Social norms as an interactive process: An agent-based cognitive modelling study

Yue Chen[1](yue.chen@ru.nl)
Tibor Bosse[1](tibor.bosse@ru.nl)
Marieke Woensdregt[2](marieke.woensdregt@ru.nl)
[1] Behavioral Science Institute, Radboud University, The Netherlands
[2] Donders Institute for Brain, Cognition, and Behavior, Radboud University, The Netherlands

Abstract

Social norms are often characterized as a system of rules that guide behavior. However, social norms also allow for flexibility; not entirely restricting individuals to one possible behavior. Here, we put forward an agent-based cognitive model that captures social norms as processes that are socially constructed through interactions between individuals. In this modelling work, we focus on the role of norm acquisition and conformity bias in both action production and inference-making. This computational cognitive model allows us to think about social norms along three dimensions: individual vs. collective, behavior vs. belief, and subjective vs. objective. Our simulation results show that increased conformity bias can induce misjudgments about the true desires of others and misalignment between different agents' perceptions of the social norm. However, if agents do not assume that others also conform in their behavior, this increased conformity bias does not necessarily lead to excessive misperceptions of the social norm.

Keywords: social norms; theory of mind; social interaction; pluralistic ignorance; agent-based modeling

Introduction

The presence of social norms is universal across human societies. From avoiding crashes by driving cars on the same side of the road, to sharing food with every member of the camp in hunter-gatherer societies after each hunt, social norms organize the collective of individuals and benefit human communities by fostering cooperation and collaboration (Binmore, 2001; Mackie, 2018). Therefore, they merit study and investigation into their nature and formation.

Social norms are often conceptualized as "a complex set of rules" or "default" behavioral rules stored in our minds, activated by social cues within specific contexts (Bicchieri, 2005; Sripada & Stich, 2006). This rule-based conceptualization is practical for explaining *prescriptive norms* that ought to be adhered to, which are identified primarily by their violations or the punishment that may follow their violations (Cialdini, Reno, & Kallgren, 1990). In contrast, *descriptive norms*, which refer to the tendency of people to conform to the most prevalent behavior in their community, are less rule-like (Hawkins, Goodman, & Goldstone, 2019).

Despite the universal presence of social norms, they manifest in diverse ways. Therefore, the study of norms would arguably benefit from a pluralistic approach (Dale, Dietrich, & Chemero, 2009; Westra & Andrews, 2022). This does not only mean studying norms from multiple disciplinary perspectives (e.g. anthropology, economics), or studying the diversity of normative contents across cultures (e.g., different dining manners), or assuming plurality in the psychological and non-psychological processes that lead to the existence of norms, but also plurality in terms of their conceptualization (Morris, Hong, Chiu, & Liu, 2015; Westra & Andrews, 2022). The rule-based conceptualization provides a valuable starting point, but may be limited in the types of phenomena it can explain; not all norms are restrictive and morally obligatory. Therefore, we argue that conceptualization approaches that are different from the rule-based approach will help gain a fuller understanding of social norms.

Here, we propose a process-based conceptualization of social norms and formalize it in the form of an *agent-based cognitive model* (Smith & Conrey, 2007). Computational models are useful because they can reveal the important essence of reality through their simplicity (Sun et al., 2016). In our model, social norms are seen as processes themselves: continuous, socially constructed, and interactive, rather than static end products of interactions. Such a process-based account may be particularly useful for understanding descriptive norms.

In this paper, we take a particular descriptive norm phenomenon—*pluralistic ignorance*—as a case study, to illustrate what we mean by a process-based conceptualization of social norms. Pluralistic ignorance arises when there is a discrepancy between what individuals believe others desire, and what others truly desire. For example, college students might believe that excessive drinking is what most desire, when actually only a few enjoy drinking that much (Prentice & Miller, 1993). Norm acquisition appears to be crucial in this process. Studies on pluralistic ignorance often describe it as a misperception that occurs in the process of norm acquisition (Miller, 2023). Various and complex processes are involved in the acquisition of social norms (Hawkins et al., 2019). One such (cognitive) process we will focus on in this model is *theory of mind* (ToM), which involves making inferences about others' mental states based on their behavior (Premack & Woodruff, 1978). In our model, we focus on ToM as the psychological mechanism underlying descriptive norm acquisition, as it has been proposed to be involved in the acquisiton of prescriptive norm (Zhi-Xuan & Ong, 2019). In addition to ToM, a tendency to conform appears to be a central driving force behind social norms (Bicchieri, 2005). Our model therefore also assumes a conformity bias in agents' behavior, as well as in how they infer what other agents desire.

