INSTITUTE FOR RESEARCH IN ECONOMIC AND FISCAL ISSUES

IREF Working Paper Series

To Follow or not to Follow the Herd? Transparency and Social Norm Nudges

Elena Kantorowicz-Reznichenko
Jaroslaw Kantorowicz

IREF WORKING PAPER NO. 201901

APRIL 2019

IN ENGLISH: EN.IREFEUROPE.ORG
IN FRENCH: FR.IREFEUROPE.ORG
IN GERMAN: DE.IREFEUROPE.ORG

INSTITUTE FOR RESEARCH IN ECONOMIC AND FISCAL ISSUES

To Follow Or Not To Follow The Herd? Transparency and Social Norm Nudges

Elena Kantorowicz-Reznichenko[1]* and Jaroslaw Kantorowicz**

IREF Working Paper February 2019

Abstract

Behavioral interventions in general and nudges in particular have become in recent years a popular regulatory instrument all around the world. Despite the excitement around this policy relevant field, some concerns were raised. Nudges utilize behavioral biases in order to direct an individual's behavior to a certain desired decision by the government. People however, are usually neither aware of the biases nor of the fact those biases are used to influence their behavior. Making nudges transparent is important in democratic societies; yet, this might inhibit their effectiveness. This is the first paper to examine the effectiveness of transparent social norm nudges. We find that unlike with defaults, where transparency has no inhibitive effects, disclosing the way social norms work and the purpose of using them eliminates the positive social norm effect. These results hold only for male participants. Given the proliferation of nudges in public policies around the world, these results call for further research on nudges and transparency.

Keywords: nudge, transparency, regulation, social norms.

1. Introduction

The use of behaviorally informed policies in general and nudges in particular (Thaler & Sunstein 2008) has increased rapidly in recent years in many countries around the world (EU Report 2016; Lunn 2014; Organisation for Economic Co-operation and Development [OECD] 2017; Sunstein 2013; Alemano & Sibony 2015). The idea of nudges, or as it also termed - choice architecture - is to structure the choice, based on known psychological mechanisms, in such a way that a person will make a decision, which is in his/her best interest or increasing the general welfare of society (Thaler & Sunstein 2008, p. 5; Bovens 2009, p. 208). Despite its popularity as an additional regulatory instrument, it also attracted criticism. Supporters of a free market challenge the legitimacy behind and the efficiency of governmental use of nudges (e.g. Glaeser 2006; Schnellenbach 2012; Rebonato 2012). Nudges lead people to make choices, which without the intervention they would avoid. Therefore, the government indirectly manipulates its citizens and meddles with their freedom of choice. However, given the fact that despite this criticism, many governments around the world employ nudges, it is important to investigate whether the use of these nudges can be made "more legitimate".

[1] * Corresponding author. Rotterdam Institute for Law and Economics, Erasmus School of Law, Erasmus University Rotterdam, Institute for Research in Economic and Fiscal issues (IREF) Fellow. ** Institute of Security and Global Affairs and Department of Economics, Leiden University. We are grateful to IREF and the Erasmus University of Rotterdam (EUR Fellowship) for their support in funding this project. We would like to thank Peter Wakker, Robert Dur, Christoph Engel, Klaus Heine, Sophie van der Zee, Jurjen Kamphorst, Enrico Colombatto and the participants of the Experiments at the Crossroads of Law and Economics Workshop in Rotterdam, The Netherlands for their useful comments. In addition, we are grateful to Liam Wells for his editorial assistance with the experimental design.

In this paper, we focus on the criticism about the covert nature of nudges, which might render them manipulative. In majority of cases, nudges utilize psychological mechanisms without the awareness of the targeted decision-maker. Notwithstanding the importance of transparency of public policies in democratic societies, disclosure of the potential influence and the motive behind nudges is rare (Johnson *et al*. 2012; Rebonato 2012, p. 104; Glaeser 2006).

Proposals to introduce meaningful transparency of the employed nudges - where people will be informed about the way the nudge works and the motive behind it - face its own problem. It is strongly assumed that the effectiveness of nudges lies in their latency (Bovens 2009;). Therefore, making the nudge transparent is expected to diminish its effectiveness, what will be called in this paper – “the transparency problem”. One explanation for this effect might derive from the psychological reactance theory. This theory suggests that when people sense their autonomy and freedom of choice is restricted, they might act against the source of this restriction to restore their sense of freedom (Brehm 1966). In the context of nudges, this might mean people will act against the attempted influence by exerting an opposite behavior to the one desired by the choice architect.

One might suggest that if the public generally supports nudges, the lack of transparency in their implementation is not a problem. In recent years, several large-scale international surveys were conducted to examine whether the public in different countries favors nudges. These surveys demonstrate a wide support for nudges in many countries around the world (see for example, Reisch & Sunstein 2016; Sunstein *et al*. 2018a; Sunstein *et al*. 2018b). Despite the general importance of the abovementioned studies for the legitimacy of publicly implemented nudges, it is insufficient to resolve the transparency problem. It is well known that people’s attitudes do not always directly reflect their actions (Wicker 1969). Therefore, it is plausible to assume people might state they support nudges, but then reject a nudge when it is being applied to them. In light of the proliferation of behaviorally informed public policies around the world, it is crucial to investigate issues pertaining to the ethicality of such interventions in a rigorous manner. Furthermore, given the evidence that people prefer transparent nudges over opaque nudges (Osman et al. 2018), it is important to investigate whether inherently covert interventions can be made more transparent while still maintaining their effectiveness.

The trade-off between transparency and effectiveness is an empirical question. Yet, to date it lacks sound empirical evidence. Furthermore, the theoretical foundation of this problem is not clear. To the best of our knowledge, the transparency problem was only investigated in the context of defaults. With this type of nudge, there seem to be a consensus that transparency does not impede the effectiveness of defaults (Loewenstein *et al*. 2015; Kroese *et al*. 2016; Steffel *et al*. 2016; Bruns *et al*. 2018; Paunov *et al*. 2018). Nevertheless, defaults are not the only type of nudges that is being used by governments. The psychological mechanisms underlying the effectiveness of this nudge are different from the psychological channels responsible for the effectiveness of other nudges. Therefore, these results might not be generalized to all types of nudges.

The transparency problem is an important question to investigate from a normative perspective. If nudges can be made transparent and remain effective, like with defaults, then there is no problem implementing them. On the other hand, if meaningful transparency removes the effect of the nudge, this might undermine the claim nudges are truly freedom-preserving interventions. Authorities can experiment with different meaningful types of disclosure to examine whether the effectiveness is removed because people reject the nudge or simply because the provided information itself has a negative effect. However, if all types of meaningful transparency lead people to go against the nudge, the legitimacy of governmental use of those nudges in a democratic society is undermined.

In light of the above, this paper makes the first step in investigating the transparency problem in the context of other nudges besides defaults. In particular, we focus on a social norm nudge. Social norms, which entail presentation of statements regarding what other people do, or think should be done, is a recognized instrument used by governments to direct behavior, e.g. in the context of increasing tax compliance (Hallsworth *et al*. 2017). Due to its importance, and potential application in different contexts (see section 2.2), it is imperative to investigate whether making this nudge transparent will impede its effectiveness. If the effectiveness of social norms is consistently diminished as a result of disclosing the use of this nudge and its psychological channel, a form of transparency which can mitigate this negative impact should be found, or the use of social norms as a regulatory tool might need to be reconsidered all together.

To investigate the transparency problem with respect to social norms, we have conducted an experiment with abstract gamble choices. Participants were presented with two types of lotteries, lottery A and lottery B, and had to choose one of them to play. Lottery A was less risky but with a lower expected return than lottery B. Given most people are risk averse, they were expected to choose lottery A, despite lottery B offering a higher expected return. In order to encourage the choice of lottery B, we have introduced a descriptive social norm. Furthermore, to test the influence of transparency on the effectiveness of the social norm, we have added two more treatments, where: (1) we have informed participants of the way and the intention of using the social norm (called in this paper "transparency" or "simple transparency"), and (2) we have also informed them about the purpose of the nudge, which was to increase their expected return (called in this paper "transparency with purpose" or "full transparency").[2] Due to the rich literature on the gender differences in the context of abstract gamble experiments (i.e. women being more risk averse than men - see, e.g., Eckel & Grossman 2008 - we also measured heterogeneity effects). Finally, we have measured psychological reactance to test for a potential channel to explain the transparency problem, if such is found.

Our findings demonstrate that the social norm was effective only for men. This is consistent with the literature on risk aversion of women in the context of simple gambles, and with the idea of a freedom-preserving nudge (women having a strong preference for a safe investment). On itself, this is an important finding. Given the growing literature on the heterogeneity effects with respect to the effectiveness of nudges (e.g. Bronchetti *et al*. 2013; Beshears *et al*. 2015; Gerber & Rogers 2009), it raises an important policy question – should nudges be adjusted to sub-groups in order to increase their cost-effectiveness? When investigating the influence of transparency on the effectiveness of the nudge (especially for men) we find that it does in fact inhibit the social norm's effectiveness. This is true even for the option of full transparency where not only the psychological mechanism is disclosed, but also the (benign) purpose behind it. Despite the limitations of the study, these are very important findings. At the very least this demonstrates that the comforting results with respect to defaults cannot be generalized to social norms. Therefore, it is crucial to

[2] By "intention" we simply mean that the choice of the social norm is intentional, i.e. given the known effect of social norms, we intentionally use it to increase the probability to choose a certain lottery (and disclose this fact). "Purpose" refers to the goal of increasing the probability of choosing a certain lottery – this lottery has a higher expected return, and thus is more beneficial for the participants.

conduct further research on the interaction between meaningful transparency and the effectiveness of different nudges.

