To follow or not to follow the herd? Transparency and social norm nudges

Elena Kantorowicz-Reznichenko1,2 | Jaroslaw Kantorowicz3

1Rotterdam Institute for Law and Economics, Erasmus School of Law, Erasmus University Rotterdam, Rotterdam, The Netherlands
2Institute for Research in Economic and Fiscal issues (IREF) Fellow, France
3Institute of Security and Global Affairs and Department of Economics, Leiden University, Leiden, The Netherlands

Abstract

Behavioral interventions in general, and nudges in particular have become in recent years a popular (soft) regulatory instrument all around the world. Despite the excitement around this policy-relevant field, some concerns have been raised. Nudges utilize behavioral biases in order to direct an individual’s behavior. People, however, are usually not aware of the fact that such biases are used to influence their behavior. Making nudges transparent is important in democratic societies; yet, this might inhibit their effectiveness. Whether transparency inhibits the effect of a nudge was examined with respect to default nudges. However, this is the first paper to examine the effectiveness of transparent social norm nudges. Using an online experiment, we find that unlike with defaults, where transparency seems not to have inhibitive effects, disclosing the way social norms work and the purpose of using them diminishes the positive social norm effect. By means of heterogeneity analysis, we show that these results (the positive effect of the nudge and the inhibitive influence of transparency) hold only for male participants. Given the proliferation of nudges in public policies...

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 Enrico Colombatto, Robert Dur, Christoph Engel, Klaus Heine, Jurjen Kamphorst, Sophie van der Zee, Peter Wakker and the participants of the European Association of Law and Economics Conference in Tel Aviv, Israel; the Experiments at the Crossroads of Law and Economics Workshop in Rotterdam, The Netherlands; the IREF workshop in Aix-en-Provence, France, for their useful comments. In addition, we are grateful to Liam Wells for his editorial assistance with the experimental design. Remaining errors are however ours. Replication files are provided in the supplementary information.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Kyklos published by John Wiley & Sons Ltd.
around the world, these results call for further research on nudges and transparency.

**KEYWORDS**

nudge, regulation, social norms, transparency

**JEL CLASSIFICATION**

D91; K20; K23

---

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 (Alemanno, 2015; Gopalan & Pirog, 2017; Lunn, 2014; OECD, 2017; Sunstein, 2013). Its importance in public policy is evident, as illustrated by the statement of the European Commission's Behavioral Insight Unit in 2016 – “The use of behavioural economics in the design and delivery of regulation is at the forefront of regulatory policy and governance” (Sousa et al., 2016). The idea of nudges, or as it is also termed - choice architecture - is to structure the choice, based on known behavioral economics tools, in such a way that a person will make a decision, which is in his or her best interest or would increase the general welfare of society (Bovens, 2009; Thaler & Sunstein, 2008). Despite its popularity as an additional (soft) regulatory instrument, it has also attracted criticism. Some scholars challenge the legitimacy behind and the efficiency of governmental use of nudges (e.g. Glaeser, 2006; Rebonato, 2012; Schnellenbach, 2012; Schnellenbach, 2019). However, given the fact that notwithstanding the criticism, many governments around the world employ nudges, it is important to investigate whether the use of these nudges can be made “more legitimate”.

The idea of utilizing people’s cognitive and behavioral “biases” is not new and have been investigated for decades, e.g. for marketing purposes (Hossain & List, 2012; Maheswaran & Meyers-Levy, 1990). The new element which appeared with the “nudge movement” is that governments, rather than private actors, should be the ones harnessing behavioral insights to shape public’s behavior. This confirms the idea put forward by Frey & Eichenberger (1994), who suggested that such biases, or “anomalies” in the sense of deviations from the neo-classical rational behavior, provide incentives for others to exploit them.

