*Behavioural Public Policy* (2023), **7**, 914–923
doi:10.1017/bpp.2023.5

ARTICLE

# Norm nudging and twisting preferences

Cristina Bicchieri

Center for Social Norms and Behavioral Dynamics, University of Pennsylvania, Philadelphia, PA, USA
Email: cb36@sas.upenn.edu

(Received 31 January 2023; accepted 31 January 2023; first published online 29 March 2023)

**Abstract**
Public policies often involve "norm nudging", the use of norm information to steer individual behavior in a prosocial direction. Analysis of social norm messaging often concentrates on the outcome measure: the potential change in behavior. The cognitive and psychological processes that underlie individuals' response to norm information, especially the inferences they draw, are usually overlooked. This knowledge gap may lead to adverse consequences, such as messages backfiring. To (factually) justify norm nudging interventions, it is essential that policymakers understand the complex mechanisms that link context, social expectations and preferences, and how the interpretation of different types of messages affects behavior.

## Norm nudging

The basic idea of nudging states that it guides individuals toward those behaviors they *would prefer* to act on were they fully informed, rational and immune to weakness of the will. Nudging would encourage them to realize their true preferences, without restricting their choice set. It is the very presence of different options that allows nudging supporters to claim that there is no infringement on freedom of choice. However, critics are quick to point out that exploiting individuals' psychological biases is, in itself, a form of hidden manipulation. As their objection goes, a person whose choice is manipulated is *unaware* that their beliefs, preferences or emotions have been influenced. Despite what the actors may feel, their choice is not the outcome of a conscious, careful consideration of which goals they should achieve or a rational deliberation about the best means to realize them. In the worst case, manipulation may lead a person to seek harmful goals or hold false beliefs, but even if the induced goals were positive and the beliefs correct, critics could still argue that the nudged individual is, in fact, no longer free to choose.

On the other hand, is awareness of being influenced by nudging necessary to ensure that one is still making a free choice? What if a light form of manipulation that leaves all options open can induce people to make better, more rational choices? Cognitive psychology tells us that we often have little direct introspective awareness or access to our higher-level cognitive processes (Nisbett & Wilson, 1977). We are not aware of the influence that stimuli we consider irrelevant have on our responses. For example, we are usually blind to contextual factors, such as position, serial order and anchoring effects. We may choose an object because of its position (say, to one's left or right), but would find it implausible to believe that such a trivial factor has influenced our choice. When asked to explain our behavior, we usually offer credible reasons and refer to evidence we are conscious of as a plausible cause of our action. This does not mean that the stimulus we found influential was indeed causally efficacious. In fact, the elements that we are unaware of, like implicit perceptions, may be the true drivers of our decisions. The 'manipulations' of nudging simply make specific stimuli available and salient. Provided alternative options are accessible and transparent, and the final choice is for the individual to make, unawareness of the stimuli's causal influence does not seem to deny people's capacity for agency.

The original nudges addressed "internalities", i.e., errors in decision making that lower individual utility. Examples of such nudges include the setting of defaults, like automatically enrolling employees into 401ks (Madrian & Shea, 2001), implementing intention prompts to increase influenza vaccinations (Milkman *et al.*, 2011) or food repositioning to encourage healthy food choices (Kroese *et al.*, 2016). Yet nudges can also be used to address "externalities", i.e., choices that affect others, especially those choices that diminish others' utilities. Here, I will exclusively focus on such nudges, commonly called "norm nudges". Such nudges may inform a target about what other people do in the same or very similar circumstances (e.g., *Your use of electricity is greater than most people's use in your neighborhood*), what other people approve of (e.g., *The vast majority of residents support recycling*) or even involve a direct injunction (e.g., *Please do not litter on the beach*). Norm nudging generally involves some form of social comparison: It informs individuals about what others do and/or approve or disapprove of with the aim of inducing or discouraging some target behavior.