Overall, we did not find support for the psychological reactance theory. Generally, there was no full consistency between people's actions (the probability of choosing lottery B) and their reported experience of reactance. This is consistent with Bruns *et al*. (2018) who examined psychological reactance with respect to defaults and did not find any effects. Therefore, a more comprehensive theoretical framework should be constructed for the transparency problem. Having such theory in place, will allow for reliable predictions and a deeper understanding of which types of nudges necessitate transparency and which types of transparency might mitigate its negative effect.

The paper is structured as follows. In section 2, we introduce the theoretical framework that will serve as the basis for our predictions. The theoretical framework part includes the psychological reactance theory, theories and evidence supporting social norms, and the literature on gender differences with respect to risky choices. This is followed by several hypotheses to be tested in the experiment. Section 3 presents the experimental design, followed by section 4, which presents the results. In section 5, we discuss the results, limitations, policy implications, and future avenues for research.

2. Theoretical Framework

2.1. Nudges and Transparency

Despite the initial excitement around nudges, a serious concern was raised. Governments in democratic societies have an obligation to make their policies transparent to the public. Yet the incorporation of nudges into public policies adds a covert element – nudges exploit cognitive or behavioral biases in order to influence a person's decision. In majority of cases, people are not aware of those biases, nor are they aware of the fact that these biases are intentionally exploited to direct their behavior. This opaque element of nudges might render them as being manipulative (e.g. Hansen & Jespersen 2013, pp. 15-16; Wilkinson 2013), and limiting individuals' autonomy to evaluate, deliberate and chose for themselves (e.g. Hausman and Welch 2010, p. 128).

One potential solution to tackle this criticism is to introduce transparency when using nudges in public policy. Such calls were already explicitly made in official reports (e.g. House of Lords 2011, p. 13; Dutch Scientific Council for Government Policy report 2014, p. 68). However, as Bovens (2009, p. 216) discussed, there can be two types of transparency in this context: (1) *type interference transparency* – the government announcing in general that they are going to use nudges to tackle certain problems, or (2) *token interference transparency* – transparency with respect to each specific nudge. The latter type of transparency is more meaningful and would provide information about the intention behind the nudge, the fact the person is being nudged, and the means through which this nudge is expected to be effective (Hansen & Jespersen 2013, p. 17). Despite its potential desirability from an accountability and legitimacy perceptive, some scholars raise a concern that a token transparency would harm the effectiveness of the nudge (Bovens 2009, p. 217).[3]

[3] Bovens (2009) suggests a compromising type of transparency – *in principle token transparency* – to design every nudge in such a way that an attentive parson would see the manipulation (p. 217).

One psychological mechanism, which may induce the transparency problem, is psychological reactance (Brehm 1966). The underlying idea behind the psychological reactance theory, is that interpersonal threat to freedom steaming from attempts to influence or pressure someone to a certain decision, might lead them to try to restore this sense of freedom (Brehm 1966, p. 10). Psychological reactance theory refers to the everyday behavioral and attitudinal choices people are used to. A threat to freedom in this context refers to the attempt of social influence, which creates a feeling of pressure to exert some behavioral change. The response according to the psychological reactance theory is a decrease in the attractiveness of the "forced" decision. The stronger is the perceived attempt to influence, thus the perceived threat to freedom, the stronger will be the resistance. In extreme cases, it might even lead to a "boomerang effect" when the decision maker will try to reinstate his freedom by going in the opposite direction to the one desired by the influencer. Several important conditions can moderate (or aggravate) the psychological reactance. First, the reactance is weaker if there was no expectation of a free choice. Second, reactance is weaker if the person does not feel sufficiently competent to make the choice, and third, if the choice is not important (Clee *et al*. 1980, pp. 390-391; Brehm & Brehm 1981, pp. 5-6).

The existence and the magnitude of the transparency problem is an empirical question. Yet, despite being theoretically discussed, the empirical evidence is scarce (Marchiori *et al*. 2017, p. 5). To the best of our knowledge, the transparency problem was directly investigated only with respect to nudges in the form of defaults. Scholars have investigated different forms of disclosure in different contexts such as, medical decisions (Loewenstein *et al*. 2015), healthy food (Kroese et al. 2016), and environment and charitable giving (Steffel *et al*. 2016; Bruns *et al*. 2018). In all these studies, transparency did not influence the effectiveness of the default. Moreover, one study in the context of course and experimental studies enrolment presented evidence that transparency may even enhance the effect of defaults (Paunov *et al*. 2018). Given the variety of contexts and methods to investigate the transparency problem, it seems safe to assume that introduction of transparency does not inhibit the effectiveness of nudges in the form of default rules.

Despite the importance of the empirical evidence on the influence of transparency in the context of defaults, it cannot be directly generalized to other types of nudges. The psychological mechanisms responsible for the effectiveness of different choice architectures are not the same. For instance, the underlying mechanisms of defaults, and the reasons people are sticking to them are because they work as recommendations, or a reference point, or simply constitute the effortless option (Dinner *et al*. 2011; Johnson & Goldstein 2003). Social norms on the other hand, exploit people's tendency for conformity (see section 2.2). Therefore, it is possible that transparency has no influence on the effectiveness of some nudges, yet inhibits such effectiveness with respect to other types of nudges. Coming back to our example, people might see defaults as recommendations by people who know more than them on a particular topic, or simply be indifferent with the respective choice and follow the status quo. In such cases, making this mechanism salient might not evoke any negative feelings that will impede the effectiveness of defaults. On the other hand, with respect to social norms, explaining to someone that he is being treated as a conformist because most people do what others do, might touch upon person's self-perception and evoke a negative reaction. Therefore, the current paper is an important first step into the investigation of the transparency problem with respect to another prominent nudge – social norms.

2.2. Social Norms

The ability of actions and opinions of others, i.e. social norms, to influence individual's decisions is a well-established phenomenon in social psychology. Its extreme power is interestingly illustrated in the famous experiment by Asch (1956) where subjects followed the opinion of the majority when it was clearly wrong. Even though traditionally, social norms are understood as only moral prescriptions, theories of social influence emphasize two different meanings of the term "norm". Norm means something which is socially desirable, but also something which is simply common and normal (Cialdini et al. 1991, p. 203). This interpretation led to the definition of two types of social norms: 1) *Injunctive norms,* which refer to the normative statements of what is moral and *ought* to be done, implying social sanctions and assisting people to determine which behavior is socially acceptable and which is not. An injunction norm would be for example, "people should pay taxes." 2) *Descriptive norms* describe what typical behavior *is*, or which actions are taken by others, and often does not have as such a moral value (Cialdini *et al*. 1990). An example of a descriptive norm would be, "90% of people already paid their taxes". The psychological mechanisms, or the motivational reasons, that underlie the effectiveness of these two types of social norms are different. One is considered to be a normative social influence and the other is informational social influence (Deutsch & Gerard 1955). Normative or injunctive norms are effective by signaling the moral rules that a person should follow. The individual motivation to follow these rules is the belief of the consequential social rewards or punishments. On the other hand, descriptive norms serve as a decisional and informational short cut. The motivation of individuals to follow what other people are doing is simply the belief the actions of the majority represent an effective and adaptive behavior. In other words, people believe in the wisdom of the crowd (Cialdini *et al*. 1990, p. 1015; Cialdini *et al*. 1991, p. 203). One should note that descriptive norms have an effect on people's behavior even when they are entirely neutral and bear no moral value. The example of Asch's (1956) experiment illustrates this. The choice of one line over the others when determining which line is longer, has no moral significance. Other studies also demonstrated the influence of morally neutral behavior of others on the actions of the induvial in, the choice of a product (Venkatesan 1966), looking at the sky (Milgram *et al*. 1969), assessing the range of movement of a light (Sherif 1963, pp. 98-117).

After many years of investigating the different elements, which are responsible for the effect and the power of social norms, it was adopted as one of the instruments used by governments (or other entities) to influence people's behavior. For example, social norms were found to be useful in increasing tax compliance (Bott *et al*. 2017; Hallsworth *et al*. 2017). Another example is the exploitation of social influence to reduce smoking (Thaler & Sunstein 2008, p. 68). Different channels may be responsible for the effect of social norms: it can be a peer effect where the decision maker assumes others have private (better) information, or conformity due to the social costs of deviating from what is perceived as acceptable behavior in the society (Beshears *et al*. 2009). Both types of social norms, injunctive and descriptive, are used as nudges.

