Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Psychology

Review

Computational modeling of social decision-making

Sarah Vahed[1,a], Elijah P. Galván[1,a] and Alan G. Sanfey[1,2]

Social decision-making is guided by a complex set of social norms. Computational modeling can play a significant role in enriching our understanding of these norms and how precisely they direct social choices. Here, we highlight three major advantages to using computational modeling, particularly models derived from Utility Theory, in the study of social norms. We illustrate how such models can help generate detailed processes of decision-making, enforce theoretical precision by delineating abstract concepts, and unpack when, and why, people adhere to specific social norms. For each benefit, we discuss a recent study which has employed modeling in the service of assessing the role of norms in decision-making, collectively revealing how computational modeling enables better prediction, description, and explanation of important social choices.

Addresses
[1] Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, the Netherlands
[2] Behavioural Science Institute, Radboud University, Nijmegen, the Netherlands

Corresponding author: Vahed, Sarah (sarah.vahed@donders.ru.nl)
[a] These authors contributed equally to this work.

Keywords
Social decision-making, Computational Modeling, Social norms, Utility models.

Introduction

Many of the most important choices we make occur in the context of highly complex social environments. These social decisions, which by definition affect others as well as ourselves, are typically guided by social norms, which prescribe actions that are perceived as appropriate to take or to avoid in a given situation [1]. However, ambiguity exists and inconsistencies remain as to how exactly we should define norms in these decision contexts, and how precisely they might affect social choice [2]. In addition, the interplay of different motives that may underlie the use of particular norms, as well as potential individual differences in norm compliance, are obstacles in more fully understanding social decisions and their underlying processes [3]. We believe that computational modeling has an important role to play in moving the needle on social norm research, through robust and replicable theory-driven steps which can address these aforementioned gaps in knowledge.

Computational modeling entails translating abstract psychological theories into precise formal frameworks [4,5]. These models are typically operationalized as mathematical equations, wherein each term or parameter reflects a crucial facet of the particular decision-making process under investigation. The resulting model is thus a formalized hypothesis about how choices are made, which can be tested against real, observed, behaviour, in turn quantifying how accurately the model's predictions represent the actual underlying decision-making process [5]. Moreover, by comparing different models, it is possible to test various potential explanations about the role of social norms in general, and specific norms in particular, in directing the choices people make in social contexts, facilitating both hypothesis testing and theory building.

Various families of computational models have been developed to capture aspects of the decision-making process, for example models of learning and valuation (reviewed in Refs. [4,6]). Of immediate relevance, the family of Utility models, grounded in economic theory, attempts to describe how people experience subjective value when choosing between options [7]. In the study of social norms, conventional computational approaches include agent-based modeling in which simulations of social interactions advance understanding of how norms emerge, change, and influence decision-making across multiple agents [8,9]. In contrast, Utility models present an opportunity to formalize the process by which people incorporate payoff differences between themselves and others, helping to unpack how specific social norms are integrated into the individual decision-making process. In the following section, we highlight three major advantages in using this type of computational modeling to enhance our understanding of social choice. Each benefit is illustrated with reference to a recent study which has employed Utility models in the

Current Opinion in Psychology 2024, **60:**101884

This review comes from a themed issue on **Norm Change (2024)**

Edited by **Cristina Bicchieri**, **Michele Gelfand**, and **Giulia Andrighetto**

For complete overview about the section, refer Generation COVID: Coming of Age Amid the Pandemic (2024)

Available online 4 September 2024

https://doi.org/10.1016/j.copsyc.2024.101884

2352-250X/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

service of assessing the role of norms in driving social decisions, thereby advancing norm theory. Overall, we believe these insights broadly illuminate the value of a formal modeling approach in studying complex social decision-making.

From conceptual to trial-specific predictions

One advantage of utilizing computational modeling to study the role of social norms in decision-making is the ability to bridge the gap from conceptual ideas to quantitative predictions: models make *a priori* predictions about how and why people decide, which can be directly validated with data.

A recent study by Gao and colleagues [10] usefully demonstrates this advance. Here, the authors sought to disentangle the different social norms which dictate how people respond to help when it is offered. Specifically, they aimed to distinguish the role of two norms: indebtedness and reciprocal altruism. Indebtedness represents a perceived obligation to repay others and is marked by negative affect, while reciprocal altruism reflects a desire to repay others due to prosocial motivations and as such is associated with positive affect. The authors proposed a conceptual model of how people respond to offers of help, which was implemented as a computational model by using mathematical expressions to represent when, and how, these different social norms factor into the decision-making process.

In a novel decision-making task, participants were subjected to electric shocks which an anonymous benefactor could pay to reduce, thereby helping the participant. Sometimes, the benefactor had a false-belief about the participants' ability to repay them in turn, thereby eliminating participants' feelings of indebtedness. Participants had two decisions to make: one, to either accept or reject the benefactor's help and, then subsequently, whether to repay the benefactor. A conventional approach to analysis would have shown that when help is given without expectations of repayment, there is a significant increase in the likelihood of this help being accepted and a decrease in the amount repaid, indicating that indebtedness plays a role in these decisions rather than altruism. However, by translating their conceptual model into a computational model, Gao and colleagues [10] instead found compelling evidence that indebtedness, altruism, and self-interest in fact all play meaningful roles in guiding help-acceptance and repayment decisions. Specifically, the authors were able to generate specific, quantitative predictions about these decisions for each trial. This study thus illustrates the potential for computational models to generate rigorous, falsifiable, trial-specific predictions which help develop and validate conceptual ideas about how norms factor into the decisions people make.