In this paper, we focus on the criticism about the covert nature of nudges, which might render them manipulative. In the majority of cases, nudges utilize psychological mechanisms without the awareness of the targeted choice-maker (Hansen & Jespersen, 2013; Hausman & Welch, 2010; Schnellenbach, 2016; Wilkinson, 2013). This situation raises concerns that the already over-powerful government is insufficiently constrained by transparency and accountability (Rebonato, 2012). Notwithstanding the importance of transparency of public policies in democratic societies (Bauhr & Marcia, 2014; Kaufmann & Belver, 2005; Roelofs, 2019), disclosure of the potential influence and the motive behind nudges is rare (Glaeser, 2006; Johnson et al., 2012). The importance of making behavioral policies transparent was also emphasized in a recent report by the OECD (2019).

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 their own problem. It is 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 potential 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 accepts nudges. These surveys demonstrate a wide support for nudges in many countries around the world (see for example, Reisch & Sunstein, 2016; Sunstein et al., 2018; Sunstein et al., 2019). 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 the 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.

The trade-off between transparency and effectiveness is an empirical question. Yet, to date it lacks sound empirical evidence (Schmidt, 2017). Furthermore, the theoretical foundation of this problem is not clear. To the best of our knowledge, the transparency problem was almost exclusively investigated in the context of defaults.1 With this type of nudge, there seems to be a consensus that transparency does not impede the effectiveness of defaults (Bruns et al., 2018; Loewenstein et al., 2015; Paunov et al., 2018; Steffel et al., 2016). 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. Investigating the effectiveness of transparent nudges brings libertarian paternalism (Thaler & Sunstein, 2003) closer to autonomy-enhancing paternalism (Binder & Lades, 2015). Even though transparent nudges may still focus on outcomes rather than the decision process, it induces more reflective choices.

In the light of the above, this paper takes the important step in investigating the transparency problem in the context of other nudges besides defaults. In particular, we focus on a (descriptive) social norm nudge. Social norms, which entail presentation of statements regarding what other people do (descriptive), or think should be done (injunctive), 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 altogether.

To investigate the transparency problem with respect to social norms, we have conducted an experiment with abstract gambling 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 that 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 In addition, there are two streams of literature which suggest that in this type of an experiment men and women will behave differently. First, it has been demonstrated in a large number of experiments in the context of abstract gambling games that women are more risk averse than men (e.g. Byrnes et al., 1999; Eckel &

---

1To the best of our knowledge, the only exception is a study by Kroese et al., 2016, which demonstrated that disclosing the reason to rearrange the food in a shop (increasing consumption of healthy food) did not diminish the effectiveness of the nudge.

2By “intention” we 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 on average for the participants.
Grossman, 2002; Eckel & Grossman, 2008; Levin et al., 1988). Second, in a recent study, it has been found that men were more responsive to the descriptive social norm than women were in their donation behavior Croson et al. (2010). Therefore, we have also performed heterogeneity analysis.

Our findings demonstrate, consistent with Croson et al. (2010), that the descriptive social norm was effective only for men. On the other hand, both genders made a similar choice in the control group, thus not confirming the stronger risk aversion of women. Of itself, this is an important finding. Given the growing literature on the heterogeneity of effects with respect to the effectiveness of nudges (e.g. Beshears et al., 2015; Bronchetti et al., 2013; 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. 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 conduct further research on the interaction between meaningful transparency and the effectiveness of different nudges.

The paper is structured as follows. In section 2, we introduce the theoretical framework that will serve as the basis for our predictions. 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 the 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 manipulative (e.g. Hansen & Jespersen, 2013; Wilkinson, 2013), and limit individuals’ autonomy to evaluate, deliberate and chose for themselves (e.g. Hausman & Welch, 2010). One might suggest that as such this is not a significant impediment given that in democratic societies there are instruments to increase the transparency of policies. For example, nudges can be brought to the public debate by opposition parties or the free media. However, it is not self-evident that the existence of such mechanisms reduces the governments’ obligation to increase the transparency of its policies, or at least to understand whether such policies enjoy wide support once properly disclosed.

Therefore, one potential solution to tackle this criticism is to introduce transparency when using nudges in public policy. Such calls have already been explicitly made in official reports (e.g. House of Lords, 2011; [Dutch Scientific Council for Government Policy report 2014]). However, as Bovens (2009) discussed, there can be two types of transparency in this context: (1) type interference transparency – the government announces 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). Despite its potential desirability from an accountability and legitimacy perceptive, some scholars raise a concern that token transparency would harm the effectiveness of the nudge (Bovens, 2009).