The evidence for the success of social norms as nudges is prevalent, yet not without challenges. Two large-scale field experiments found a substantial increase in timely tax payment as a result of social norm messages. In particular, in the first experiment, the authors in this study (Hallsworth *et al*. 2017) tested different variations of messages sent to people in the UK who were delayed in paying their taxes. They found that a descriptive social norm with moral implications ("*Nine out of ten people in the UK pay their tax on time. You are currently in the very small minority of people who have not paid us yet*") increased tax payment by 5.1 percentage points, which translates to

£4.9 million during the testing period. In the second experiment, they compared the influence of descriptive (e.g. "*The great majority of people in the UK pay their tax on time*") versus injunctive (e.g. "*The great majority of people agree that everyone in the UK should pay their tax on time*") social norms and found that descriptive norms had significantly larger effect on tax compliance as compared to injunctive norms. Another large scale and successful implementation of social norms are the OPOWER energy conservation programs. In these programs, OPOWER sent letters to a large number of households across the U.S. providing them information about their energy usage as compared to their neighbors, with a statement whether this household is more or less efficient than their neighbors (descriptive norm). In addition, it categorized the household by stating one of the words "Great", "Good" or "Below average" (an injunctive norm). Allcott (2011) evaluated these programs and found that social norms decreased energy consumption by approximately 2%. The author estimated that this effect is equivalent to the effect of increasing the price of energy by 11-20% in the short run or by 5% in the long run. The effectiveness of social norms was found also in other domains, such as voting (Gerber & Rogers 2009), and charitable giving (Frey & Meier 2004).

Contrary to those studies, Richter *et al*. (2018), applying social norms in the context of food consumption and sustainability in Norway and Germany and Silva and John (2017) examining social norms in the context of late tuition fees payment in the UK found no evidence of the effectiveness of descriptive social norms in enhancing desirable behavior. Moreover, Beshears *et al*. (2009) found a boomerang effect of a descriptive social norm, which led to the opposite behavior than the (desired) peer behavior. In their study, the authors examined the effect of a descriptive social norm about the participation in a savings plan and the size of the savings. They found that a sub-group of subjects in fact decreased their savings when they were exposed to the description of their peers who saved more.

One potential explanation for the mixed results with respect to the effectiveness of social norms is the different context. It is plausible that social norms work in some contexts but not in others. Furthermore, there might be some heterogeneity effects where some groups have stronger preferences against a particular nudge than others. Consequently, this might seem as if the nudge has generally "failed" when in fact it was effective for one group but not for the other. All things considered, and given the wide application of social norms, it is important to investigate the influence of transparency on it effectiveness.

2.3 Gender Differences in the Context of Risky Decisions

The context of our experiment involves risky choices (an abstract gamble experiment). Given the rich literature on the differences between men and women with respect to their risk attitudes, and in particular the finding that women are more risk averse than men, this paper also puts forward predictions and tests for heterogeneity effects. Therefore, this section presents the literature on the gender differences with respect to risk attitudes.

In a meta-analysis of 150 studies, Byrnes et al. (1999) found general differences in risk taking behavior between men and women. In particular, they found that in most domains, men are significantly more risk taking than women. Eckel and Grossman (2008, chapter 113) provide a literature survey on the gender differences with respect to risk aversion. This review indicates a stronger risk aversion among women in the context of abstract gamble tasks. Couple of example studies can be mentioned. Eckel and Grossman (2002) conducted an experiment to examine gender differences with respect to the level of risk aversion and loss aversion (in the context of investment choices in the lab). They found that women were more risk averse than men, and this difference was statistically significant. Such difference was not found with respect to loss aversion. In addition, the authors demonstrated that women were also *perceived* to be more risk averse than men, both by other women and by men. These results were confirmed in a later study by Eckel and Grossman (2008) where in a simple gamble-choice task women demonstrated significantly stronger risk aversion than men, irrespective of the frame (gain, loss or neutral). In a different study, by Levin *et al*. (1988), the authors examined the interaction effect between the framing of the gamble (as a gain or as a loss) and the gender of the decision maker. Participants received a set of gambles that varied in the stakes, probabilities and framing. Subsequently, they had to decide whether they would take the gamble or not. Men were found to be more favorable of the gambles than women, especially in the gain condition, and this difference was statistically significant.

Not many empirical studies on this topic try to provide theoretical explanation for this gender difference. Nevertheless, Eckel and Grossman (2002) suggest that evolutionary psychology might explain this difference. They argue that this difference might be derived from the different "returns to alternative investment in reproductive success" (p. 282). In other words, for women the dominant strategy might be to lower the risk for themselves and their offspring in order to achieve successful parenting. For male, on the other hand, competition for mating and taking more risks to acquire better resources, might lead to a better mate, therefore, increasing their return. This channel is tentative and was not examined empirically. In this paper we will not be testing for the channel for the gender differences, but given the prevalence of literature suggesting the existence of such differences in the specific context of gambles, we think gender is an important factor to take into consideration.

With respect to the influence of social norms, the perception, which prevailed in the past, was that women are more prone to be influenced by the actions or opinions of others. However, this perception was found to be wrong in a meta-analysis of experimental studies on the topic (Eagly 1978). Furthermore, a recent study by Croson *et al*. (2010) examined the joint effect of social norms and gender on donations (to public radio). They found that men were more responsive to the descriptive social norm than women were in their donation behavior. This finding strengthens the need to account for gender in our study.

2.4 Hypotheses

Based on the theoretical framework presented in the previous sections, this part puts forward hypotheses to be tested in the experiment.

Main effect hypotheses

H1. ***If participants are confronted with a social norm, the probability of choosing a riskier lottery but with a higher expected return will increase*** **[social norm effect]**

As have been discussed in section 2.2, descriptive norms tend to influence the choices of people by providing information on the actions of others. One of the channels responsible for this effect is a simple belief in the wisdom of the crowd.

H1a. ***A social norm statement will increase the probability of choosing a riskier lottery but with a higher expected return, to a larger extent for men than for women*** **[heterogeneity effect]**

As discussed in section 2.3, women seem to be more risk averse than men, especially in the context of abstract gambles. This might suggest that women have a stronger preference for less risky financial decisions (i.e. preferring an option with higher probability to win, but lower expected return). According to the nudge literature, those interventions are freedom preserving and thus should work only when the person has no strong preference to the contrary (Thaler & Sunstein 2008, pp. 5, 178; Sunstein 2017, p. 8). Given the stronger risk aversion of women, we can expect that they will follow the nudge (social norm) to a lesser extent than men will.

With respect to the influence of transparency, we rely on the psychological reactance theory as discussed in section 2.1, to form predictions. Since there is no theoretical ground for gender differences in psychological reactance levels, this part will be exploratory.

H2. ***If participants receive information on the way social norms work, the probability of choosing a riskier lottery but with a higher expected return will decrease*** **[transparency effect]**

The descriptive social norm is a form of social influence, which is expected to induce a certain level of pressure to follow the majority. The sense of threat to freedom of choice is expected to increase with the introduction of disclosure about the way this nudge works (through expectation of conformity) because it makes salient the intention of social influence. With respect to the moderating factors, several outcomes are possible. On the one hand, there is a clear expectation of freedom of choice (the task is to choose between two lotteries), and due to the simplicity of the task we assume participants will not feel incompetent to make the choice. Therefore, the sense of threat is not moderated. On the other hand, due to the simplicity of the task and the small stakes, participants most probably would not perceive the choice as an important decision for them, thus mitigating the sense of reactance.

H3. ***If participants receive information on the way social norms work and the purpose to use it, the probability of choosing a riskier lottery but with a higher expected return will increase*** **[full transparency effect]**

Providing participants with an explanation that the chosen social norm is meant to increase their individual welfare (encouraging to choose the lottery with a higher expected return), might mitigate the negative influence of the transparency. This additional information makes the goal of the nudge(r) explicit.

Psychological reactance hypotheses

To measure the psychological channel, which is potentially responsible for a transparency effect (if one is found), we also put forward hypotheses for the interaction between the psychological reactance measures and the different nudge and transparency treatments. State reactance investigates how the framing of the choice affects participants' perception of this choice (Dillard & Shen 2005). Trait reactance suggests that people, who are inherently more prone to reactance, will react stronger against the attempted influence when faced with perceived limits on their freedom of choice (Hong & Page 1989; Hong & Faedda 1996). Here as well, since there is no theoretical ground for gender differences in psychological reactance levels, this part will be exploratory.

H4a. ***If participants receive information on the way social norms work, experience of state reactance will be higher compared to when they do not receive such information*** **[state reactance]**

H4b. ***If participants receive information on the way social norms work and the purpose to use it, experience of state reactance will be lower compared to when they receive information on the way social norms work only, but higher as compared to when they receive no additional information*** **[state reactance]**

The social norm itself may evoke psychological reactance. However, if people are indeed not aware of their biases and the attempt to influence their behavior, such attempted influence is made more salient with the disclosure. Therefore, we expect the transparency (and full transparency) treatment to increase the reported sense of limited freedom.

H5. ***The higher is the participant on the trait reactance measure; the lower will be his probability to choose the option encouraged by the social norm when he receives information on the way social norms work only, or combined with purpose*** **[transparency trait reactance].**

Also with trait reactance we expect that if people were not fully aware of the social norm effect and the reasons it is used, disclosure would make the social influence more salient. Therefore, even though there might be differences in the choices of high and low trait reactance participants in the social norm treatment, those differences are expected to be exacerbated in the transparency treatments.