The model itself not only provides a process-based con-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

ceptualization but also allows us to think about social norms along three dimensions: individual vs. collective, behavior vs. belief, and subjective vs. objective. Therefore, it can help unify different views on which level of explanation is the right one for studying social norms, such as individual versus collective level (Legros & Cislaghi, 2020). Using pluralistic ignorance as a case study also sheds light on the question of whether people assume that others behave purely out of their own desire, or assume that others work just like them (i.e., also being influenced by their perception of the social norm and some tendency to conform). Assuming that others act purely out of their own desire may mistakenly lead to pluralistic ignorance if others' private attitudes are different from their behavior because they are acting under the influence of a social norm (Prentice & Miller, 1993). We ran computer simulations to compare these two different inference assumptions and their effect on social norm processes.

Model

Description

The model *population* consists of a fixed total of 50 agents (N = 50) that are categorized into 5 different *groups* (k = 5), each with different (drinking) preferences. The initial proportion of the different groups in the total population constitutes the *initial population social norm*, and is represented by a set of k weights $w_{initial}$. Each (drinking) group is modelled as a Gaussian distribution of which the mean represents the *group identity* (μ^i_{group}) and all groups have the same standard deviation (σ_{group}), as shown in Eq. 1. It gives the probability of different possible desires that agents in this group may have.

$$Group_i \sim \mathcal{N}(\mu^i_{group}, \sigma_{group}) \tag{1}$$

Each agent has four attributes: (i) a group identity, (ii) a true desire, (iii) a subjective social norm $w_{subjective}$, and (iv) a conformity bias c. To capture that an agent with a particular desire doesn't act exactly the same way each time, we model desire as a Gaussian distribution as well; with a unique mean μ^j_{desire} per agent and shared standard deviation σ_{desire} for all agents, as shown in Eq. 2. An agent's unique μ^j_{desire} is sampled from the group distribution $Group_i$ (given their group identity), and is sampled only once to initiate the population (❶ in Fig. 1). An agent's desire distribution represents the probability of them producing different actions (e.g., consuming x drinks) given their desire.

$$Desire_j \sim \mathcal{N}(\mu^j_{desire}, \sigma_{desire}) \tag{2}$$

Inspired by Falandays and Smaldino (2022) and Toscano and McMurray (2010), all (drinking) groups' distributions are combined and weighted by $w_{subjective}$, which together constitutes a Mixture of Gaussians distribution (MOG) representing an agent's *subjective norm perception* (*SNP*), as shown in Eq. 3. $w_{subjective}$ represents an agent's belief about the proportion of different groups in the population. An agent's subjective norm perception as an MOG is a generalized abstraction of descriptive norms. It represents the agent's perception of the probability of different μ^j_{desire} values in the population.

$$SNP \sim \sum_{i=1}^{k} w^i_{subjective}\ \mathcal{N}(\mu^i_{group}, \sigma_{group}) \tag{3}$$

Populations are connected in a social network that determines which agents can interact with each other. At each time step in our simulations, all agents first select a receiver to produce an action signal for. The agent could produce its action directly from its desire; however, in this model we assume that the tendency to conform to the subjective norm perception influences the agent's action production (❷ in Fig. 1). Therefore, the action signal is sampled from another MOG that combines the agent's desire and social norm perception, weighted by their conformity bias c, as shown in Eq. 4.

$$P(Action|Desire, SNP, c) = (1-c) * Desire + c * SNP \tag{4}$$

We compare two different strategies for inferring others' desires based on their actions: (i) a 'simple' strategy which assumes others act purely out of their own desire and (ii) a 'complex' strategy which assumes others are also influenced by the social norm, just like the agent itself. Both strategies make use of Bayesian inference with the aim of inferring the sender's $Desire_s$. Because the receiving agent doesn't have access to the sender's true group $Group_s$, we assume they use their own subjective norm perception (SNP_r) as a prior over possible desires. An agent using the simple inference strategy (red agent in ❸ in Fig. 1) infers the sender's desire ($Desire_s$) by assuming that they act completely out of their own desire with no influence of a conformity bias, as shown in Eq. 5.