One psychological mechanism, which may induce the transparency problem, is psychological reactance (Brehm, 1966). The underlying idea behind the psychological reactance theory, is that an interpersonal threat to freedom stemming from attempts to influence or pressure someone to a certain decision, might lead them to try to restore this sense of freedom (Brehm, 1966; Brehm & Brehm, 1981; Clee & Wicklund, 1980).
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). To the best of our knowledge, the transparency problem was directly investigated almost exclusively 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), the environment, and charitable giving (Bruns et al., 2018; Steffel et al., 2016). In all these studies, transparency did not influence the effectiveness of the default. Moreover, one study in the context of courses 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 the introduction of transparency does not inhibit the effectiveness of nudges in the form of default rules.

Despite the importance of this empirical evidence, 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; Sunstein & Reisch, 2013). Social norms on the other hand, harness 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 to 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 or she is being treated as a conformist because most people do what others do, might touch upon the 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.3

2.2 | Social norms

The ability of actions and opinions of others, i.e. social norms, to influence an individual’s decisions is a well-established phenomenon in social psychology. Even though traditionally, social norms are understood as only moral prescriptions, theories of social influence emphasize two different meanings of the term “norm”. A norm means something which is socially desirable, but also something which is simply common and normal (Cialdini et al., 1991). Normative or injunctive norms are effective because they signal the moral rules that a person should follow. The individual motivation to follow these rules is the belief in 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 that 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; Cialdini et al., 1991; Deutch & Gerard, 1955). One should note that descriptive norms have an effect on people’s behavior even when they are entirely neutral and bear no moral value (e.g. Asch, 1956; Milgram et al., 1969; Sherif, 1963; Venkatesan, 1966).

After many years of investigating the different elements, which are responsible for the effect and the power of injunctive and descriptive social norms, they were adopted as one of the instruments used by governments (or other entities) to

---

3One should note that despite this experimental investigation being novel in terms of examining the legitimacy of social norm nudges, it can also be connected to a long discussion in behavioral economics about competing frames. A nudge in essence is a choice of framing. The idea that the way a message is framed affects people’s choices has been known already from the 1970s. The introduction of the prospect theory by Kahneman & Tversky (1979) sparked research into the ways people’s behavior can be affected by framing in different fields (e.g. Hossain & List, 2012; Maheswaran & Meyers-Levy, 1990; Rothman & Salovey, 1997). However, even though the effect of isolated frames is clear by now, it is unclear what happens when there are competing frames. Introducing transparency can be viewed as a competing frame to the isolated nudge.
influence people’s behavior. For example, social norms were used to increased tax compliance (Bott et al., 2017; Hallsworth et al., 2017) or reduce smoking (Thaler & Sunstein, 2008).

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 (Hallsworth et al., 2017). The effectiveness of social norms was found also in other domains, such as energy conservation (Allcott, 2011), 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 & John (2017) examining social norms in the context of late payment of tuition fees in the UK found no evidence of the effectiveness of descriptive social norms in enhancing desirable behavior. Moreover, Beshears et al. (2015) found a boomerang effect of a descriptive social norm, which led to the opposite of the (desired) peer behavior.

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 heterogeneity of effects where some groups have stronger preferences against a particular nudge than others. Consequently, it 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 their effectiveness.

2.3 | Theoretical predictions

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

2.3.1 | 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.

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 the introduction, women might react differently to the nudge either because descriptive social norms have a different effect on them (Croson et al. (2010), or because they are more risk averse (e.g. Eckel & Grossman, 2008).

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 are no theoretical grounds 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].

4For a review see Kantorowicz-Reznichenko and Wells (2021, forthcoming).
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.

H3. If participants receive information on the way social norms work and on the purpose of using 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.

