

# Exploring the mechanisms driving conformity – A replication study

Virginia Jagusch, Sarah Neuhoff, Emily Campos Sindermann

September 14, 2021

## **Abstract**

The descriptive norm effect refers to the phenomenon that individuals prefer behaving similarly to others [1]. The question here becomes whether whether people still conform to others, when they do not identify with their ideas and beliefs in general. Self-categorization theory proposes that an individual's identity is linked to their corresponding group identity. Hence, in order to ensure ingroup norm conformity, self-categorization theory proposes that people will actively avoid conforming to the norms of an outgroup. In two separate experiments, Pryor, Perfors and Howe [1] explored how subjects behaved when presented with a descriptive norm of an outgroup. Their findings suggest that the descriptive norm effect is robust against different group identities, meaning that people seem to conform to a majority irrespective of group identity. Here, we replicated experiment 1 of [1]. The present study did not yield expressive results and could thereby not confirm the findings of [1]. The lack of conclusive results might be due to changes in the experimental design as well as limited sample size.

## Introduction

Our behavior is highly susceptible to outside influence. More precisely, individuals appear to adjust their behavior according to the behavior of others. This phenomenon is known as the descriptive norm effect and has been observed in a variety of contexts [2, 3, 4, 5, 6, 7]. However, the mechanisms underlying this willingness to conform seem to remain unclear [1].

One promising explanation would be that people have an intrinsic desire to maintain ingroup identity and hence adjust their behavior accordingly. This goes in line with self-categorization theory which proposes that an individual's identity is linked to their corresponding group identity [1]. The degree to which individuals would conform to norms of a specific group would therefore depend on how strongly individuals identify with the group's ideas and beliefs. Furthermore, self-categorization theory predicts that in order to retain a sense of ingroup identity, individuals will actively avoid conforming to the norms of an outgroup.

In contrast, other explanations make the more general claim that individuals tend to conform to the majority, disregarding group identity [8].

In order to gain more insight into the mechanisms driving the descriptive norm effect, Pryor, Perfors and Howe [1] tested the prediction of the self-categorization theory, that people specifically avoid conforming to an outgroup, against the more general (alternative) hypothesis that people simply conform to the most popular opinion, irrespective of group identity. In two separate experiments, they observed what effect ingroup and outgroup descriptive norms had on the subjects' conforming behavior. Their findings suggest that the descriptive norm effect is robust against different group identities. People seem to conform to a majority, irrespective of whether that majority pertains to the same ingroup. This argues for a more general mechanism underlying the descriptive norm effect, able to overpower an ingroup vs. outgroup mentality.

Within the scope of this project, we aim to replicate experiment 1 [1]. In the following we will present the methodologies and the results. Finally, we will discuss the results as well as point out limitations of the present study.

## Methods

**Participants:** A total of 72 participants were recruited primarily via distribution of the link to the experiment in group chats of the local Cognitive Science students from the University of Osnabrück. The experiment was hosted on netlify. Participants could interrupt the experiment at any point during the experiment, participant data was only saved and collected at the very end. Participants were not compensated for participating in any way.

**Procedure:** The experiment was implemented using \_magpie, hosted on its servers and distributed via netlify. Participants could complete the experiment on desktop or on mobile.

Upon first clicking on the link to the experiment, participants were greeted by a Welcome screen thanking them for their participation and providing them with information about the experiment. The initial information about the experiment was kept vague, simply stating that the experiment was about "a moral dilemma and how you feel making that choice". Before the beginning the experiment participants were asked to provide basic demographic information (age, gender, level of education and native languages).

Afterwards, participants were asked about which of a list of topics they cared about the most and confirmed their choice by clicking on it. The topics the participants could choose from were "Abortion", "Animal Testing", "The Vaccine against Covid-19", "Protesting" and "Political Orientation". These are not the social issues the participants of the original experiment [1] could choose from, but we came to the conclusion that the social issues used in the initial experiment were too centered on US-related politics and social issues and would feel weaker to the target demographic we were likely to reach with the experiment. Therefore, we used issues we believed would be more present in the minds of German students of today.

Once the participants had chosen their social issue, they were asked to rate their agreement to a simple statement (for example "Abortion should be illegal") about their chosen issue on an 11-point Lickert scale ranging from "-5" "Strongly disagree" to "+5" "Strongly agree".

Once they made a choice, participants were shown another instructions screen, informing them that they were going to be presented with a moral dilemma on the next screen and should pick which course of action they most strongly agreed with, and that they would be asked about their emotional state afterwards.

The moral dilemma presented was the same for all participants. We used the

same dilemma as the one described in the original experiment [1], about a robber stealing money but donating it all to an orphanage. Participants needed to decide whether they would "call the police and report the robbery" or "Nothing, and let the robber be on their way" and how sure they felt about their choice on a 6-point Lickert scale. Underneath the dilemma, participants were presented with one of four possible combinations of sentences. Which sentence condition the participants would be shown was random. The sentences were different combinations about the other participants' choices, and whether they had previously agreed with the participants opinion on their chosen social issue.