Differentiating related underlying motives

A second advantage of computational modeling is that it makes it possible to test diverse psychological motives which may simultaneously impact social decision-making. A recent study by Li and colleagues [11] showcases this in their investigation of decisions concerning distributive justice, that is, the allocation of resources across people. Decisions in this context have been associated with different motives. Specifically, while it is well-known that fairness principles play a role in choices about sharing wealth [12], it has also been shown that people generally tend to avoid causing harm to others [13], and are concerned about changing existing rank or status structures [14].

In order to investigate the conflicting motives of equality-seeking, harm-mitigation, and hierarchy-maintenance, and specifically how these factors may concomitantly impact resource distribution, the authors utilized a new redistribution paradigm and developed four computational models to formalize different weightings of these motives. Importantly, the experiment was designed to ensure that each of the aforementioned motivations could be distinguished by patterns of choice. The models themselves ranged from a simple model focused solely on inequality aversion to three more complex versions considering the combinations of all motives, and therefore this study sought to evaluate how well each model embodied the underlying process guiding third-party resource allocation.

Testing the models' predictions against participants' actual choices, and comparing each model's performance, the results revealed that redistributive decisions in this context are best explained by accounting for the concurrent use of all three prosocial motives in guiding choice [11]. The findings thus highlight how individuals, during wealth redistribution, consider their aversion for inequality, harm and rank reversal. Computational modeling can thus extend and refine influential theories of, for example, fairness norms, by clarifying the role of distinct social motivations which modulate individuals' distributive decisions. In this way, the findings illustrate an important advantage of computational modeling over conventional approaches: it makes it possible to test multiple motives which may concurrently impact choice. This approach can therefore yield deeper insights into the intricate interplay of social norms which shape individual choices in social contexts.

Identifying individual differences

A third advantage of utilizing computational models to study social norms is that they facilitate a focus away from group-level tendencies and, instead, move towards identifying interindividual differences in the application of different norms.

An experimental study conducted by van Baar and colleagues [15] aimed for greater insight into why people reciprocate trust, even when this may often not be in one's own best financial interest. For example, when faced with a situation where someone has placed their trust in us and we are given the opportunity to exploit that person for financial gain, most people choose to reciprocate trust, even towards a stranger. The authors identified two potential norms that could lead people to reciprocate. The first is the principle of inequality-aversion, as failing to reciprocate trust would result in an overall unequal distribution of resources. The second norm is guilt-aversion: failing to reciprocate violates the expectations of those who have originally trusted us, which typically results in feelings of guilt. Through a modeling approach, the authors assessed the role of both social norms in the decision-making process. Thus, rather than seeking to detect which role exerts a stronger influence on average, the approach enabled the authors to identify, at an individual level, which people rely on which norm.

Since people typically expect others to create equality in distributions, van Baar and colleagues [15] developed an experimental task in which the participant knows that their game partner, who has placed trust in them, has a false belief about the ability of the participant to reciprocate the trust. Thus, the participant is forced to choose between either meeting the partner's expectations or ensuring equality between themselves and their partner: they cannot do both. A conventional approach to analyzing the participants' reciprocity behavior would show that inequality-aversion, as opposed to guilt-aversion, motivates reciprocity at a group level. However, computational modeling demonstrated that only 40% of participants were inequality-averse, while 10% were exclusively motivated by guilt-aversion and a further 10% were motivated by greed. Importantly, 40% of participants switched between inequality-aversion and guilt-aversion across the experiment as a function of which social norm was 'cheaper', that is, what was to their greater financial benefit. Capturing each individual's specific motivation for reciprocity demonstrates that different people use different social norms in this task, offering a more nuanced, accurate, and insightful characterization of the data. In a broader sense, this study illustrates one relative advantage of computational modeling over traditional, more descriptive approaches: it treats individual differences as meaningful, rather than merely noise to be regressed out in order to identify an average, group-level, effect.

Conclusion

Computational models have significant potential to advance our understanding of the important role which social norms play in decision-making processes. The benefits of using models as tools for robust hypothesis generation and theory development are clearly established [5]. Here, we extend this by specifically highlighting how this approach can also generate insights into questions otherwise challenging to address via traditional means of investigation. They can help to identify the critical norms which impact social choice, clarify how diverse norms interact, and determine how individuals employ norms differently in their decision-making. This is the value we see in the use of computational models broadly, and Utility models in particular, in short, not only to better describe and predict distinct social choices, but to also better *explain* them.