3. Experimental Design

3.1 The Experiment

In order to examine the influence of transparency on the effectiveness of social norms, we have designed a simple lottery choice between-subjects experiment.[4] Given that the general population is risk averse (on average), the experiment was designed in such a way that the lottery with the highest expected return was not the one which a risk-averse person would choose. Namely, lottery A was less risky, but the expected return was lower (2/3 chance to win £7, expected return = £4.7). Lottery B was riskier, but also offered a higher expected return (1/3 chance to win £20, expected return = £6.7). All participants have received instructions how the lotteries work and were asked to choose which of the two lotteries to play. After this choice, they have played their chosen lottery and received the results. In order to incentivize genuine choices, participants were informed that on top of the participation fee they all receive, each of them has 1/20 chance to receive payment of the actual lottery. For the presented lotteries payoffs, see Figure 1.

[4] The design is inspired by Billion and Desmet (2018).

Figure 1: The Lotteries Payoffs

	Blue	Yellow	Red
Lottery A	7	7	0
Lottery B	20	0	0

In order to encourage people to choose the lottery with the higher expected return, we have introduced a nudge in the form of a descriptive social norm. The social norm statement informed the participants that around 90% of participants in a similar previous experiment have chosen lottery B. The chosen social norm is clearly descriptive since it describes the behavior of others without any moral implications, and it was chosen over an injunctive social norm in light of some evidence of a stronger effect of descriptive norms (Hallsworth *et al*. 2017). In addition, we chose to express the social norm in percentages (90%) rather than in a fraction ("great majority") due to empirical evidence that the former yields a stronger effect than the latter (Hallsworth *et al*. 2017, p. 24).

In addition to the social norm manipulation, we have included two treatments with two different levels of transparency. The choice of the transparency form was meant to restore participants' autonomy in the sense of giving them "back" the control over evaluating and choosing between different options (see the explanation of how nudges restrict autonomy in e.g. Hausman & Welch 2010, p. 128). Our goal was to examine a meaningful type of transparency, the token interference transparency, that would provide participants with information about the intention behind the nudge, the fact they are nudged, and the means through which this nudge is believed to be effective (Hansen & Jespresen 2013, p. 17). However, we were also interested to examine whether the influence of such transparency (if exists) can be mitigated by explaining people the purpose of using the nudge and how it is meant to benefit them. The design of the experimental groups is presented in Table 1.

Table 1: Experimental Groups

Experimental Group	Provided information (Manipulation)	N	% Female
Control	*No information*	196	62%
Social Norm	*In a recent almost identical study,* ***around 90% of participants chose Lottery B*** *when given a choice between Lottery A and Lottery B. In other words, most people preferred a one in three chance of earning £20 to a two in three chance of earning £7.*	188	62%
Social Norm + Transparency	*In a recent almost identical study,* ***around 90% of participants chose Lottery B*** *when given a choice between Lottery A and Lottery B. In other words, most people preferred a one in three chance of earning £20 to a two in three chance of earning £7.* ***Please note, the reason you are presented with the information about the choice of majority of participants in a similar study is to influence your decision. The choice of presenting this information follows evidence from behavioural studies that demonstrate people are strongly influenced by the actions and beliefs of other people.***	182	59%
Social Norm + Transparency + purpose	*In a recent almost identical study,* ***around 90% of participants chose Lottery B*** *when given a choice between Lottery A and Lottery B. In other words, most people preferred a one in three chance of earning £20 to a two in three chance of earning £7.* ***Please note, the reason you are presented with the information about the choice of majority of participants in a similar study is to influence your decision. The choice of presenting this information follows evidence from behavioural studies that demonstrate people are strongly influenced by the actions and beliefs of other people. The goal of providing you with this information is to help you to make the best monetary decision (i.e. choose the lottery with the higher expected return).***	182	61.5%

Note: the respective N excludes participants who failed the attention check or responded wrong to both control questions. Therefore, it includes only participants who were part of the analysis.

In this experiment, we sought to examine the influence on a lottery choice of: a) a descriptive social norm as such; b) a descriptive social norm combined with an explanation of the intention and the channel through which social norms affect people's behavior; and c) a descriptive social norm combined with an explanation of the intention, the channel through which social norms affect people's behavior, and the purpose to use this nudge. In addition, we examined whether the effects differ depending on the gender of the participant. Our dependent variable is the probability to choose lottery B. The baseline for the social norm effect is the control group, and the baseline for the transparency effect is the social norm group.

After the lottery task, participants were requested to answer questions measuring state and trait reactance. With state reactance, participants had to report on a 5-point Likert scale to which extent they felt that the presented social norm: (1) threatened their freedom to choose; (2) tried to make a decision for them; (3) tried to manipulate them; (4) tried to pressure them. Furthermore, they were asked to indicate how irritated they were with regard to the given social norm statement. For trait reactance, participants were requested to state the level of their agreement on a 5-point Likert scale with 14 statements (e.g. "*Regulations trigger a sense of resistance in me*").[5] For the full psychological reactance questionnaire, which was presented to the participants, see Appendix 1. This part was meant to identify the psychological mechanism that might induce the transparency problem.

3.2 Procedure

Participants (N = 936) were recruited via the online platform Prolific Academic and the experiment was programmed and data collected via a software called Qualtrics. An online platform gives the advantage of recruiting more demographically diverse participants, which potentially increases the generalizability of the results as compared to the classical laboratory sample of participants. Yet the quality of data is comparable to university laboratory experiments (Buhrmester et al. 2011, Peer at al. 2017). Therefore, the use of online platforms to conduct experiments is becoming more common.

The following restrictions were set for choosing the participants: 1) UK nationality; 2) fluent English; 3) employed (full or part time); 4) participants with more than 95% completion rate.[6] Only participants, who gave their explicit consent to participate in the study, could continue to next screens. All participants received identical instructions and two test questions to measure their attentiveness.

Participates were randomly allocated to the experimental groups. In the treatment groups, immediately after making the choice of the lottery, participants were presented with a simple manipulation check question (on a separate screen). The question asked about the social norm information ("*In the previous screen you have been provided with information on the percentage of people who chose Lottery B in a recent study. What was this percentage? 75% / 90% / 95%*"). Only participants who answered this question correctly could continue with the study. The manipulation check was meant to guarantee participants paid attention to the presented manipulation. After excluding people who failed the manipulation check and the attention checks, the remaining sample, and the one used for the analysis, was N=748.

Following the presentation of the lottery results, the participants were requested to fill out a questionnaire measuring state (only treatment groups) and trait reactance, and provide basic

[5] State reactance questions we based on elements from Dillard & Shen (2005) to measure the experience of restricted freedom evoked by the attempted social influence. The trait reactance sentences are based on Hong and Page (1989) and Hong and Faedda (1996).

[6] Given the instructions were in English, we chose an English speaking sample. The unemployment rate in the UK is low (4%), and given the possibility the unemployed participants might be overrepresented in prolific we tried to reduce the chance of a biased result. Furthermore, with the employment condition we tried to reduce the number of "career" participants. We set a high completion rate to increase the probability of "good quality" participants whose submissions are not rejected by other researchers due to lack of attention.

demographic information. Each participant who completed the study received £1 for participation, and one out of every 20 participants (randomly selected) was paid according to the results of his/her chosen lottery (0, £7 or £20).

4. Results

We discuss the results in two parts, whereas the first part deals with the main effect hypotheses (H1-3), the second part zooms in on the psychological reactance hypotheses (H4-5). Before that we however briefly describe subjects' characteristics.

Subjects' characteristics

Our sample totals 748 participants, i.e. 80% of participants passing the attention checks (initial sample had 936 responses). Of these 748 participants, 458 are females and 290 are males. The average age in the sample equals 34.8 years. We also observe that 59% of the respondents were highly educated. These characteristics are quite balanced across the experimental groups. The only exception is the underrepresentation of highly educated people in the control group. To account for this unbalanced distribution of education level, we control for it in the multiple regression models. For further details, see Table 2.