The possible sentence conditions were:

1. "Approximately 60% of the other participants who agree with you on your chosen issue chose to let the robber go."
2. "Approximately 60% of the other participants who agree with you on your chosen issue chose to report the robber."
3. With 1. as the first part followed by "Approximately 85% of participants in a previous study who disagreed with you on your chosen issue chose to report the robber."
4. With 2. as the first part followed by "Approximately 85% of participants in a previous study who disagreed with you on your chosen issue chose to let the robber go."

After choosing their course of action in the moral dilemma, participants were asked to click on the correct statement from a series of statements asking about the experiment itself. This was done to see whether they paid attention to the questions and to rule out any participants who just clicked through the answers randomly. Subsequently, participants were asked how they feel about their chosen action in the moral dilemma and chosen to rate their feelings as either "happy" or "sad" or something in between. This question was irrelevant and it's answer was not considered for the purposes of this experiment.

## Results

We excluded 12 participants from the analysis for either failing the understanding check and/or rating their attitude towards their chosen social issue as neutral. The distribution of responses for the remaining 60 participants is shown in Figure 1.

**Model comparison** To find out the extent to which the self-categorization theory provides a better or worse explanation of our data, we directly compared the two theories using the models that were set up by Pryor, Perfors and Howe [1]. Each of the two models represent Bayesian versions of ordinal logistic regression, which predicts the proportions of responses on an ordinal scale. We assume that certain variables, the descriptive norms, change the odds of making higher or lower responses on the scale. A detailed explanation of the models and their prior assumptions can be found in [1].

We used a Bayes Factor (BF) to assess the relative evidence for the self-categorization model and the alternative model, which was calculated with the "Bridge Sampling" package in R [9]. The Bayes Factor represents the probability of the observed data occurring under the alternative model relative to the probability of the data occurring under the self-categorization model.

$$BF = \frac{p(data|alternative)}{p(data|self-categorization)}$$

We only found a BF of 0.89, which means that the observed data is 0.89 times more likely under the alternative model than under the self-categorization model. This outcome does not support the assumption that the alternative model fits the data better than the self-categorization model unlike the original study suggests.

**Effect sizes** We also ran a frequentist ordinal logistic regression to measure the effect sizes of the parameters. However, we did not find a significant effect of the *ingroup descriptive norm* (estimate = -0.05, std error = 0.07, z value = -0.07,  $\Pr(>|z|) = 0.95$ ). There was also no significant shift in bias for the participants that were *both norms shown* (estimate = -0.08, std error = 0.7, z value = -0.11,  $\Pr(>|z|) = 0.91$ ). Lastly, the interaction between *ingroup descriptive norm* and *both norms shown* was not significant (estimate = 0.46, std error = 0.99, z value = 0.46,  $\Pr(>|z|) = 0.64$ ). Hence, there is a lack of clear evidence that the parameters are needed.

As you can see in Figure 2 both models have almost the same posterior densities suggesting that there is no difference in the predictive power of the two models as they would have the same parameter values. Therefore, we could not reproduce the results of the original paper here.

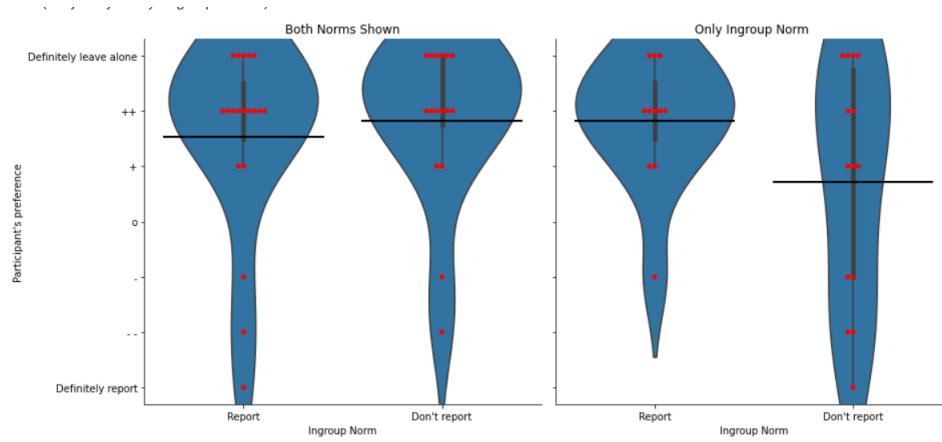


Figure 1: Superimposed violinplot representing responses to the moral dilemma in each condition of the experiment. The red dots represent the relative proportion of each response in each condition. The horizontal black lines represent the mean response in each condition in order to give a better sense of how the pattern of responses changed in each condition.