## Norm nudging: a factual justification

The main debate on norm nudging, not unlike the discussions surrounding individual nudges, is often focused on the *justification* of such interventions. Clearly, whether a policy is justified depends on our views of freedom, autonomy and wellbeing, and how much weight we cast on each of these values. Yet justification should also involve a *factual judgment* concerning, among other things, the capability of a policymaker to know how to effectively correct people's mistakes and their competence in enacting sound policies, such as norm nudging, that aims to steer people toward prosocial behavior. A factual judgment goes beyond simply questioning the effectiveness of an intervention. Even if an intervention is deemed to be morally justified, and even if a (positive) behavior change has occurred, do we know why it occurred? Do we know if it will last or eventually reverse? How can we be sure it will not soon backfire? Answering these questions requires a deep understanding of the mechanisms of norm nudging, and awareness of the pitfalls one may encounter and try to avoid.

Let's start by looking at the social comparisons entailed by norm nudging. When we inform individuals about what others do and/or approve or disapprove of, we implicitly

https://doi.org/10.1017/bpp.2023.5 Published online by Cambridge University Press

assume that the target behaviors are *interdependent*, i.e., one's choice is a function, among other things, of what one believes others do and/or approve of. Without such interdependence, informing people about others' actions or normative attitudes would have little effect on behavior. More precisely, interdependence means that the preference for performing or abandoning a target behavior is *conditional* on the kinds of social expectations we hold (Bicchieri, 2006, 2016). To foster behavior change, we thus need to change individuals' social expectations in order to change their preferences. Though an individual's initial behavior may be selfish, the policymaker can hope to induce a new, prosocial behavior by changing the social expectations the target behavior is conditional on, shaping new preferences. This, however, is no easy feat.

The policymaker must learn what kind of social expectations underlie the specific behavior they are hoping to change or induce and which type of message framing is best suitable for their intended purpose. Unfortunately, much of the literature on norm nudging does not explicitly acknowledge expectations and conditional preferences as essential to the mechanisms that facilitate behavioral change. A deeper analysis of their function and interaction within the context of norm nudging is necessary if we wish to understand the conditions under which such nudging may be (factually) justified (Bicchieri & Dimant, 2019; Bicchieri, 2022). It is my hope that by elaborating on the intricate but incredibly relevant nuances of the relationship between expectations and norm change, we will clarify the fundamental role this relationship plays in behavioral interventions and incorporate it, in greater detail, into future studies, policies and discussions.

## Norms: descriptive and social

A better understanding of which social expectations may matter to behavior, and how, also helps to decide which norms we are talking about. Unfortunately, the word 'norm' is used quite casually in the nudging literature. It may mean what is commonly done, what is commonly perceived as acceptable, and even what is commonly deemed to be moral. Norm nudging may involve messages about what others do, such as Alcott's (2011) comparison of one's electricity consumption to neighbors; messages about what is the right thing to do, such as "April is 'Keep Arizona Beautiful' month. Please do not litter" (Cialdini *et al.*, 1991); or messages about what is commonly approved of, such as "Shoppers in this store believe that reusing shopping bags is a worthwhile way to help the environment" (De Groot *et al.*, 2013). If we hold the assumption that norm nudging always refers to interdependent behaviors, then we are left with only two types of norms to work on: descriptive and injunctive. This distinction is usually quite coarse and would benefit from some refinement. Usually by 'descriptive norm' is meant what people normally do, and this may encompass drinking coffee in the morning as well as stopping at red lights and passing at green ones when driving. By 'injunctive norm' is commonly meant what is considered right, approved or prescribed. This could encompass Kashrut laws about food as well as the common rule of reciprocating favors. Yet such rough distinctions are not helpful when we need to decide (a) whether the target behavior is effectively grounded on social expectations (do they play a causal role?) and (b) which expectations we should aim to change.

https://doi.org/10.1017/bpp.2023.5 Published online by Cambridge University Press

A finer, more useful distinction might be the following (Bicchieri, 2006). A *descriptive norm* is a pattern of behavior such that individuals prefer to conform to it on the condition that they believe that most people in their reference network conform to it. This means the behavior is only conditional on empirical expectations, and such expectations refer to a specific group that matter to the decision-maker in the specific circumstances of the choice. So collective habits, or customs, are excluded here, since social expectations do not play a causal role in supporting, say, the Italian morning coffee habit, and changing this habit might involve learning new factual information (coffee contains unhealthy substances, it increases the heart rate, causes dependency, etc.) It may also happen that changing a custom involves *creating* new social expectations and preferences. For example, the Indian open defecation custom can be changed by informing people about the increasing number of households that build and use toilets (Ashraf *et al.*, 2021). Customs, in other words, may be changed by creating new descriptive norms, but the preexisting social expectations (about the frequency of open defecation, for example) were not influencing the behavior.