$$\begin{aligned} &P_{simple}(Desire_s|Action_s) \\ &\quad \propto P(Action_s|Desire_s) * P(Desire_s|SNP_r) \end{aligned} \tag{5}$$

In contrast, an agent using the complex inference strategy (the blue agent in ❸ in Fig. 1) assumes that the sender acts out of a combination of their true desire and their subjective norm perception, mediated by their conformity bias (as is in fact the case for all agents in our simulations). This is shown in Eq. 6. The complex inference agent assumes that the sender has the same conformity bias and subjective norm perception as themselves (i.e., they model others in an egocentric way).

$$\begin{aligned} &P_{complex}(Desire_s|Action_s) \\ &\quad \propto P(Action_s|Desire_s, SNP_r, c_r) * P(Desire_s|SNP_r) \end{aligned} \tag{6}$$

After the inference step, the receiving agent selects the maximum a posteriori (MAP) desire from the posterior distribution, which represents the most likely μ_{desire} of the sender. Then, this inferred μ_{desire} is compared to the receiving agent's subjective norm perception to identify the (drinking) group that the inferred μ_{desire} most likely came from (i.e., where it has highest probability). The receiving agent then updates their subjective norm perception ($w_{subjective}$) from the previous timestep $t-1$ to the current t. This is done by increasing the weight of the group it inferred the sender came from, with update rate ϕ_{update} multiplied by the number of senders n_s^i from that group (a given receiver can receive multiple actions per timestep, because receivers are chosen probabilistically). This is followed by a normalisation step, as shown in Eq. 7.

$$w_{subjective_t} = \frac{\phi_{update} * n_s^i + w^i_{subjective_{t-1}}}{\sum_{i=1}^{k} \phi_{update} * n_s^i + w^i_{subjective_{t-1}}} \tag{7}$$

In sum, a timestep in our simulations consists of each agent selecting a receiver and producing an action (e.g., consuming x drinks), followed by inferring the desire of any sender from which they received an action, assigning that sender to a particular (drinking) group based on this inferred desire, and updating their own subjective social norm accordingly.

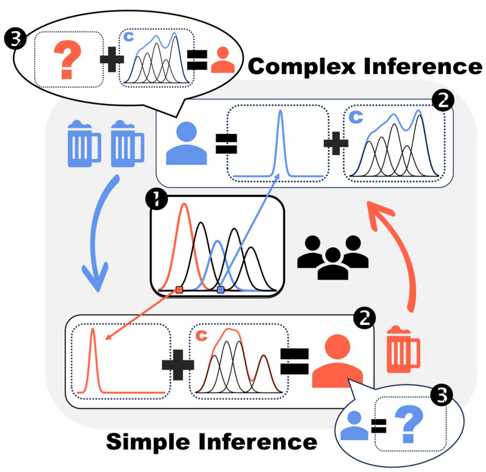

Figure 1: An illustration of agents' interaction. ❶ Initializing a population of agents from their corresponding group distributions. ❷ Both the red and the blue agent generate actions (consuming x drinks). ❸ Red and blue agents with simple and complex inference strategies interact. Note that in our simulations reported below, we only looked at homogeneous populations of agents with all the same inference strategy (so simple and complex agents don't interact with each other there).

Simulation Details[1]

All our simulated populations are organised as a *fully connected network*, in which all agents have equal opportunity to interact with each other. We compare three different strengths of the conformity bias (0.2, 0.5 and 0.8), and two inference strategies, yielding six conditions in total. Each simulation consists of a population that is homogeneous in terms of both conformity bias and inference strategy, and 5,000 timesteps.

[1]Link: `https://osf.io/y93sv/` for code and data on OSF.

Table 1: Model parameters

Parameter	Description	Value(s)
N	Population Size	50
k	Drinking Groups	5
μ_{group}	Group Means	0.1, 0.3, 0.5, 0.7, 0.9
σ_{group}	Group Standard Deviation	0.1
σ_{desire}	Desire Standard Deviation	0.05
ϕ_{update}	Update Amplitude	0.01
$w_{initial}$	Initial Population Social Norm	[0.1, 0.1, 0.1, 0.1, 0.1]
$w_{subjective}$	Initial Subjective Social Norm	[0.1, 0.1, 0.1, 0.1, 0.1]
network	Network Type	fully-connected
step	Simulation Timesteps	5,000
run	Independent Runs per Condition	20
ToM	Inference Strategies	Simple, Complex
c	Conformity Bias	0.2, 0.5, 0.8

During each timestep, every agent sends one action signal. We ran 20 independent simulations per condition. All simulations start with an evenly distributed initial group social norm (i.e., same number of agents per group). All agents start out with a uniform subjective social norm. All simulations are initialized by sampling the μ_{desire} from each agent's group distribution. See Table 1 for all parameter settings.