3 | EXPERIMENTAL DESIGN

3.1 | The experiment

In order to examine the influence of transparency on the effectiveness of social norms, we designed a simple, between-subjects, lottery choice experiment. Given that the general population is expected to be 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). The logic of our experiment followed the simple economic assumption that “more is better”. Therefore, we aimed to give participants an option most of them would prefer (less risky) and then to nudge them to choose the option that on average is expected to provide a higher payoff (riskier). If the nudge has an effect, we then have an opportunity to investigate what happens to its influence once it becomes transparent.

All participants received instructions how the lotteries work and were asked to choose which of the two lotteries to play. After this choice, they 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.

In order to encourage people to choose the lottery with the higher expected return, we 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

<table>
<thead>
<tr>
<th></th>
<th>Blue</th>
<th>Yellow</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lottery A</td>
<td>7</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Lottery B</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

FIGURE 1 The lotteries payoffs
[Colour figure can be viewed at wileyonlinelibrary.com]

5The design is inspired by Billion & Pieter (2018).
In the 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). In addition to the social norm manipulation, we included two treatments with two different levels of transparency (simple transparency, with only the intention and the mechanism of the nudge, and full transparency with the additional information about the goal of the nudge). The design of the experimental treatments is presented in Table 1.

Besides examining the main effects of the nudge and the additional transparency, we also examined whether these effects differ depending on the gender of the participant. Our dependent variable is a dummy indicating whether a respondent chose the risky alternative (1) or not (0). The baseline for the social norm effect is the control group, and the baseline for the transparency effect is the social norm group.

### 3.2 Procedure

Participants (N = 936) were recruited via the online platform Prolific Academic and the experiment was programmed and data collected via the software Qualtrics between December 2018 and January 2019. An online platform gives the advantage of recruiting more demographically diverse participants than in experimental labs, which potentially increases the external validity 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 et al., 2017).

### TABLE 1 Experimental groups

<table>
<thead>
<tr>
<th>Experimental group</th>
<th>Information provided (manipulation)</th>
<th>N</th>
<th>% female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>No information</td>
<td>196</td>
<td>62%</td>
</tr>
<tr>
<td>Social Norm</td>
<td>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.</td>
<td>188</td>
<td>62%</td>
</tr>
<tr>
<td>Social Norm + Transparency</td>
<td>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.</td>
<td>182</td>
<td>59%</td>
</tr>
<tr>
<td>Social Norm + Transparency + Purpose</td>
<td>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).</td>
<td>182</td>
<td>61.5%</td>
</tr>
</tbody>
</table>

Note: the respective N excludes participants who failed the manipulation check or responded wrongly to both attention check questions. Therefore, it includes only participants who were part of the analysis.
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. 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. Participants were randomly allocated to the experimental groups (right after reading the instructions), and those who answered the manipulation check question correctly could continue with the study. The manipulation check was meant to guarantee participants paid attention to the experimental 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. 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 or her chosen lottery (0, £7 or £20).

4 | RESULTS

4.1 | Subjects’ characteristics

Our sample totals 748 participants (referred to as the “entire sample”), 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 under-representation of highly educated people in the control group. Statistical tests suggest, however, that these differences are not significant. To check for the robustness of the results, we nonetheless control for these additional covariates in the multiple regression models. For further details, see Table 2.

4.2 | 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 the participants chose lottery B (riskier but higher payoff choice) in the control group and 31% of the participants chose the same lottery in the (social norm) nudge treatment. By using bivariate non-parametric chi-square test, we find that this difference is not statistically significant ($\chi^2 = 1.934, p = 0.164$).

| TABLE 2 | Participants’ characteristics per experimental group |
|----------|-----------------------------|-----------------|----------------|
| Group    | % female | Average age | % highly educated |
| Control  | 62.2%    | 34.6        | 52.0%           |
| Social Norm | 61.7%  | 34.6        | 60.6%           |
| Transparency | 59.3%  | 35.6        | 62.1%           |
| Transparency + purpose | 61.5% | 34.3        | 61.5%           |

Note: Tests for differences across groups in terms of (1) female share $\chi^2(3) = 0.38, p$-value $= 0.944$; (2) age $F = 0.54, p$-value $= 0.654$; (3) shares of highly educated respondents $\chi^2(3) = 5.33, p$-value $= 0.149$.