Traffic rules, on the contrary, are the exemplary case of descriptive norms. On 3 September 1967, Sweden switched from driving on the left-hand side of the road to the right. Apart from the needed changes in traffic lights, road signs, intersections and road lines, a massive media campaign was enacted not only to inform people of the move, but to coordinate the massive change in empirical expectations that was needed.

A *social norm* instead is a rule of behavior such that individuals prefer to conform to it on condition that they believe that (i) most people in their reference network conform to it and (ii) that most people in their reference network believe they ought to conform to it (Bicchieri, 2006, 2016). In a social norm, the behavior is conditional on both empirical and normative expectations. Note, again, that expectations refer to a *reference network*, a group of people that matters to the agent when she has to make a choice. Prosocial collective behaviors are often social norms. Actively rejecting corruption, not cheating on tax payments, and recycling materials, are all behaviors that – for many of us – require the expectation that others do not cheat, bribe and actively recycle. Is this reassurance enough? If good behavior has a cost, as it usually has, the temptation to be selfish and free ride is often present. That is why the 'push' of normative expectations is often needed. We need to know that – even in the absence of formal sanctions – our selfish behavior will be disapproved, reprimanded and chastised.

If norm nudging aims to influence behavior via social information, the policy maker must decide which message best fits the situation at hand. In other words, which expectations does she want to induce or change? On the one hand, we have empirical expectations, i.e., what people expect other people to do. On the other hand, we have normative expectations, i.e., what people expect other people to approve or disapprove of. Integrating norm change interventions into policy requires carefully framing the nudging message to appropriately target the specific interdependence that exists (or we want to create) in a real-world setting.

## How information affects behavior

In order for an intervention to be effective, a policymaker must take into account several elements. First, one needs to correctly identify the mechanism through which

https://doi.org/10.1017/bpp.2023.5 Published online by Cambridge University Press

different types of information affect behavior. Sometimes purely descriptive information is effective, as when informing college students of the real percentage of drinking on campus (Perkins, 2003). This is the case when we are trying to change a descriptive norm. At other times, a normative message may succeed in inducing better behavior, as when Illinois residents are told that the vast majority of them support energy conservation (Bolsen, 2013). And finally, there are times where the combination of descriptive and normative information proves successful, as when energy consumption information is accompanied by a smiling emoticon (Schultz *et al.*, 2007). This latter combination helps when we want to change or bring about a new social norm.

Descriptive information about the true number of people who perform a specific (positive) action may turn out to be very useful, especially if there is underestimation of how frequent or common the action is. In cases like these, descriptive information, if trusted, helps people decide to engage in more prosocial behavior, such as paying taxes on time (Hallsworth *et al.*, 2017). Likewise, information about how many people do in fact disapprove of a specific behavior, such as drinking alcohol, practicing unsafe sex or bullying, may also be useful in steering people away from actions they actually dislike but cannot openly criticize (Miller & McFarland, 1987; Perkins, 2003). More often than we can imagine, the perceived norm (descriptive or social) and the actual norm do not overlap, and transparent communication about real endorsements or just real frequencies is not possible. Take for example underage drinking on a college campus. When students overestimate it and feel pressure to conform, they engage in a behavior they may not want to partake in but find hard to avoid. This decision may even be accompanied by feelings of shame, guilt and fear. Pluralistic ignorance can be hard to discern on the surface, which is why policymakers must take the time to measure individuals' underlying expectations and compare these beliefs with the true prevalence or endorsement of a given behavior.

