unsolved. *Foundations of Insurance Economics* pp. 49-82. Springer, Dordrecht.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531-542.

Messick, D. M., and Sentis, K. (1983). Fairness, preference, and fairness biases. *Equity Theory: Psychological and Sociological Perspectives*, *61*, 94.

Millon, T., Lerner, M. J., and Weiner, I. B. (2003). *Handbook of Psychology: Volume 5, Personality and Social Psychology. New Jersey: John Wiley & Sons, Inc*.

North, D. C. (1981). Structure and change in economic history. *Norton*.

Nowak, M., and Highfield, R. (2011). Supercooperators: Altruism, evolution, and why we need each other to succeed. *Simon and Schuster*.

OECD. (2017). Behavioural insights and public policy: Lessons from around the world. *OECD*.

Pennacchi, G. (2008). Theory of Asset Pricing pp. 3-26. *Boston*, *Massachusetts: Pearson Education.*

Perez-Truglia, R., and Troiano, U. (2018). Shaming tax delinquents. *Journal of Public Economics*, 167, 120-137.

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., and Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science*, *325*(5945), 1272-1275.

Rege, M., and Telle, K. (2001). An experimental investigation of social norms (No. 310). *Discussion Papers*.

Rege, M., and Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics*, *88*(7-8), 1625-1644.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., and Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, *35*(2), 395-405.

Rosenberg, M. (1960). An analysis of affective-cognitive consistency. Teoksesta Hovland, C. and Rosenberg, M.(toim.) *Attitude organization and change*.

Rosenthal, R., and Rosnow, L (1969). Artifact in behavioural research. *New York Academic Press*.

Roth, A. E. (1995). Bargaining experiments. *Handbook of Experimental Economics*, 253-348.

Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, 317-344.

Smith, A. (1776). 1776. The Wealth of Nations. *Oxford University Press, Oxford*.

Soetevent, A. R. (2005). Anonymity in giving in a natural context—a field experiment in 30 churches. *Journal of Public Economics*, *89*(11-12), 2301-2323.

Study Copenhagen (n.d.). Taxes – 'SKAT'. Retrieved June 13, 2020, from https://studycph.dk/money/taxes/.

Sutherland, R. (2019). Alchemy: The Dark Art and Curious Science of Creating Magic in Brands, Business, and Life. *HarperCollins.*

Tangney, J. P. (1995). Shame and guilt in interpersonal relationships.

Tangney, J. P., and Dearing, R. L. (2003). Shame and guilt. *Guilford Press.*

Sunstein, C., and Thaler, R. (2008). Improving Decisions About Health, Wealth, and Happiness.

Titmuss, R. M. (1970). The Gift Relationship. *London: Allen and Unwin.*

Triandis, H. C. (1977). Interpersonal behavior. *Brooks/Cole Pub. Co..*

Ullmann-Margalit, E. (1977). The Emergence of Norms. *Oxford University Press.*

Von Neumann, J., and Morgenstern, O. (1945). The Theory of Games and Economic Behaviour, *Journal of Farm Economics,* 27(3), 725-726

Voss, T. (2001) Game theoretical perspectives on the emergence of social norms. In Social Norms (Hechter, M. and Opp, K.D., eds), pp. 105-138, *Russell Sage Foundation*.

West, S. A., Griffin, A. S., and Gardner, A. (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, *20*(2), 415-432.

Yoeli, E., Hoffman, M., Rand, D. G., and Nowak, M. A. (2013). Powering up with indirect reciprocity in a large-scale field experiment. *Proceedings of the National Academy of Sciences*, *110* (Supplement 2), 10424-10429.

Zanna, M. P., and Rempel, J. K. (2008). Attitudes: A new look at an old concept.