The possible research directions and application of this approach are broad. As just one example, computational models can create the foundation for critical tests of the plausible biological bases of distinct motives underlying norms, including by investigating the relationship between model parameters and patterns of neural activity [10,11,15]. In addition, detailed and comprehensive conceptual models can be combined and validated with other useful response types, including large-scale self-report data and questionnaires [10]. Finally, while we have specifically focused here on examples from models derived from Utility Theory, other types of computational models can yield useful information about the role of learning, uncertainty, or attention in decision-making, allowing for greater understanding of the impact of norms in social behavior [4].

While we believe such computational models offer clear benefits, they also require care in their use and interpretation. Models necessitate both rigorous and exhaustive testing, with careful evaluation of aspects such as model fit, model comparison, and assessment of observed variances. Additionally, given the vast number of model variants which could be potentially employed for a given question of interest, it is crucial for researchers to be mindful of the models they construct and seek to compare. It is important to recognize that each model remains an approximation and, as such, is inherently limited in capturing all nuances of complex decision behaviour. However, when developed with thought and consideration, models can indeed provide important insights for the study of social behaviour. In this regard, computational models enhance our understanding of abstract concepts regarding norms by providing process models of complex decision-making, enforcing greater theoretical precision, and unpacking when, and why, certain people adhere to particular social norms.

Credit author statement

Sarah Vahed: Conceptualization, Writing - Original draft preparation, Writing - Reviewing and Editing; **Elijah P. Galván:** Conceptualization, Writing - Original draft preparation, Writing - Reviewing and Editing; **Alan G.**

Sanfey: Conceptualization; Writing - Reviewing and Editing, Funding acquisition.

Funding
Preparation of this manuscript was supported by a Dutch Science Foundation (NWO) Grant 406.21.GO.015, to A.G.S.

Declaration of competing interest
None of the authors declares any conflicts of interest.

Data availability
No data was used for the research described in the article.

References
References of particular interest have been highlighted as:

* of special interest

1. Rilling JK, Sanfey AG: **The neuroscience of social decision-making**. *Annu Rev Psychol* 2011, **62**:23–48.
2. Legros S, Cislaghi B: **Mapping the social-norms literature: an overview of reviews**. *Perspect Psychol Sci* 2020, **15**:62–80.
3. Gavrilets S, Tverskoi D, Sánchez A: **Modelling social norms: an integration of the norm-utility approach with beliefs dynamics**. *Philos Trans R Soc B* 2024, **379**, 20230027.
4. * Cushman F: **Computational social psychology**. *Annu Rev Psychol* 2024, **75**:625–652.
5. Guest O, Martin AE: **How computational modeling can force theory building in psychological science**. *Perspect Psychol Sci* 2021, **16**:789–802.
6. Konovalov A, Hu J, Ruff CC: **Neurocomputational approaches to social behavior**. *Curr Opin Psychol* 2018, **24**:41–47.
7. Loewenstein GF, Thompson L, Bazerman MH: **Social utility and decision making in interpersonal contexts**. *J Pers Soc Psychol* 1989, **57**:426.
8. Jackson JC, Rand D, Lewis K, Norton MI, Gray K: **Agent-based modeling: a guide for social psychologists, Soc. Psychol. Personal**. *Sci* 2017, **8**:387–395.
9. Conte R, Andrighetto G, Campennl M: *Minding norms: mechanisms and dynamics of social order in agent societies*. USA: Oxford University Press; 2014.
10. * Gao X, Jolly E, Yu H, Liu H, Zhou X, Chang LJ: **The psychological, computational, and neural foundations of indebtedness**. *Nat Commun* 2024, **15**:68.
11. * Li Y, Hu J, Ruff CC, Zhou X: **Neurocomputational evidence that conflicting prosocial motives guide distributive justice**. *Proc Natl Acad Sci USA* 2022, **119**, e2209078119.
12. Tricomi E, Rangel A, Camerer CF, O'Doherty JP: **Neural evidence for inequality-averse social preferences**. *Nature* 2010, **463**:1089–1091.
13. Crockett MJ, Kurth-Nelson Z, Siegel JZ, Dayan P, Dolan RJ: **Harm to others outweighs harm to self in moral decision making**. *Proc Natl Acad Sci USA* 2014, **111**:17320–17325.
14. Xie W, Ho B, Meier S, Zhou X: **Rank reversal aversion inhibits redistribution across societies**. *Nat Human Behav* 2017, **1**. 0142.
15. van Baar JM, Chang LJ, Sanfey AG: **The computational and neural substrates of moral strategies in social decision-making**. *Nat Commun* 2019, **10**:1483.

Further information on references of particular interest

4. * This review highlights progress in employing various formal models within social psychology, This review highlights progress in employing various formal models within social psychology, demonstrating the potential of computational models to enhance our understanding of processes such as inference, attribution, value-based learning, decision-making, and multi-agent choice.

10. * This study employs a multifaceted approach, integrating a large-scale online questionnaire, an interpersonal game, computational modeling, and neuroimaging, to validate a conceptual model of indebtedness.

11. * This study presents behavioral, computational, and neuroimaging evidence on distributive decisions, offering a comprehensive framework for understanding the cognitive and biological mechanisms coordinating multiple prosocial motives in the brain which guide redistribution behaviors.