Outcome Measures

As agents generate actions at each timestep and make inferences to update their subjective norm perception, measurements could be constructed along three dimensions, as shown in Fig. 2a. However, one of these axes is pre-defined in this model: the individual and collective objective beliefs, which both remain constant. In this paper, we focus on two measures: (i) *individual subjective belief* (the updated subjective norm perception of each agent), and (ii) *collective subjective belief* (constructed by aggregating the group with the highest weight in agents' individual norm perceptions across all agents). The behavior axis will not be covered in this paper.

Inference Accuracy To evaluate the accuracy of inferences made about senders' desires and their respective (drinking) groups, we compare each inference made given an action signal to the sender's true group. We compute inference accuracy as the proportion of correct inferences out of the total number of inferences made, for each simulation run.

Individual and collective subjective belief Both individual and collective subjective belief are measured in Shannon Entropy, which quantifies the uncertainty in a probability distribution. When everything is equally possible, this yields maximum uncertainty and therefore maximum entropy. Thus, the uniformly distributed initial population social norm (collective *objective* belief; not shown in our figures) and the initial individual subjective belief have maximum entropy. In our case with five (drinking) groups, the highest possible entropy is 1.61 nats. In sum, both measures stand for agents' beliefs about the diversity of (drinking) norms in the population; either from the individual or from the collective perspective.

As entropy decreases from the maximum, it signifies that agents begin to deviate from the "everything is equally pos-

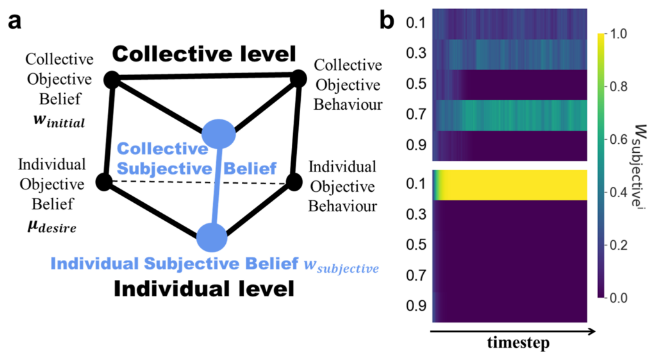

Figure 2: Measure Dimensions and Heatmaps **a**) Three measure dimensions, with each node representing a specific measure (with the ones we focus on in this paper in blue). **b**) Individual subjective norm perception as it changes over time. The upper heatmap shows an agent who believes that the social norm is diverse (that many different desires are possible). The lower heatmap shows an agent who believes in a dominant norm: that all other agents belong to group 0.1.

sible" view and start favouring certain (drinking) groups as more probable (upper heatmap in Fig. 2b). At the extreme, agents may end up believing that one (drinking) group dominates the whole population with a weight of 1, resulting in entropy approaching 0 (lower heatmap in Fig. 2b).

The entropy of each agent's individual subjective belief is measured at each timestep. This yields a time series of entropy measures for all agents within a given condition, which we average over agents and independent simulation runs. The entropy of the aggregate measure of collective subjective belief is also averaged over independent simulation runs.

In-Group Misalignment We use Bray–Curtis dissimilarity to assess the degree of misalignment between two agents' individual subjective belief; a measure commonly used in ecology (Bray & Curtis, 1957). It shows how different two agents' individual subjective beliefs are across all timesteps, with 1 meaning totally different and 0 meaning the same. The in-group misalignment measure is a dyadic measure and can be seen as asking how different two heatmaps (like in Fig. 2b) are, so that it quantifies the difference between these heatmaps as a whole, rather than for each timestep.

Results

In this section, the three conformity bias strengths of 0.2, 0.5 and 0.8 will be referred to as low, mid and high, respectively. For example, we call the condition where agents use the complex inference strategy and have conformity bias 0.5: 'complex-mid'. All simulation results are shown in Fig. 3.