Other policy interventions need not assume there is a biased misperception of behavior or normative attitudes that need to be corrected. These are instances in which a norm is not already present (or believed to be present). For example, this happens with public goods provision, where individual and collective welfare can be at odds and need to be reconciled. Often one must choose between being the fool or the knave, contributing when not enough people do, or exploiting the good will of contributors. Norm nudging in this situation should highlight the commonness of prosocial behavior, and the general approval that accompanies it. The problem is that often bad behavior happens to be common behavior. For instance, conveying information that corruption is common and disapproved of backfires (Cheeseman & Peiffer, 2022), because its commonality normalizes the behavior, making it more acceptable even to those who would otherwise refrain from it. Corruption campaigns that stress the prevalence of corruption exacerbate the problem precisely because the associated normative claim is undermined. This is a relatively common situation when empirical and normative information are incongruent. In this case, we know that empirical information about negative behavior will trump any normative claim put forth by the policymaker (Bicchieri & Xiao, 2009). One way to address this problem is to highlight the good behavior of the minority, and the advantages that accrue to such behavior. In the case of energy consumption, Nolan *et al.*

https://doi.org/10.1017/bpp.2023.5 Published online by Cambridge University Press

(2008) added a positive emoticon to the reported energy consumption of low users, helping them feel proud of their choices, and at the same time informing their peers that it is possible to consume less and save money.

Paying attention to the context in which a target behavior occurs is another element that may "make or break" a policy intervention. Context, here, refers to what people can observe, and the beliefs that they hold about their environment. As we shall see, much data supports the view that context plays a crucial role in people's *inferences* from norm messages. For example, Goldstein *et al.* (2008) found that hotel guests of a midsize hotel in the US reduced their towel usage when receiving empirical information that the majority of fellow guests staying in the same hotel reused their towels. However, when using the same framing on a European population, Reese *et al.* (2014) found no effect on hotel towel reusing. As this example suggests, a more cost-effective approach would have been to pilot the message with various framings prior to scaling-up the intervention. Another important element to consider is the correct choice of the reference group. In Goldstein's experiment, when the norm nudging message referred to towel reuse among guests not just in the same hotel, but in the same room as the subject, the intended effect was even greater. This suggests that sharing the same circumstances may create a common identity and establish a successful reference group.

Suppose for a moment that all the above-mentioned difficulties have been overcome. The policymaker is correct in identifying the behaviors that depend on social expectations, and happens to know which expectations, empirical or normative, matter to the target behavior. Even in this fortunate case, a main worry about the effective ability of the norm nudging policymakers to get a significant behavioral change remains. The concern stems from the lack of a clear understanding of the psychological and cognitive process underlying diverse message framings. Norm nudging behavioral studies rely on the assumption that people's behaviors are dependent on their social expectations, which the norm messages aim to elicit or change. However, how the receivers *interpret* the messages is largely ignored. Rather than simply assuming that social expectations influence behaviors, we should be mindful of the complex mental processes occurring in response to the messages.

## Social inferences

The cognitive process to conjecture about other people's attributes, actions and mental states is known as *social inference* (Hastie, 1983), and norm nudging relies heavily on such inferences. Current literature has primarily focused on social inferences regarding actions or mental states of another individual, yet there is still little evidence about inferences regarding *collective others* (Jenkins, 2014; Pegado *et al.*, 2018). Investigating individuals' mental representations about the actions and mental states of social groups is important, for at least two main reasons.

First, the information provided in norm messaging is often ambiguous (e.g., many Philadelphia residents got the flu vaccine this year). The ambiguity of the message leaves room for the receiver to give a self-serving interpretation (Dana *et al.*, 2007). Does "many" mean a majority or just a significant number? And even if one were more explicitly told that "the majority of Philadelphia residents got the flu vaccine

https://doi.org/10.1017/bpp.2023.5 Published online by Cambridge University Press

this year", the majority could be interpreted as 51%, a little over half people got the flu vaccine, or it could be interpreted as 90%, almost all got the flu vaccine. The former interpretation could indicate a weak norm whereas the latter interpretation could signal a much stronger one. Depending on the receiver's prior belief, these two types of interpretations could lead to drastically different impacts on the receivers' behavior. This is an important concern when norm messages conflict with people's self-interest. In this case, individuals may attempt to exploit the "wiggle-room" that was present in the information (e.g., interpret the message to signal a weak social norm) and avoid conforming to the desired behavior (Bicchieri & Kuang, 2022). Social inferences are not just sensitive to social contexts, they are also quite flexible.