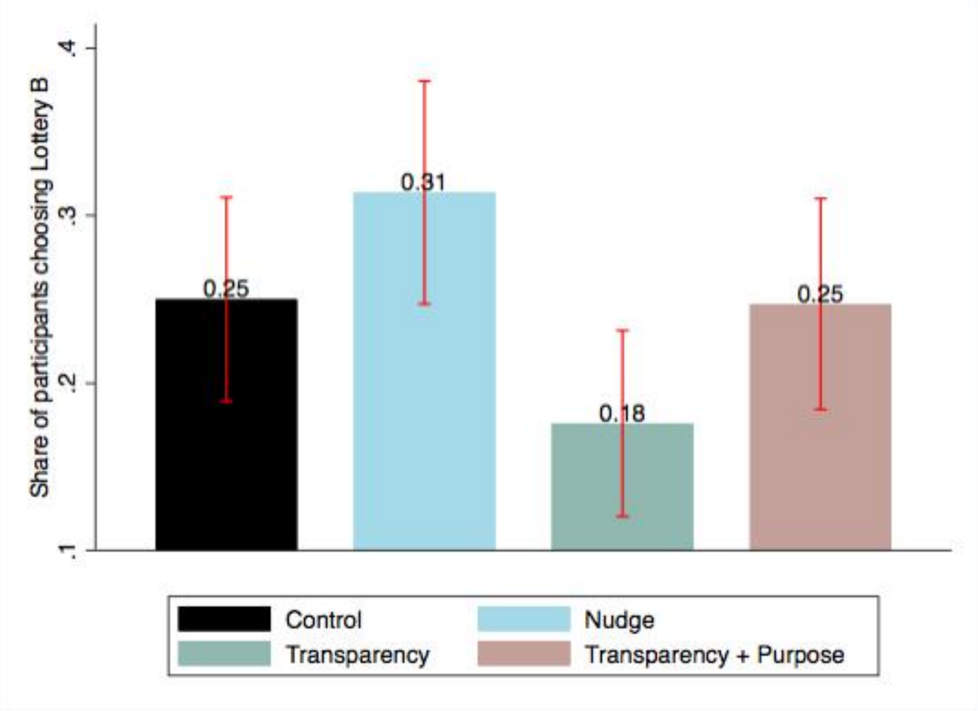
Table 2. Participants Characteristics per Experimental Group

Group	% Female	Average Age	% Highly educated
Control	61.9%	34.5	51.8%
Social Norm	62.4%	34.6	60.8%
Transparency	59.4%	35.5	62.3%
Transparency + purpose	62.3%	34.4	61.5%

Main effects

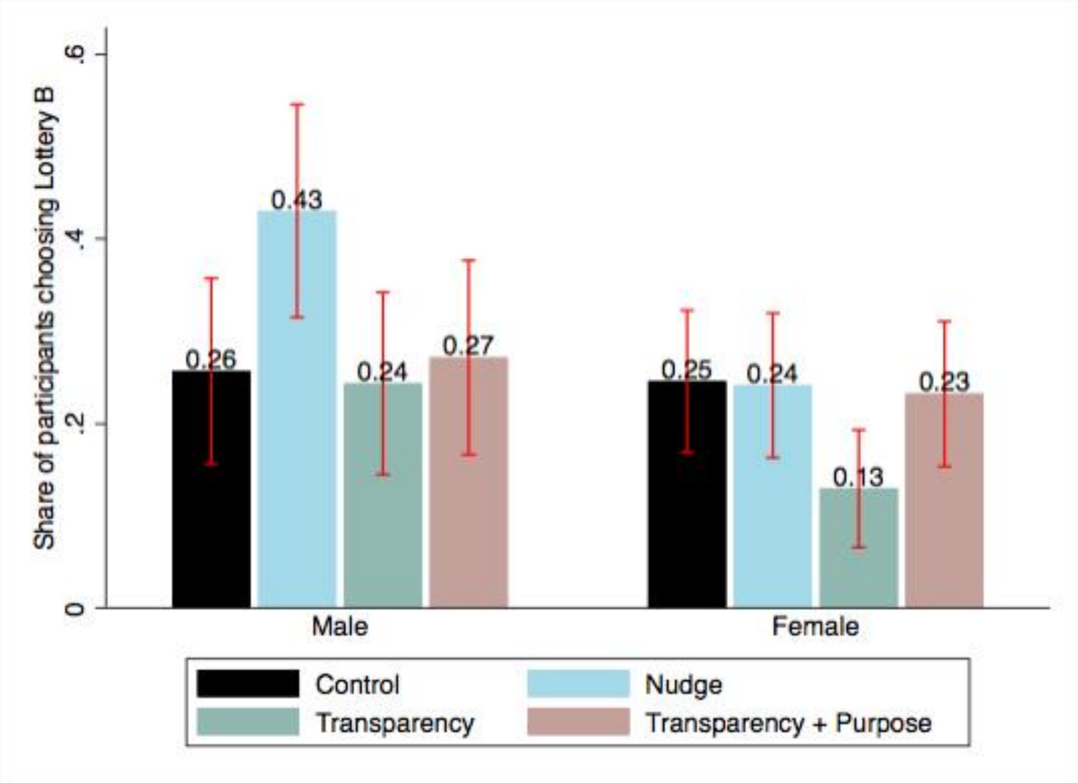
First and foremost, we are interested in testing the overall effect of a social norm nudge (H1) and comparing treatment effects between men and women (H1a). In Figure 2 it can be seen that 25% of participants chose lottery B (riskier but higher payoff choice) in the control group and 31% of participants chose the same lottery in the (social norm) nudge treatment. This difference is nonetheless not statistically significant ($\chi^2(1)=1.934$, p=0.164).

Figure 2. Differences between the Control and Treatment Groups for the Entire Sample



Turning to the variation between treatment effects, we are able to observe striking differences between treatment effects for women and men (see Figure 3). While women have roughly the same propensity to choose lottery B in the control and the nudge group (25% and 24% respectively), men are much more likely to choose lottery B under the (social norm) nudge condition. 26% of male participants chose lottery B under the control condition and 43% in the nudge treatment group. This difference is statistically significant ($\chi^2(1)=4.895$, p=0.027). This therefore provides evidence in favor of H1a, which stipulates that given the tendency of women to be more risk averse, the social norm should be more effective for men as compared to women. This is consistent with the nudge literature, which suggests the behavioral intervention is effective only when the target of the nudge does not have a strong preference to the opposite (the freedom-preserving element of the nudge). To further substantiate these results we run a set of logistic regressions, where our dependent variable is a binary variable coded with 1 in case the participant chose lottery B and 0 otherwise. We regress this variable on the nudge treatment (Table 3, model 1) and the nudge treatment along with gender variable and their interactions (Table 3, model 2). We observe similar results to the non-parametric test.

Figure 3. Differences between the Control and Treatment Groups across Females and Males



After showing the heterogeneous treatment effects, we now turn to testing the effects of transparency (H2) and the effects of transparency combined with purpose (H3). Since we were able to discover the variation in treatment effects for women and men, besides comparing the choices between the nudge and transparency conditions for the whole sample, we also present the results for women and men separately. We start the analysis by first looking at the differences graphically. Figure 3 shows there are indeed differences between the nudge and transparency conditions. Transparency alone seems to decrease the "effectiveness" of the nudge. While in the nudge condition 31% of participants decided to choose lottery B, in the transparency condition this share went drastically down to 18%. The χ^2 test confirms that this difference is statistically significant ($\chi^2(1)=9.497$, p=0.002). See Table 3 Model 3, which further substantiate these results via logistic regression. These disparities between the nudge and the simple transparency conditions can also be observed for women ($\chi^2(1)=4.585$, p=0.032) and men ($\chi^2(1)=5.742$, p=0.017), respectively. The results from the logistic regression (Table 2 Model 4) point out into the same direction. Overall, this provides evidence in favor of H2, which predicts that transparency will hinder the "effectiveness" of the social norm nudge, the so-called transparency effect.

One should also note the differences between the control groups and the simple transparency (Figures 2 and 3). For the entire group and for men, this difference is not statistically significant ($\chi^2(1)=3.084$, p=0.079 and $\chi^2(1)=0.036$, p=0.849 respectively). This means, simple disclosure of the social norm, brought the probability of choosing Lottery B back to the level of the control group (for men). Moreover, for female, simple disclosure led in fact to a boomerang effect since the probably to choose lottery B is now lower than in the control group, and this difference is statistically significant (χ^2=5.006, p=0.025).

This result seems to be consistent with the psychological reactance theory. Pure disclosure of the intent to use a social norm and the way it operates increases the salience of the attempted social influence. This in turn, might have increased the sense of restricted freedom, and thus evoked reactance against the nudge. Reducing the probability to choose lottery B may be viewed in light of the psychological reactance theory as participants' attempt to restore their freedom.

The question remains whether revealing the purpose of the nudge will reinstate some of its effectiveness (H3). To this end, we compare the distribution of lottery choices across the nudge and full transparency condition. In Figure 2, it is noticeable that the explanation of purpose restores propensity of choosing lottery B. However, for the entire sample, these differences in the nudge and full transparency groups are not statistically significant (χ^2=2.029, p=0.154). We again observe heterogeneous effects between women and men. While for women the full transparency restores the likelihood of choosing lottery B to its initial level (compare 24% and 23% of participants choosing lottery B in the nudge and full transparency groups, respectively), for men this effect does not occur (the nudge effect is not restored). The difference between the nudge (43% chose lottery B) and the full transparency (27% chose lottery B) condition is statically significant (χ^2=3.940, p=0.047). At the same time, the difference between control and full transparency group for men is not statistically significant (χ^2=0.040, p=0.842).