Inference Accuracy Results The subjective norm perception is constructed by making inferences about others' desires. Intuitively, when an agent infers that another agent is drinking excessively to conform to the norm, the inference will be more accurate if the other agent is indeed doing so in order to conform. Thus, we expect the complex strategy will yield more accurate inferences, since its inference structure is consistent with others' action production process.

Fig. 3a shows that the inference accuracies of complex agents decrease nonlinearly as conformity bias increases, and are lower than those of the corresponding simple conditions (except complex-low). Simple agents' accuracies decrease linearly as conformity bias increases. Contrary to our expectation, complex agents are overall less accurate than simple agents in inferring others' desires, even though complex agents' inference model mirrors how actions are truly produced (and all agents are initiated with the correct prior).

Below, we will provide an explanation for this counterintuitive result. But for this to make sense, it helps to view the difference between the two types of agents as whether they are consistent in their conformity bias between the behavioral level (action production) and belief level (inference-making). Through this lens, we can think of simple agents as using a complex inference strategy with an inference conformity bias of 0.0, while in their behavior they actually conform to a low, mid or high degree (i.e., inconsistent). In contrast, complex agents have a low, mid or high conformity bias in their action production, and a corresponding conformity bias in their inference-making (i.e., consistent).

For simple agents, the discrepancy between their action-production and inference-making processes increases as their action conformity bias increases across conditions, while their inference conformity bias remains zero. Their inference strategies no longer explain well how actions are actually produced, leading to more errors. The linearly decreasing inference accuracy reflects this linearly growing discrepancy.

Complex agents' conformity bias in inference operates differently. A complex agent's own norm perception becomes self-reinforced, because in the inference process they (i) assume (egocentrically) that others share their own norm perception and (ii) use their own norm perception as the prior (see Eq. 6). As a result, small differences in the initial interactive experiences of complex agents are amplified by their egocentric assumptions, mediated by their conformity bias. Consequently, this self-reinforcing inference structure takes the complex agent further and further away from objective truth and causes a nonlinear decrease in their accuracy. Thus, the correct inference structure and initial prior does not necessarily result in accurate inferences. This is a result of our assumption that agents do not know each other's perception of the social norm, and therefore use their own norm perception as a proxy, in combination with our assumption that agents use their own norm perception as a prior over possible desires (which needs to be defined for Bayesian inference).

Individual Subjective Belief Results Fig. 3b shows the mean entropy of individual agents' subjective belief (i.e., how diverse they believe the population is in terms of agents' desires). We expect that higher inference accuracy will lead to agents more accurately capturing the objective social norm (which starts out as maximally diverse) in their subjective belief. In Fig. 3b, we see that all simple conditions, together

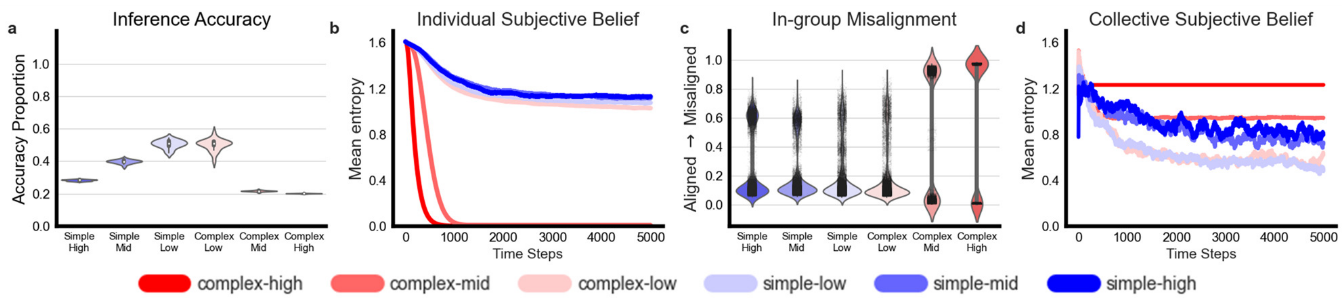

Figure 3: Simulation results across conditions (inference strategy*conformity bias strength): **a**) Inference accuracy; **b**) Mean entropy of individual subjective belief; **c**) In-group misalignment; **d**) Mean entropy of collective subjective belief.

with the complex-low condition, overlap and are somewhat below the maximum. This means that agents in these conditions believe that their population is quite diverse in their (drinking) preferences. Contrary to our expectations, the accuracy of inferences in simple conditions (which decreased linearly with increasing conformity bias) did not affect the subjective norm perception (in terms of perceived diversity).