---

6Given that the instructions were in English, we chose an English-speaking sample. The unemployment rate in the UK was low (4%) at the time of conducting the experiment and given the possibility of unemployed participants being greatly over-represented 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.
Turning to the variation between treatment effects, we are able to observe striking differences between treatment effects for women and for 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 ($\chi^2 = 4.895, p = 0.027$). This therefore provides evidence in favor of H1a. Furthermore, given the fact that the vast majority of participants in the control group have chosen the less risky lottery (75%), the initial assumption for the experimental design that participants would generally be risk averse seems to be supported.

Nevertheless, the similar propensity for risk aversion between men and women in the control group (26% of man and 25% of female, respectively, chose the less risky lottery) suggests that the underlying drive of the difference is variation in the response to the descriptive social norm, consistent with the findings of Croson et al. (2010). It does not confirm on the other hand, the explanation that women will react differently due to their higher risk aversion as suggested by experimental studies using abstract gambling games (e.g. Eckel & Grossman, 2008). Our results cannot apprise on the reasons why women are less susceptible to social norms than men. However, this question goes beyond the focus of this study, which is the influence of transparency on the effectiveness of nudges.

Moving beyond bivariate comparisons and to further substantiate these results, we run a set of multiple logistic regressions, where our dependent variable is a binary variable coded with 1 if 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 chi-square tests reported above.

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 entire sample (Table 3, Model 3), we also present the results for women and men separately (Table 3, Model 4). We start the analysis by first looking at the differences graphically. Figure 2 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 non-parametric chi-square test confirms that this difference is statistically significant ($\chi^2(1) = 9.497, p = 0.002$). See Table 3 Model 3, which further substantiates these results via multiple 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 3 Model 4) point in 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 women, simple disclosure led in fact to a boomerang effect since the probability of choosing lottery B is now lower than in the control group, and this difference is statistically significant ($\chi^2 = 5.006$, $p = 0.025$).

This result is potentially 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 of choosing lottery B could be viewed in the light of the psychological reactance theory as participants’ attempt to restore their freedom.

The question remains of 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 the propensity of choosing lottery B. However, for the entire sample, these differences in the nudge and full transparency groups are not statistically significant ($\chi^2 = 2.029$, $p = 0.154$). 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 ($\chi^2 = 3.940$, $p = 0.047$). At the same time, the difference between control and full transparency group for men is not statistically significant ($\chi^2 = 0.040$, $p = 0.842$).

Likewise, these results can be observed in Table 3 Model 4, which presents the results from multiple logistic regression. Notice that the coefficients on the “Transparency+Purpose” variable and the interaction term “Transparency+Purpose * 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 (the full transparency effect). As can be seen in Model 4, full transparency also reduces the propensity of men to choose lottery B 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.
**TABLE 3**  Main effects: Logistic regressions