Second, increasing evidence indicates that there exists a cognitive link between empirical and normative expectations (Eriksson *et al*., 2015; Lindström *et al*., 2018). If empirical and normative messages convey information about one another, which information we present first may have a major effect on subsequent behavior. An intervention that emphasizes the commonness of a behavior usually implies high approval, whereas emphasizing high approval would seem to imply high frequency (Blanton *et al*., 2008). Indeed, people show a tendency to mix normative and descriptive information in memory recall tasks (Eriksson *et al*., 2015). People also judge the norm transgressor as more punishable when socially desirable behavior becomes more common (Trafimow *et al*., 2004; Welch *et al*., 2005; Lindström *et al*., 2018). Further behavioral evidence shows that the spontaneous inference from one type of social expectation to another can lead to remarkably different behavioral results depending on which social expectation is elicited first (Cialdini *et al*., 1990).

Despite the above-mentioned studies showing the tendency to infer one expectation from the other, the relationship between the two types of information is not simple. In a recent study (Bicchieri *et al*., 2023), participants' empirical (normative) expectations were elicited, and then they were asked to make an inference about their normative (empirical) expectations. Here by *inference* I refer to the reasoning process by which an individual derives the information about a social behavior or beliefs from a summary representation of a group's beliefs or behavior. Participants received a description of the dice task (Fischbacher & Heusi, 2013), where subjects have to anonymously report the result of a roll of the die. They get paid in case they report a five, get nothing otherwise. When individuals received the normative information message "The majority of participants in the dice task disapproved of lying. How many participants did not lie for their own benefit?", they inferred that a considerable portion of participants did not do what they preached. Conversely, when the participants were given the empirical information message "The majority of participants in the dice task did not lie for their own benefit. How many participants disapproved of lying?", they tended to also believe that the participants held parallel normative beliefs. This resulting pattern is one of asymmetry, wherein individuals are inferring normative expectations from empirical expectations but not vice versa. This asymmetry is quite significant and influences the subsequent decision to engage or not in lying.

To complicate matters further, the valence, positive or negative, of the reported behavior is also important in altering people's response to a message (Farrow *et al*., 2017). There is evidence suggesting that individuals may form different

https://doi.org/10.1017/bpp.2023.5 Published online by Cambridge University Press

summary representations of the collective behaviors/beliefs depicted in the norm messages. Depending on the type of social expectation being elicited and the valence of the underlying behavior, inferential processes may lead to very different conclusions. In a recent experiment (Bicchieri & Kuang, 2022), we analyzed 23 prosocial/ antisocial behavioral domains commonly used in norm nudging experiments. Participants were randomly assigned to either (a) judge the approval rates of a behavior among residents of a hypothetical city after receiving empirical information regarding the commonality of the behavior or (b) to judge the prevalence of a behavior after receiving normative information regarding residents' endorsement of it. We manipulated the valence (positive or negative) of the behavior for each domain to test whether it moderated the effect of the type of norm-based information on the social inference.

We found a significant double asymmetry in people's norm inferences. Empirical information about positive behavior leads to a parallel, strong normative inference (i.e., most people donating to charities implies that most people approve of such donations). Normative information about positive behavior, on the other hand, does not lead to a strong empirical inference (most people approving of donating to charities does not imply most people do it). This is the same pattern discussed earlier. However, when the normative information is about a negative behavior, there is a strong, parallel empirical inference (e.g., most people approve of driving above the speed limit implies most people do it). Our evidence supports the mental association of "common" and "acceptable" (Eriksson *et al.*, 2015) when the empirical message is positive or prosocial. When the normative message is negative, the relation between common and acceptable is reversed, and individuals infer others' undesirable behavior to a greater prevalence based on their normative attitudes. In addition, we found the pattern of norm inference could be context-dependent. Specifically, a few behaviors act as "outliers" that do not follow the general double asymmetrical patterns we observed. For example, two of the outliers, such as taking advantage of a job position to collect additional income and bribing to obtain government contracts, resulted in no significant difference in inferences from empirical or normative information in both positive and negative behavior conditions. To understand why some behaviors fall out of the common patterns of norm inference, we did a follow-up experiment and found that outliers can be explained by people's perception of how socially damaging a behavior could be, and not by their baseline rating, frequency or observability. In the large majority of cases, however, inferences are doubly asymmetric, something to keep in mind when designing norm nudges, especially those that aim to create positive empirical conclusions from normative information alone.