An explanation for this is that the initial population norm (collective objective belief) is set to be maximally diverse (same number of agents per group). When agents produce actions based on these desires with a probability of $1-c$ in each round, and a simple agent performs inference over these actions (assuming that the actions reflect true desires), the inferred belief will preserve diversity (similar to the agents' initial individual subjective belief). When simple agents act out of their norm perceptions (with probability c), which are initially set to be diverse, they produce a wide variety of actions. These diverse actions are perceived as resulting from diverse desires, because the simple inference agents assume that others act only according to their own desires. As a result, simple agents make a wide range of errors in inferring the desires of others. However, these errors align with the overall diversity produced by diverse norm perceptions, once again preserving diversity. The interplay between diverse norm perception, inference errors, and desires forms a feedback loop that maintains diversity in simple agents' individual subjective beliefs.

The subjective norm diversity in both the complex-mid and complex-high condition decreases to zero (more quickly for complex-high than complex-mid). This rapid entropy decrease indicates that agents in these conditions quickly start believing that one norm is dominant (i.e., that all other agents belong to this group). The fact that this happens more quickly when the conformity bias is high, is a result of the self-reinforcing process for complex agents described above, which gets stronger as the conformity bias increases. This explains why complex-mid and complex-high populations end up believing in one dominant norm, while complex-low populations do not. In sum, inference accuracy has an effect on the diversity of individual subjective norm perceptions only in complex conditions, but not in simple conditions.

In-group Misalignment Results For the individual subjective belief results, we looked at an average over the entropy of subjective norm perceptions. This operation will inevitably lose information as we compress a full distribution to one value, making some different subjective norm perceptions indistinguishable. For example, [0, 0.2, 0, 0, 0.8] and [0, 0.8, 0, 0, 0.2] have the same entropy, but they stand for very different beliefs about the social norm. Therefore, we use in-group misalignment to measure how different agents' beliefs about norms are. Intuitively, we expect that an increase in conformity bias would lead to a decrease in misalignment.

Contrary to our expectations, Fig. 3c shows that a stronger conformity bias leads to an increase in misalignment, especially in the complex conditions. Both the mid- and high-complex conditions look polarized and have much more misalignment than the other conditions. All simple inference conditions show lower levels of in-group misalignment than their complex counterparts. This happens because agents of both types conform to their own subjective norm perception (more strongly so as the conformity bias increases), which means they do not conform to 'the objective truth'. So, the higher the conformity bias, the more unique agents' subjective beliefs become, leading to more misalignment. In complex agents, the egocentric assumptions in their inference process reinforce their unique perspective even more (modulated by the conformity bias), leading to even more misalignment.

Collective Subjective Belief Results Before delving into the results, we should make clear that this measure is the result of aggregation and averaging. Our collective subjective belief measure does not imply "emergence" in the complexity science sense, but is simply an aggregated measure that gives us insight into simulated activity over time.

Fig. 3d shows that the mean entropy of the collective subjective belief increases with conformity bias in all complex conditions. The simple-low condition overlaps with the complex-low condition, and in the simple-mid and simple-high conditions, entropy levels are very similar, and in both cases lower than in their corresponding complex conditions. This is surprising because it is the exact opposite of what happens at the individual level: Fig. 3b and Fig. 3d show a reversal in their trends. At the individual level, simple agents retain more diversity, while at the collective level, complex agents are more diverse in their belief. Furthermore, the overlap between the different simple conditions at the individual level also disappears at the collective level (*high* = *mid* > *low*), similarly to what happens for inference accuracy.

To explain this, we have to look back at the in-group misalignment results. If all agents would align on the same individual belief about which group is the dominant one, that group would become dominant at the collective level as well, resulting in a collective subjective belief with zero entropy. However, complex agents in the mid and high conditions who perceive one norm as dominant, tend to misalign on which group is the dominant one. Thus, their subjective beliefs are more diverse at the collective than at the individual level.