These results can be likewise observed in Table 3 Model 4. Notice that the coefficients on the "Trans+Purp" variable and the interaction term "Trans+Purp#Female" also entirely cancel out. Given that the social norm nudge worked only for men, it seems that there is not much evidence in favor of H3 (full transparency effect). As can be seen in Model 4, full transparency also reduces the propensity of men to choose lottery B (10% statistical significance) and the difference between this propensity in the control group and in the full transparency group for men is not statistically significant. This finding suggests that even adding the purpose behind the nudge did not restore (male) participants' sense of freedom to the extent they would follow the social norm. Model 5 and model 6 in Table 3 demonstrate the difference between the control group and all experimental conditions (model 5) along with heterogeneity effects for women and men (model 6). It is clear from the latter that the nudge effect can be identified for men (see the positive and statistically significant coefficient next to the "Social Norm" condition), but not for females (see the negative and statistically significant coefficient next to the interaction term "SN#Female"). Model 7 controls for age and finds no differences.

Table 3. Main Effects: Logistic Regressions

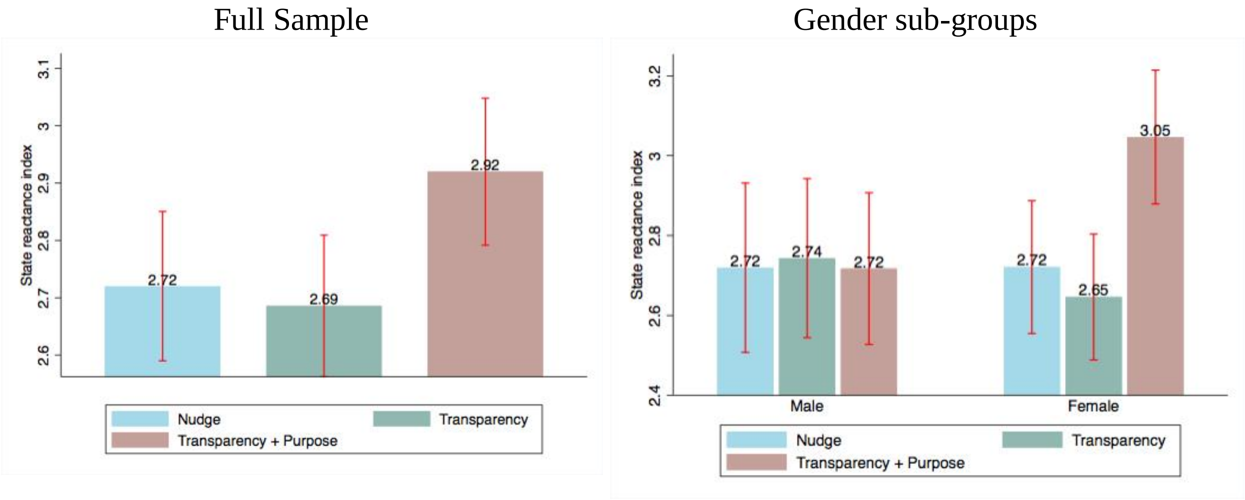
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Control v. SN	Control v. SN	SN v. Transparency	SN v. Transparency	Control v. all treatments	Control v. all treatments	Control v. all treatments
	Full sample	Gender interaction	Full sample	Gender interaction	Full sample	Gender interaction	Full set of controls
Social norm (SN)	0.296	0.769*			0.303	0.773*	0.779*
	(1.29)	(2.15)			(1.33)	(2.16)	(2.18)
Trans			-0.766**	-0.847*	-0.462^{+}	-0.0745	-0.0749
			(-3.06)	(-2.35)	(-1.81)	(-0.20)	(-0.20)
Trans+Purp			-0.333	-0.700^{+}	-0.0296	0.0727	0.0724
			(-1.43)	(-1.95)	(-0.12)	(0.19)	(0.19)
Female		-0.0447		-0.862**		-0.0486	-0.0532
		(-0.13)		(-2.67)		(-0.14)	(-0.16)
SN#Female		-0.816^{+}				-0.813^{+}	-0.824^{+}
		(-1.74)				(-1.74)	(-1.76)
Trans#Female				0.0764		-0.736	-0.727
				(0.15)		(-1.41)	(-1.39)
Trans+Purp#Female				0.640		-0.172	-0.174
				(1.34)		(-0.35)	(-0.36)
High_Educ	0.253	0.243	0.155	0.172	0.157	0.169	0.172
	(1.09)	(1.03)	(0.75)	(0.82)	(0.90)	(0.96)	(0.97)
Age							-0.00602
							(-0.73)
_cons	- 1.234***	-1.201***	-0.877***	-0.388	-1.182***	-1.158***	-0.950*
	(-5.90)	(-4.01)	(-4.32)	(-1.42)	(-6.21)	(-4.06)	(-2.37)
N	384	384	552	552	748	748	748
Pseudo R^2	0.007	0.023	0.017	0.035	0.012	0.026	0.027

Note: *z* statistics in parentheses. Statistical significance: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Psychological reactance

This section serves to test for potential transmission channels of transparency. Particularly, we are interested to verify to what extent the receipt of information on the way the social norm works translates into a greater level of psychological reactance. To this end, we first test the differences in experience of *state* reactance between the social norm group (no transparency) and the transparency group (H4a). We further examine whether explaining the purpose of using the nudge alleviates the experience of state reactance (H4b). We test these hypotheses with OLS regression as the experience of state reactance is a (continuous) indicator obtained by calculating the average of five scores employed in measuring state reactance (see Appendix 1). For graphical illustration of results, see Figure 4.

Figure 4: State Reactance



The results presented in Table 4 do not provide evidence for neither of the hypotheses. In Model 1 (the model without gender heterogeneity) the highest state reactance score was observed in the group with full transparency. The difference between this group and the nudge group was statistically significant at 0.05 level. However, this difference is entirely driven by the female participants. The reported experience of state reactance in the transparency group is statistically indistinguishable from the state reactance in the nudge condition. For male participants, who are now our benchmark group since only they were affected by the social norm, none of the transparency conditions evoked reactance (as reported by them).[7]

[7] Surprisingly, women's reported experience of reactance is contrary to their choices. Highest reported experienced reactance in the full transparency treatment would predict the less frequent choice of lottery B. Yet the lowest frequency of choosing lottery B is in the simple transparency treatment. The frequency of choosing lottery B is then the same in the full transparency, the social norm and the control experimental groups (female participants).

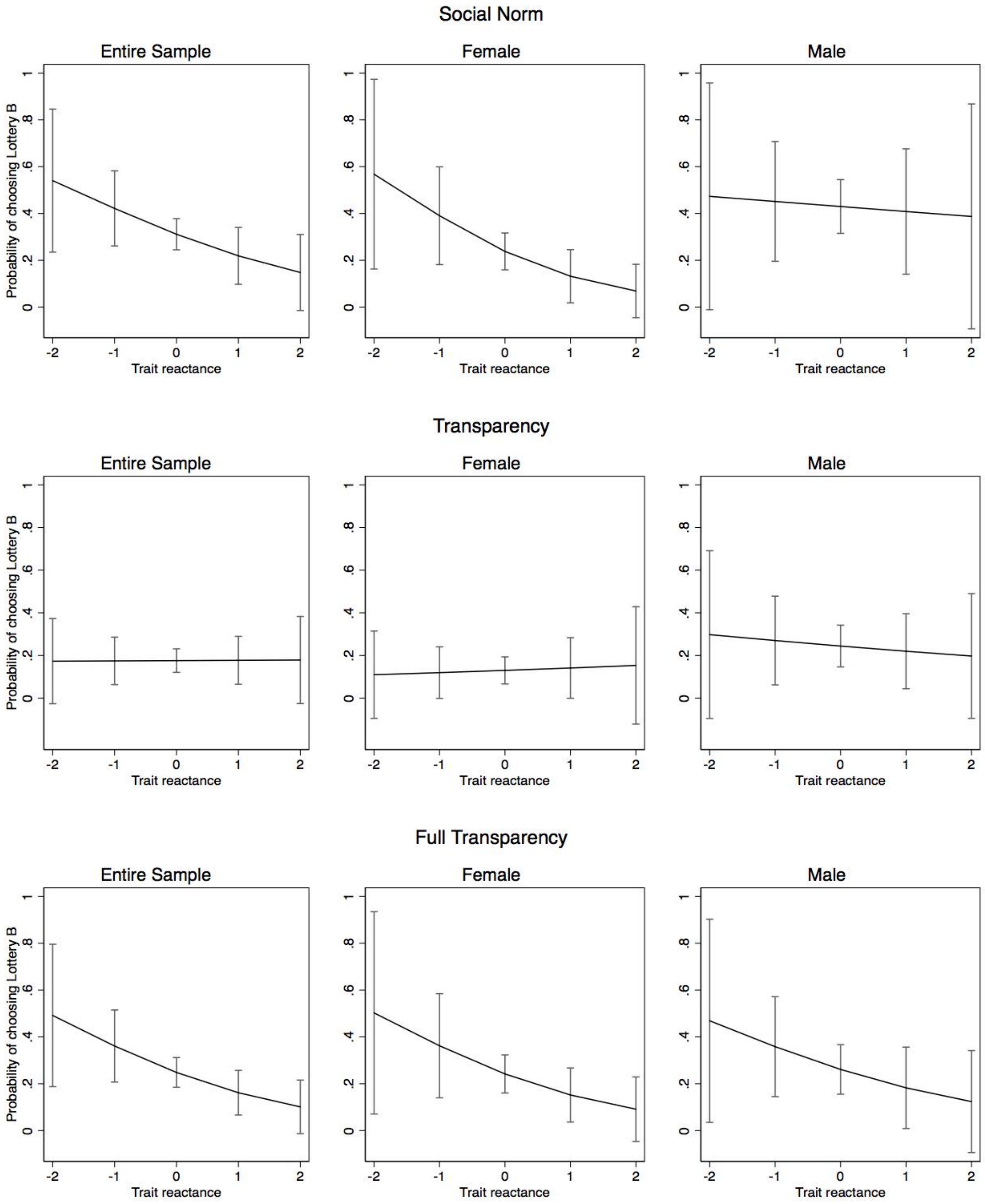
Table 4. State reactance OLS regressions