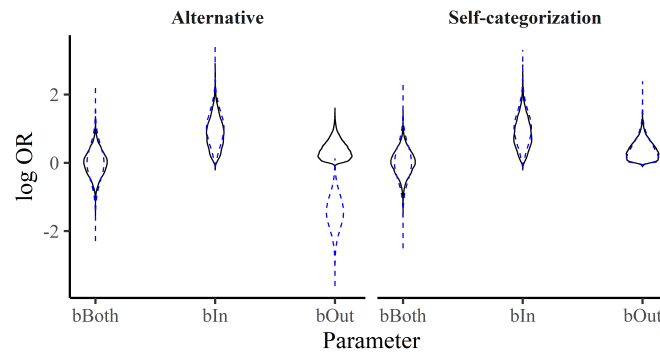


Figure 2: Violin plots of the prior (dashed blue lines) and posterior (solid black lines) density for each of the parameters in both the self-categorization model and the alternative model.

## Discussion

Overall, the findings of the present study do not support the findings of [1]. While the findings of [1] suggest that the descriptive norm effect can be explained by more general group identity independent mechanisms rather than self-categorization theory, our findings cannot confirm this. We only found a very weak BF of 0.89 compared to the BF of the original study ranging from 30.04 to 522.62 [1]. Furthermore, no significant effect was found for the *ingroup descriptive norm*, nor was there a significant shift in bias for participants in the *both norms shown* condition. The interaction between *ingroup descriptive norm* and *both norms shown* also did not show significance. Therefore, we did not achieve to replicate the results of [1].

The present study differed from the original study in mainly two aspects. Firstly and most importantly, the issues participants were presented in the beginning of the experiment, based on which ingroup and outgroup were determined, might not have been polarizing enough to create a strong ingroup vs. outgroup mentality. The original study used mostly political orientation in order to distinguish ingroups from outgroups by using statements which represented a clear republican or democratic stance towards a topic. As this study mostly included German participants, and political orientation in Germany is not as polarizing as in the US, an ingroup vs. outgroup mentality among the participants might not have been as clear as in the original study. Secondly, the present study conveyed information about how other participants decided slightly different than the original study. More specifically, when participants were informed about how other participants behaved, the specific issue in which they agreed/disagreed was not mentioned. For example, in the present study, if participant X indicated that they cared most about abortion, they might have been told that "approximately 60% of the other participants who agree with you **on your chosen issue** chose to call the police and report the robber". In contrast, in the original paper this statement would have been presented as follows: "approximately 60% of the other participants who agree with you **on illegal abortion** chose to call the police and report the robber". This generalization of the information in the present study, might not have conveyed a sense of ingroup as strongly, as the more specific information provided in the original study. Additionally what might be worth mentioning is that a specific p-value was not given in the original study, therefore we assumed the standard p-value of 0.05.

A limitation of the present study was the limited sample size. While the original study comprised 301 participants, the present study only included the data of 60 participants, which was most probably due to time constraints and the inability to provide any (monetary) incentives for participation.

All in all, the present study could not confirm the findings of the original study [1]. We did not find a meaningful BF, nor significant effect sizes of the parameters. Possible explanations could be changes in the experimental design as well as limited sample size. Ultimately, future studies are necessary in order to confirm the observed effect in [1] and therefore establish whether the descriptive norm effect can indeed be explained by general mechanisms instead of group-specific preferences as proposed by self-categorization theory.

## References

- [1] C. Pryor, A. Perfors, and P. D. Howe. “Conformity to the descriptive norms of people with opposing political or social beliefs”. In: PloS one 14.7 (2019), e0219464.
- [2] M. Wenzel. “Misperceptions of social norms about tax compliance: From theory to intervention”. In: Journal of Economic Psychology 26.6 (2005), pp. 862–883.
- [3] P. W. Schultz, J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius. “The constructive, destructive, and reconstructive power of social norms”. In: Psychological science 18.5 (2007), pp. 429–434.
- [4] B. I. Team. “Applying behavioural insight to health”. In: London: Cabinet Office (2010).
- [5] K. Abbink, E. Freidin, L. Gangadharan, and R. Moro. “The effect of social norms on bribe offers”. In: The Journal of Law, Economics, and Organization 34.3 (2018), pp. 457–474.
- [6] N. C. Köbis, J.-W. Van Prooijen, F. Righetti, and P. A. Van Lange. ““Who doesn’t?”—The impact of descriptive norms on corruption”. In: PloS one 10.6 (2015), e0131830.
- [7] C. Bicchieri and E. Xiao. “Do the right thing: but only if others do so”. In: Journal of Behavioral Decision Making 22.2 (2009), pp. 191–208.
- [8] R. N. Rimal, M. K. Lapinski, R. J. Cook, and K. Real. “Moving toward a theory of normative influences: How perceived benefits and similarity moderate the impact of descriptive norms on behaviors”. In: Journal of health communication 10.5 (2005), pp. 433–450.
- [9] Q. F. Gronau, H. Singmann, and E.-J. Wagenmakers. “bridgesampling: An R Package for Estimating Normalizing Constants”. In: Journal of Statistical Software 92.10 (2020), pp. 1–29. DOI: 10.18637/jss.v092.i10.