In contrast, simple agents tend to favour certain groups over others in their diverse individual subjective beliefs (see upper heatmap in Fig. 2b). Most of them are more consistent (low misalignment) in the norm views they hold than in the corresponding complex conditions (except complex-low), so that what they all believe to be the relatively more or less dominant groups become even more so collectively. Thus, they have a lower diversity of subjective beliefs at the collective level than at the individual level, as well as than their corresponding complex conditions at the collective level. Moreover, even though their individual norm perceptions have the same degree of diversity (entropy overlap), simple agents disagree more about which groups are more dominant as the conformity bias increases (increased misalignment). As a result, their collective subjective beliefs no longer overlap.

Summary Our simulations show that increased behavioral conformity can induce misjudgments about the true desires of others, and misalignments of norm perceptions among agents. This is strongest in complex inference agents, even though their model of how others produce actions is correct. Due to their egocentric bias, their individually constructed norm perceptions deviate greatly from the objective belief. However, collectively they are more in line with the objective belief. Interestingly, an increase in misjudgments of desires by simple inference agents does not necessarily lead to large misperceptions of the norm; but their misalignment does lead them to collectively deviate from the objective belief.

Discussion

In this paper, we formalized a process-based conceptualization of descriptive social norms, in the form of an agent-based cognitive model. We modelled descriptive norms, which are considered trivial in norm psychology (in contrast to prescriptive norms) because they are thought to be acquired easily through direct observation of statistical regularities in actions, to lack a "sense of should" and to be "normal rather than normative" (Kelly & Davis, 2018; Zhi-Xuan & Ong, 2019; Theriault, Young, & Barrett, 2021). If so, then the phenomenon of *pluralistic ignorance*, which we take as a case study, cannot be fully explained, since it is not prescriptive and cannot be accounted for by looking only at regularities in behavior. Some behaviors are regular because they are common sense, like drinking more water when the weather is hot, but that does not make them social norms. Descriptive norms are in the awkward position of being neither fully attributable to statistical regularities nor to rules. They appear to be statistical regularities; however, simply conceptualizing and studying descriptive norms as such is not sufficient for explaining them. Meanwhile, it seems that we cannot completely discard the rule-like flavor of descriptive norms, because what is behaviorally common is sometimes perceived as normative in a prescriptive way, or becomes prescriptive in the course of development and interaction (Heyes, 2022; Knobe, 2023).

Therefore, we choose to place interaction at the center of social norms, conceptualizing them as processes in which individuals interact with each other (Dingemanse et al., 2023). This conceptualization is not mutually exclusive with a rule-based account. On the contrary, it is a more inclusive approach that allows us to see descriptive norms — prescriptive norms as a continuum in which neither rules nor regularities are the focus, but rather the process of the various interactions themselves. Through this continuum perspective, we could identify under what circumstances descriptive norms arise, and when approval and punishment may come in, transforming a descriptive into a prescriptive norm. In line with interactionism, agents in our model construct their own subjective perception of norms as their social realities through interaction, which influences how they act and how they presume others will act (Mirski & Bickhard, 2021).

The constructive norm process is a multi-dimensional construct across the individual/collective, belief/behavior, and subjective/objective levels. Our simulation results reveal intricate interactions that have to do with: (i) Individual/Collective: individually vs. collectively believed subjective norms, and the misalignment between agents' individual norm perception within a population; (ii) Behavior/Belief: conformity bias in one's own actions vs. the assumed conformity bias of others in inference-making, and the misalignment between the two. Most of our results describe steady states, but they provide us with insight into the process.

The model presented here also has several limitations. First, we assume that agents have fixed desires. However, desires can be influenced and altered based on an agent's individual subjective norm perception, as modelled in Zhi-Xuan and Ong (2019). Second, we also assume five (drinking) groups as pre-given with fixed boundaries. However, groups and categories might emerge themselves in interactions and might have different boundaries and sizes, like the category attractors simulated in Falandays and Smaldino (2022). Finally, there are several factors we have not explored in the current study; such as social network structure, means of communication, and memory capacity of the agents.

In conclusion, conceptualizing and modelling descriptive social norms as socio-interactive processes with conformity at their essence reveals complex dynamics across multiple dimensions. Studying descriptive norms in this way contributes to a more comprehensive understanding of social norms.

Acknowledgments

We would like to thank Olivia Guest, Iris van Rooij, Todd Wareham, Mark Blokpoel, Laura van de Braak, and all other members of the Computational Cognitive Science research group at Radboud University for their invaluable feedback at several stages of this project. We would also like to thank Fred Hasselman for his feedback on the master thesis by Yue Chen that this paper is based on. Finally, we would like to thank the anonymous reviewers and meta-reviewer for their helpful and insightful comments. MW was supported by the Dutch Research Council (NWO) under Gravitation Grant 024.001.006 to the Language in Interaction Consortium (postdoc position within Big Question 5).