<table>
<thead>
<tr>
<th>Dependent variable: Choice of lottery B</th>
<th>(1) Control v. SN</th>
<th>(2) Control v. SN</th>
<th>(3) SN v. transparency</th>
<th>(4) SN v. transparency</th>
<th>(5) Control v. all treatments</th>
<th>(6) Control v. all treatments</th>
<th>(7) Control v. all treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control v. SN</td>
<td>No interaction</td>
<td>Gender interaction</td>
<td>No interaction</td>
<td>Gender interaction</td>
<td>No interaction</td>
<td>Gender interaction</td>
<td>Full set of controls</td>
</tr>
<tr>
<td>Social norm (SN)</td>
<td>0.316 (1.39)</td>
<td>0.783 † (2.19)</td>
<td></td>
<td></td>
<td>0.316 (1.39)</td>
<td>0.783 † (2.19)</td>
<td>0.779 † (2.18)</td>
</tr>
<tr>
<td>Transparency</td>
<td>−0.763 †† (−3.05)</td>
<td>−0.855 † (−2.37)</td>
<td>−0.446 † (−1.75)</td>
<td>−0.0721 (−0.19)</td>
<td>−0.0749 (−0.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparency + Purpose</td>
<td>−0.331 (−1.42)</td>
<td>−0.708 (−1.97)</td>
<td>−0.0147 (−0.06)</td>
<td>0.0755 (0.20)</td>
<td>0.0724 (0.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>−0.0577 (−0.17)</td>
<td>−0.866 †† (−2.69)</td>
<td></td>
<td></td>
<td>−0.0577 (−0.17)</td>
<td>−0.0532 (−0.16)</td>
<td></td>
</tr>
<tr>
<td>Social norm (SN) * Female</td>
<td>−0.808 † (−1.73)</td>
<td></td>
<td></td>
<td></td>
<td>−0.808 † (−1.73)</td>
<td>−0.824 † (−1.76)</td>
<td></td>
</tr>
<tr>
<td>Transparency * Female</td>
<td></td>
<td>0.0963 (0.19)</td>
<td></td>
<td></td>
<td>−0.712 (−1.37)</td>
<td>−0.727 (−1.39)</td>
<td></td>
</tr>
<tr>
<td>Transparency + Purpose * Female</td>
<td></td>
<td>0.657 (1.38)</td>
<td></td>
<td></td>
<td>−0.151 (−0.31)</td>
<td>−0.174 (−0.36)</td>
<td></td>
</tr>
<tr>
<td>Higher education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.172 (0.97)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−0.006 (−0.73)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>−1.099 ††† (−6.66)</td>
<td>−1.063 ††† (−3.99)</td>
<td>−0.782 †† (−4.98)</td>
<td>−0.280 (−1.17)</td>
<td>−1.099 ††† (−6.66)</td>
<td>−1.063 ††† (−3.99)</td>
<td>−0.950 † (−2.37)</td>
</tr>
<tr>
<td>N</td>
<td>384</td>
<td>384</td>
<td>552</td>
<td>552</td>
<td>748</td>
<td>748</td>
<td>748</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.004</td>
<td>0.020</td>
<td>0.016</td>
<td>0.034</td>
<td>0.012</td>
<td>0.025</td>
<td>0.027</td>
</tr>
</tbody>
</table>

Note:

+ $p < 0.10$,

* $p < 0.05$,

** $p < 0.01$,

*** $p < 0.001$. Z statistics in parentheses.
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 women (see the negative and statistically significant coefficient next to the interaction term “SN * Female”). Model 7 provides a further robustness check as it controls for two additional variables, age and higher education, and it shows that results are stable.

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 descriptive social norm nudge. Our findings demonstrate two main outcomes. First, the descriptive social norm in the context of a lottery choice was found to be effective only for the male participants. Therefore, it raises an important policy question – should nudges be adjusted to (sub)groups to be more effective? The heterogeneity of effects 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). 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, which 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 & Rogers (2009) demonstrated that high turnover descriptive social norms affected only infrequent voters.

Second, and the main finding of our study is that meaningful transparency has the potential to inhibit the effectiveness of a social norm nudge. Not only did both types of transparency, with purpose, and without, reduce the desired choice for male participants, but 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 choice architecture used 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 of 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 architecture. On the other hand, nudges that 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, thorough and comprehensive research must be conducted.

Our study also faces some limitations. First, its narrow context limits the ability to have a broader conclusion regarding social norm nudges and transparency. However, this is also not our claim. We only stress that our findings provide tentative evidence that transparency have a potential to influence the effectiveness of social norms. Therefore, and especially in light of the importance of transparency for public policies, more research is needed. Second, the use of Prolific participants limits the external validity of the results. Hence, other samples as well should be used in future research. In our view, those limitations do not diminish the significance of our results, but emphasize the importance of further research in this area.
REFERENCES


---

**How to cite this article:** Kantorowicz-Reznichenko E, Kantorowicz J. To follow or not to follow the herd? Transparency and social norm nudges. Kyklos. 2021;74:362–377. [https://doi.org/10.1111/kykl.12274](https://doi.org/10.1111/kykl.12274)