## Conclusions

To summarize: norm "nudgers", to be successful, need to verify first that preferences will indeed be influenced by social expectations. The same behavior may be motivated by internal values, be a habitual custom, a descriptive or even a social norm. Diagnosing the nature of the target behavior is the first necessary step in trying to change it. Furthermore, the policymaker must also consider the context and reference group of those it aims to influence. Knowing that a target behavior is carried out by a

https://doi.org/10.1017/bpp.2023.5 Published online by Cambridge University Press

group one strongly identifies with will be much more efficacious than knowing it is performed by even a larger but generic population. Finally, the choice of which social expectations to elicit requires understanding and exploiting the cognitive and psychological processes that underlie people's response to norm information and the inferences they draw, as well as being aware of inference asymmetries.

Messages can and do backfire, and such negative outcomes may reduce trust in the messenger, usually the government that relays them. Avoiding adverse consequences requires the competence to enact programs that call for significant amounts of knowledge and testing out. As is the case with simple nudges that draw on psychological biases, norm nudges make use of our tendency to be influenced by social information, and process inferences about collective behavior or normative attitudes that may not be rationally warranted, as when we consider what is commonly done as also commonly approved of. The goal, in both cases, is to guide individuals to behave in ways that are better either for them individually or for the community at large.

## References

Allcott, H. (2011), 'Social norms and energy conservation', *Journal of Public Economics*, **95**(9–10): 1082–1095.

Ashraf, S., C. Bicchieri, M. Delea, U. Das, K. Chauhan, J. Kuang, P. McNally, A. Shpenev and E. Thulin (2021), 'Design and rationale of the Longitudinal Evaluation of Norms and Networks Study (LENNS): a cluster-randomized trial assessing the impact of a norms-centric intervention on exclusive toilet use and maintenance in peri urban communities of Tamil Nadu', *JMIR Research Protocols*, **10**(5). https://doi.org/10.2196/24407.

Bicchieri, C. (2006), *The Grammar of Society: The Nature and Dynamics of Social Norms*, New York: Cambridge University Press.

Bicchieri, C. (2016), *Norms in the Wild: How to Diagnose, Measure and Change Social Norms*, Oxford: Oxford University Press.

Bicchieri, C. (2022), 'Norm Nudging: How to Measure What We Want to Implement', in N. Mazar and D. Soman (eds), *Behavioral Science in the Wild*, Toronto: University of Toronto Press, 82–107.

Bicchieri, C. and E. Dimant (2019), 'Nudging with care: the risks and benefits of social information', *Public Choice*, **191**: 443–464.

Bicchieri, C. and J. Kuang (2022), *Asymmetries in the Inference from Social Norms Messages*. Philadelphia, CSNBD Paper.

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

Bicchieri, C., E. Dimant and S. Sonderegger (2023), "It's Not A Lie if You Believe the Norm Does Not Apply: Conditional Norm-Following and Belief Distortion". *Games and Economic Behavior*, **138**: 321–354.

Blanton, H., A. Köblitz and K. D. McCaul (2008), 'Misperceptions about norm misperceptions: descriptive, injunctive, and affective 'social norming' efforts to change health behaviors', *Social and Personality Psychology Compass*, **2**(3): 1379–1399.

Bolsen, T. (2013), 'A light bulb goes on: norms, rhetoric, and actions for the public good', *Political Behavior*, **35**(1): 1–20.

Cheeseman, N. and C. Peiffer (2022), 'The curse of good intentions: why anti corruption messaging can encourage bribery', *American Political Science Review*, **116**(3): 1081–1095.

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

Cialdini, R. B., C. A. Kallgren and R. R. Reno (1991), 'A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior', in *Advances in Experimental Social Psychology*, Vol. **24**, New York, Academic Press, 201–234.

https://doi.org/10.1017/bpp.2023.5 Published online by Cambridge University Press

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

De Groot, J. I., W. Abrahamse and K. Jones (2013), 'Persuasive normative messages: the influence of injunctive and personal norms on using free plastic bags', *Sustainability*, **5**(5): 1829–1844.