References

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Binmore, K. (2001). How and why did fairness norms evolve? In *The origin of human social institutions* (pp. 149–170).

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern wisconsin. *Ecological monographs*, *27*(4), 326–349.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, *58*(6), 1015.

Dale, R., Dietrich, E., & Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cognitive Science*, *33*(5), 739–742. doi: 10.1111/j.1551-6709.2009.01042.x

Dingemanse, M., Liesenfeld, A., Rasenberg, M., Albert, S., Ameka, F. K., Birhane, A., ... others (2023). Beyond single-mindedness: A figure-ground reversal for the cognitive sciences. *Cognitive science*, *47*(1), e13230.

Falandays, J. B., & Smaldino, P. E. (2022). The emergence of cultural attractors: How dynamic populations of learners achieve collective cognitive alignment. *Cognitive science*, *46*(8), e13183.

Hawkins, R. X., Goodman, N. D., & Goldstone, R. L. (2019). The emergence of social norms and conventions. *Trends in Cognitive Sciences*, *23*(2). doi: 10.1016/j.tics.2018.11.003

Heyes, C. (2022). Rethinking norm psychology. *Perspectives on Psychological Science*, 17456916221112075.

Kelly, D., & Davis, T. (2018). Social norms and human normative psychology. *Social Philosophy and Policy*, *35*(1), 54–76.

Knobe, J. (2023). Capacity for social norms, or statistical and prescriptive hybrid? *Perspectives on Psychological Science*, 17456916231187400.

Legros, S., & Cislaghi, B. (2020). Mapping the social-norms literature: An overview of reviews. *Perspectives on Psychological Science*, *15*(1), 62–80.

Mackie, G. (2018). Social norms of coordination and cooperation. *Social Philosophy and Policy*, *35*(1), 77–100. doi: 10.1017/s0265052518000109

Miller, D. T. (2023). A century of pluralistic ignorance: what we have learned about its origins, forms, and consequences. *Frontiers in Social Psychology*, *1*, 1260896.

Mirski, R., & Bickhard, M. H. (2021). Conventional minds: An interactivist perspective on social cognition and its enculturation. *New Ideas in Psychology*, *62*, 100856.

Morris, M. W., Hong, Y.-y., Chiu, C.-y., & Liu, Z. (2015). Normology: Integrating insights about social norms to understand cultural dynamics. *Organizational Behavior and Human Decision Processes*, *129*, 1–13. doi: 10.1016/j.obhdp.2015.03.001

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. doi: 10.1017/s0140525x00076512

Prentice, D. A., & Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *Journal of Personality and Social Psychology*, *64*(2), 243–256. doi: 10.1037/0022-3514.64.2.243

Smith, E. R., & Conrey, F. R. (2007). Agent-based modeling: A new approach for theory building in social psychology. *Personality and social psychology review*, *11*(1), 87–104.

Sripada, C. S., & Stich, S. (2006, 01). A Framework for the Psychology of Norms. In . S. S. P. Carruthers S. Laurence (Ed.), *The Innate Mind: Volume 2: Culture and Cognition* (pp. 280–301). Oxford University Press.

Sun, Z., Lorscheid, I., Millington, J. D., Lauf, S., Magliocca, N. R., Groeneveld, J., ... Buchmann, C. M. (2016). Simple or complicated agent-based models? a complicated issue. *Environmental Modelling & Software*, *86*, 56–67. doi: https://doi.org/10.1016/j.envsoft.2016.09.006

Theriault, J. E., Young, L., & Barrett, L. F. (2021). The sense of should: A biologically-based framework for modeling social pressure. *Physics of Life Reviews*, *36*, 100–136.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive science*, *34*(3), 434–464.

Westra, E., & Andrews, K. (2022). A pluralistic framework for the psychology of norms. *Biology & Philosophy*, *37*(5), 40. doi: 10.1007/s10539-022-09871-0

Zhi-Xuan, T., & Ong, D. C. (2019). Bayesian inference of social norms as shared constraints on behavior. In *Proceedings of the 41st annual meeting of the cognitive science society* (pp. 2919–2925).