	(1) SN v. Transparency treatments Full sample	(2) SN v. Transparency treatments Gender interaction
Trans	-0.0371	0.0338
	(-0.41)	(0.23)
Trans+Purp	0.198^{*}	0.00698
	(2.17)	(0.05)
Female		0.00648
		(0.05)
Trans#Female		-0.119
		(-0.64)
Trans+Purp#Female		0.310^{+}
		(1.66)
High_Educ	0.177^{*}	0.173^{*}
	(2.31)	(2.26)
_cons	2.613^{***}	2.611^{***}
	(33.05)	(22.99)
N	552	552
R^2	0.023	0.034

Notes**:** *t* statistics in parentheses.
Statistical significance: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We further test for the interactive effect of *trait* reactance; i.e. we verify whether under the conditions where participants receive information on the way social norms work and the way social norms work combined with their purpose (H5), the social norm effect on participants with high trait reactance will be lower than on participants with lower trait reactance. To this end, we run a set of logistic regressions for the overall sample and subsamples of women and man. In each regression we control for treatments, trait reactance and their interactions. The trait reactance variable was obtained by calculating the average of 14 scores employed in measuring trait reactance (see Appendix 1). In the regression model we use the trait reactance variable centered at its mean to make the interpretation of the results meaningful. Even though we did not hypothesize about trait reactance in the social norm group, we are including it in the figure to demonstrate better the distinctive effect of transparency. For graphical illustration of the results see Figure 5. This figure presents three panels to demonstrate the relationship between reported trait reactance scores and the outcome variable in each of the treatments, for the entire sample, for male and for female participants.

Figure 5: Trait reactance



As illustrated in Table 5, we do not find evidence for trait reactance hypothesis (H5). For the entire sample (Model 1) and for men (Models 3), who are our benchmark, we do not see the negative relationship between the reported trait reactance and the outcome variable (probability of choosing lottery B), in none of the treatment groups. Although it was not stated as a hypothesis, we see weak evidence (10% level of statistical significance) for the negative relationship between reported trait reactance and the outcome variable in the nudge treatment for the female sample. However, given the fact we did not find the nudge effect for women, this is not our benchmark group for investigating the psychological reactance.

Table 5. Trait Reactance: Logistic Regressions

	(1) Full Sample	(2) Female	(3) Male
Trans	-0.754^{**}	-0.755^{*}	-0.840^{*}
	(-3.00)	(-2.08)	(-2.32)
Trans+Purp	-0.316	0.00955	-0.750^{*}
	(-1.34)	(0.03)	(-2.04)
Trait	-0.479	-0.722^{+}	-0.0879
	(-1.54)	(-1.70)	(-0.18)
Trans#Trait	0.488	0.818	-0.0484
	(1.05)	(1.22)	(-0.07)
Trans+Purp#Trait	-0.0594	0.143	-0.371
	(-0.14)	(0.24)	(-0.55)
High Education	0.192	0.271	0.141
	(0.92)	(0.92)	(0.46)
_cons	-0.912^{***}	-1.331^{***}	-0.372
	(-4.44)	(-4.62)	(-1.21)
N	552	336	216
Pseudo R^2	0.026	0.033	0.030

Note: *z* statistics in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5. Discussion

In this paper, we have dealt with the important question of ethical use of regulatory instruments. In particular, we have investigated for the first time the transparency problem with respect to a social norm nudge. Our findings demonstrate two main outcomes. First, the social norm in the context of a lottery choice was found to be effective only for the male participants. This result is in line with the literature on the larger risk aversion of women, and the freedom-preserving nature of nudges (i.e. nudges should only be effective when there are no strong preferences to the contrary). Therefore, it raises an important policy question – should nudges be adjusted to (sub)groups to be more effective? The heterogeneity effect in our study is consistent with other studies finding differences in the levels of effectiveness of different nudges on sub-groups (Johnson *et al*. 2012, pp. 496-7). Bronchetti *et al*. (2013), for instance, found that defaults encouraging savings were not effective for low-income people. The potential explanation is the strong preference those people hold to use the amount of money, which was the target of the nudge. Beshears *et al*. (2015) found a boomerang effect driven by low-income employees when nudged with a descriptive social norm to save more. Finally, Gerber and Rogers (2009) demonstrated that high turnover descriptive social norms affected only infrequent voters.

Second, our study demonstrates that meaningful transparency can indeed inhibit the effectiveness of a social norm nudge. Not only that both types of transparency, with purpose, and without, reduced the desired choice for male participants, simple transparency also reduced such behavior for female participants as compared to the control group. This is initial evidence and more studies should be conducted before providing conclusive policy recommendations with respect to social norms and transparency. However, the main significance of these results is that they cast doubt on the ability to generalize the findings from defaults and their lack of susceptibility to the influence of transparency. Therefore, our findings stress the importance of further investigating the problem of transparency with respect to different nudges.

The general importance of investigating the transparency problem with respect to the used choice architecture by governments is clear. Nudges are advocated as freedom-preserving interventions. Therefore, the behavioral reaction of the target individuals to meaningful transparency of such interventions is a good test whether they indeed do not serve as manipulative instruments. In order to maintain their legitimacy, governments should adopt the following general rule. Nudges, which maintain their effectiveness even when made transparent, can be used. If people keep following the nudge even when they know it is employed and how it works, it signals their lack of objection to the specific choice. On the other hand, nudges which lose their effectiveness with different types of meaningful transparency become an illegitimate (if not transparent) or ineffective (with transparency) instrument of governmental intervention. However, to give clear policy recommendations with respect to each type of nudges, a thorough and comprehensive research must be conducted.

We did not find evidence for the psychological reactance as the mechanism behind the influence of transparency. This is consistent with Bruns *et al*. (2018) who also measured psychological reactance and did not find any evidence (even though there transparency did not influence the default's effectiveness). The existence of the transparency influence (for men) on the one hand, and its inconsistency with the reported experience of psychological reactance on the other hand, might suggest that additional psychological channels for the transparency effect should be examined. This leads to the conclusion a more comprehensive and specific theory should be constructed for the transparency problem.

This study is the first step in a broader project that will construct a comprehensive theory for nudges and the problem of transparency. Such theory needs to include different elements of human psychology, and especially aspects affecting person's self-perception. By modeling a more comprehensive theory, the different psychological mechanisms underlying the effectiveness of different nudges can be accounted. Consequently, one would be able to predict which nudges are more probable to be influenced by transparency. These predictions will be then tested. Finally, a step further will be taken to examine whether disclosure can be framed in a meaningful and yet less harmful way to increase the legitimacy of covert nudges. However, if people will still reject transparent nudges, policy makers should reconsider the use of such nudges as part of their regulatory toolkit.

References

Alemanno Alberto, Sibony Anne-Lise (eds.) (2015) *Nudge and the Law: A European Perspective*. Hart Publishing, Oxford and Portland.

Allcott Hunt (2011) Social Norms And Energy Conservation. *Journal of Public Economics* 95, 1082–1095.

Asch Solomon E. (1956) Studies of Independence and Conformity: I. a Minority of One against a Unanimous Majority. *Psychological Monographs: General and Applied* 70(9), 1-70.

Beshears John, Choi James, Laibson David, Madrian Brigitte, Milkman Katherine (2015) The Effect of Providing Peer Information on Retirement Savings Decisions. *The Journal Of Finance* LXX(3), 1161-1201.

Billion Stephen, Desmet Pieter (2018) Nudging to Regret: How Defaults And Peer Information Affect Anticipated Regret. Working Paper.

Bott Kristina Maria, Cappelen Alexander W., Sorensen Erik, Tungodden Bertil (2017) You've Got Mail: A Randomised Field Experiment on Tax Evasion. Doi: http://dx.doi.org/10.2139/ssrn.3033775.

Bovens Luc (2009) The Ethics Of Nudge. In: Grüne-Yanoff T, Hansson SO (eds) *Preference Change: Approaches From Philosophy, Economics And Psychology*, pp. 207–219. Springer Science & Business Media, Dordrecht, The Netherlands.

Bronchetti Erin Todd, Dee Thomas S., Huffman David B., Magenheim Ellen (2013) When A Nudge Isn’t Enough: Defaults And Saving Among Low-Income Tax Filers. *National Tax Journal* 66(3), 609–634.

Brehm Jack Williams (1966) Theory of Psychological Reactance. *Academic Press Inc*, New York, NY.