Eriksson, K., P. Strimling and J. C. Coultas (2015), 'Bidirectional associations between descriptive and injunctive norms', *Organizational Behavior and Human Decision Processes*, **129**: 59–69.

Farrow, K., G. Grolleau and L. Ibanez (2017), 'Social norms and pro-environmental behavior: a review of the evidence', *Ecological Economics*, **140**: 1–13.

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

Goldstein, N. J., R. B. Cialdini and V. Griskevicius (2008), 'A room with a viewpoint: using social norms to motivate environmental conservation in hotels', *Journal of Consumer Research*, **35**(3): 472–482.

Hallsworth, M., J. A. List, R. D. Metcalfe and I. Vlaev (2017), 'The behavioralist as tax collector: using natural field experiments to enhance tax compliance', *Journal of Public Economics*, **148**: 14–31.

Hastie, R. (1983), 'Social inference', *Annual Review of Psychology*, **34**(1): 511–542.

Jenkins, R. (2014), *Social identity*. New York: Routledge.

Kroese, F. M., D. R. Marchiori and D. T. De Ridder (2016), 'Nudging healthy food choices: a field experiment at the train station', *Journal of Public Health*, **38**(2): e133–e137.

Kuang, J., M. Delea, E. Thulin and C. Bicchieri (2020), 'Do descriptive norms messaging interventions backfire? Protocol for a systematic review of the boomerang effect', *Systematic Reviews*, **9**: 1–7. https://doi.org/10.1186/s13643-020-01533-0.

Lindström, B., S. Jangard, I. Selbing and A. Olsson (2018), 'The role of a "common is moral" heuristic in the stability and change of moral norms', *Journal of Experimental Psychology: General*, **147**(2): 228.

Madrian, B. C. and D. F. Shea (2001), 'The power of suggestion: inertia in 401 (k) participation and savings behavior', *The Quarterly Journal of Economics*, **116**(4): 1149–1187.

Milkman, K. L., J. Beshears, J. J. Choi, D. Laibson and B. C. Madrian (2011), 'Using implementation intentions prompts to enhance influenza vaccination rates', *Proceedings of the National Academy of Sciences*, **108**(26): 10415–10420.

Miller, D. T. and C. McFarland (1987), 'Pluralistic ignorance: when similarity is interpreted as dissimilarity', *Journal of Personality and Social Psychology*, **53**(2): 298.

Nisbett, R. and T. Wilson (1977), 'Telling more than we can know: verbal reports on mental processes', *Psychological Review*, **84**(3): 231–259.

Nolan, J. M., P. W. Schultz, R. B. Cialdini, N. J. Goldstein and V. Griskevicius (2008), 'Normative social influence is underdetected', *Personality and Social Psychology Bulletin*, **34**(7): 913–923.

Pegado, F., M. H. Hendriks, S. Amelynck, N. Daniels, J. Bulthé, H. L. Masson and H. O. de Beeck (2018), 'Neural representations behind "social norms inferences in humans"', *Scientific reports*, **8**(1): 1–11.

Perkins, H. W. (Ed.) (2003), *The Social Norms Approach to Preventing School and College Age Substance Abuse: A Handbook for Educators, Counselors, and Clinicians*. New York: John Wiley & Sons.

Reese, G., K. Loew and G. Steffgen (2014), 'A towel less: social norms enhance pro-environmental behavior in hotels', *The Journal of Social Psychology*, **154**(2): 97–100.

Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein and V. Griskevicius (2007), 'The constructive, destructive, and reconstructive power of social norms', *Psychological Science*, **18**(5): 429–434.

Trafimow, D., P. Sheeran, B. Lombardo, K. A. Finlay, J. Brown and C. J. Armitage (2004), 'Affective and cognitive control of persons and behaviors', *British Journal of Social Psychology*, **43**(2): 207–224.

Welch, E. W., C. C. Hinnant and M. J. Moon (2005), 'Linking citizen satisfaction with e-government and trust in government', *Journal of Public Administration Research and Theory*, **15**(3): 371–391.

**Cite this article:** Bicchieri C (2023). Norm nudging and twisting preferences. *Behavioural Public Policy* **7**, 914–923. https://doi.org/10.1017/bpp.2023.5