Brehm Jack Williams, Brehm Sharon (1981) *Psychological Reactance: A Theory of Freedom and Control*. Academic Press, New York, NY.

Bruns Hendrik, Kantorowicz-Reznichenko Elena, Klementd Katharina, Jonsson Marijane Luistro, Rahali Bilel (2018) Can nudges be transparent and yet effective? *Journal of Economic Psychology* 65, 41–59.

Buhrmester Michael, Kwang Tracy, Gosling Samuel D. (2011) Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6(1), 3-5.

Byrnes James P, Miller David C, Schafer William D (1999) Gender Differences In Risk Taking: A Meta-Analysis. *Psychological Bulletin* 125(3), 367-383.

Cialdini Robert B, Reno Raymond R, Kallgren Carl A (1990) A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. *Journal of Personality and Social Psychology* 58(6), 1015-1026.

Clee Mona A, Wicklund Robert A (1980) Consumer Behavior and Psychological Reactance. *Journal of Consumer Research* 6(4), 389-405.

Croson RT, Handy F, Shang J (2010) Gendered Giving: The Influence Of Social Norms On The Donation Behavior Of Men And Women. *International Journal of Nonprofit and Voluntary Sector Marketing* 15(2), 199-213.

Dinner Isaac, Goldstein Daniel G., Johnson Eric J., Liu Kaiya (2011) Partitioning Default Effects: Why People Choose Not to Choose. *Journal of Experimental Psychology: Applied* 17(4), 332–341.

Eagly Alice H. (1978) Sex Differences in Influenceability. *Psychological Bulletin* 85(1), 86-116.

Eckel Catherine C., Grossman Philip J. (2002) Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk. *Evolution and Human Behavior* 23, 281–295.

Eckel Catherine C, Grossman Philip J (2008) Chapter 113: Men, Women And Risk Aversion: Experimental Evidence. In: Plott Charles, Smith Vernon (eds) *Handbook of Experimental Economics Results* vol. 1, Part 7, pp. 1061-1073. Elsevier.

Frey Bruno S, Meier Stephan (2004) Social Comparisons and Pro-Social Behavior: Testing "Conditional Cooperation" in a Field Experiment. *American Economic Review* 94(5), 1717-1722.

Glaeser, Edward (2006) Paternalism and Policy. *University of Chicago Law Review* 73: 133-56.

Gerber Alan S, Rogers Todd (2009) Descriptive Social Norms and Motivation to Vote: Everybody's Voting and so Should You. *Journal of Politics* 71(1), 178-191.

Hallsworth Michael, List John A, Metcalfe Robert D, Vlaev Ivo (2017) The Behavioralist As Tax Collector: Using Natural Field Experiments To Enhance Tax Compliance. *Journal of Public Economics* 148, 14–31.

Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the Manipulation Of Choice: A Framework For The Responsible Use Of The Nudge Approach To Behaviour Change In Public Policy. *European Journal of Risk Regulation*, 4, 3–28.

Hausman Daniel M, Welch Brynn (2010) Debate: To Nudge or Not to Nudge. *The Journal of Political Philosophy* 18(1), 123–136.

Hong Sung-Mook, Page Sandra (1989) A Psychological Reactance Scale: Development, Factor Structure And Reliability. *Psychological Reports* 64, 1323-1326.

House of Lords, Science and Technology Select Committee (2011) *Behaviour Change*. London, UK.

Johnson Eric J, Goldstein Daniel (2003) Do Defaults Save Lives? *Science* 302(5649), 1338-1339.

Johnson Eric J, Shu Suzanne B, Dellaert Benedict G.C, Fox Craig, Goldstein Daniel G, Häubl Gerald, Larrick Richard P, Payne John W, Peters Ellen, Schkade David, Wansink Brian, Weber Elke U (2012) Beyond Nudges: Tools of A Choice Architecture. *Marketing Letters* 23, 487–504.

Kroese Floor M., Marchiori David R., de Ridder Denise T. D. (2016) Nudging Healthy Food Choices: A Field Experiment At The Train Station. *Journal of Public Health* 38(2), e133–e137.

Loewenstein George, Bryce Cindy, Hagmann David, Rajpal Sachin (2015) Warning: You are About To Be Nudged. *Behavioral Science & Policy* 1(1), 35-42.

Lunn Pete (2014) *Regulatory Policy and Behavioural Economics*. OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264207851-en.

Marchiori David R, Adriaanse Marieke A, De Ridder Denise TD (2017) Unresolved Questions In Nudging Research: Putting The Psychology Back In Nudging. *Social and Personality Psychology Compass* doi: 10.1111/spc3.12297.

Reisch, LA, Sunstein CR (2016) Do Europeans Like Nudges? *Judgement and Decision Making* 11(4), 310–25.

OECD (2017) *Behavioural Insights and Public Policy: Lessons from Around the World*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264270480-en.

Osman Magda, Fenton Norman, Pilditch Toby, Lagnado David, Neil Martin (2018) Whom Do We Trust on Social Policy Interventions? *Basic and Applied Social Psychology* 40(5), 249-268.

Paunov Y, Wänke M, Vogel T (2018) Transparency Effects On Policy Compliance: Disclosing How Defaults Work Can Enhance Their Effectiveness. *Behavioural Public Policy* doi:10.1017/bpp.2018.40.

Peer Eyal, Brandimarte Laura, Samat Sonam, Acquisti Alessandro (2017) Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research. *Journal of Experimental Social Psychology* 70, 153–163.

Rebonato Riccardo (2012) *Taking Liberties: A Critical Examination of Libertarian Paternalism*. Palgrave Macmillan, New York, NY.

Richter Isabel, Thøgersen John, Klöckner Christian A (2018) A Social Norms Intervention Going Wrong: Boomerang Effects from Descriptive Norms Information. *Sustainability* 10, 2848 doi:10.3390/su10082848.

Sherif Muzafer (1963) *The Psychology Of Social Norms*. Harper & Brothers Publishers, New York, NY.

Silva Antonio, John Peter (2017) Social Norms Don't Always Work: An Experiment To Encourage More Efficient Fees Collection For Students. *PLoS One* 12(5) https://doi.org/10.1371/journal. pone.0177354.

Schnellenbach Jan (2012) Nudges and norms: On the political economy of soft paternalism. *European Journal of Political Economy* 28, 266–277.

Sousa Joana, Ciriolo Lourenço Emanuele, Almeida Sara Rafael, Troussard Xavier (2016) *Behavioural Insights Applied to Policy: European Report 2016*. Joint Research Centre, the European Commission, Brussels, Belgium. [European Report, 2016].

Steffel M, Williams EF, Pogacar R (2016) Ethically Deployed Defaults: Transparency and Consumer Protection through Disclosure and Preference Articulation. *Journal of Marketing Research* 53(5), 865–880.

Sunstein CR (2013) *Simpler: The Future of Government*. Simon & Schuster, New York, NY.

Sunstein Cass (2017) Nudges that Fail. *Behavioural Public Policy* 1(1), 4-25 doi:10.1017/bpp.2016.3.

Thaler Richard H, Cass R. Sunstein (2008) *Nudge: Improving Decisions About Health, Wealth, And Happiness*. Penguin Books, London, England.

Wetenschappelijke Raad voor het Regeringsbeleid (WRR) (2014) Met Kennis van Gedrag Beleid Maken, Rapportnummer 92. Amsterdam University Press, The Hague, Amsterdam. [The Dutch Report].

Wicker Allan W (1969) Attitudes versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects. *Journal of Social Issues* 25, 41-78.

Wilkinson TM (2013) Nudging and Manipulation. *Political Studies* 61, 341–355.

Appendix 1: Psychological Reactance Questionnaire

State reactance

Please indicate to what extent do you agree with each of the following statements on a 5-point response scale (where 1 means "strongly disagree", and 5 means "strongly agree").

- The social norm statement (the choice of 90% of participants in another study) threatened my freedom to choose
- The social norm statement (the choice of 90% of participants in another study) tried to make a decision for me
- The social norm statement (the choice of 90% of participants in another study) tried to manipulate me
- The social norm statement (the choice of 90% of participants in another study) tried to pressure me

Please indicate how irritated you were with regard to the given social norm statement.
Scale on slide from “Not irritated at all “ to “Very irritated”

Trait reactance

Please indicate to what extent do you agree with each of the following statements on a 5-point response scale (where 1 = "strongly disagree", and 5 = "strongly agree").

1. Regulations trigger a sense of resistance in me.
2. I find contradicting others stimulating.
3. When something is prohibited, I usually think, ‘‘that’s exactly what I am going to do’’.
4. The thought of being dependent on others aggravates me.
5. I consider advice from others to be an intrusion.
6. I become frustrated when I am unable to make free and independent decisions.
7. It irritates me when someone points out things, which are obvious to me.
8. I become angry when my freedom of choice is restricted.
9. Advice and recommendations usually induce me to do just the opposite.
10. I am content only when I am acting of my own free will.
11. I resist the attempts of others to influence me.
12. It makes me angry when another person is held up as a role model for me to follow.
13. When someone forces me to do something, I feel like doing the opposite.
14. It disappoints me to see others submitting to standards and rules.