

Proximate and ultimate drivers of norms and norm change

Ingela Alger, Sergey Gavrilets and Patrick Durkee



Proximate and ultimate drivers of norms and norm change

Ingela Alger^{*}, Sergey Gavrilets[†], and Patrick Durkee[‡]

March 2024

Abstract

We describe a formal model of norm psychology that can be applied to better understand norm change. The model integrates several proximate drivers of normative behavior: beliefs and preferences about a) material payoffs, b) personal norms, c) peer disapproval, d) conformity, and e) authority compliance. Additionally, we review interdisciplinary research on ultimate foundations of these proximate drivers of normative behavior. Finally, we discuss opportunities for integration between the proposed formal framework and several psychological sub-fields.

Keywords: social norms, preferences, beliefs, evolutionary foundations, ultimate drivers, proximate drivers, interdisciplinary research.

1 Introduction

To understand norm change, we must understand norm psychology [1]. We describe a formal model of proximate factors underpinning normative behavior. Additionally, we summarize theoretical research from evolutionary anthropology and biology, and economics, that offers insights into ultimate explanations for these proximate factors.

^{*}Toulouse School of Economics, CNRS, University of Toulouse Capitole, and Institute for Advanced Study in Toulouse, Toulouse, France, and CEPR. ingela.alger@tse-fr.eu

[†]University of Knoxville, Tennessee, USA. gavrila@utk.edu

[‡]California State University, Fresno, USA. pdurkee@csufresno.edu

Finally, we explore opportunities for psychologists to incorporate, and contribute to, the proposed interdisciplinary framework.

2 Conceptual framework

The framework posits – in line with economists' view – that individuals choose behavior based on their preferences, given their beliefs about the situation they face. Many drivers of normative behavior and change have been identified [2, 3, 4, 5, 6, 7]. Accordingly, several alternative preferences are proposed in the literature, each focusing on a subset of factors [8]. A full understanding of norm psychology may, however, require gathering all the factors into one single framework. In the framework we adopt here, an individual's action may be simultaneously influenced by material payoffs, a desire to act in accordance with one's personal norm, a desire to avoid peer disapproval, a desire to conform with peers' actions, and a desire to comply with actions prescribed by authorities (educational, cultural, religious, political, or administrative). This combination of motivations together define the individual's preferences, which can be formalized by the following *utility function* [9]:

$$u = \underbrace{\pi(x, \tilde{x})}_{\text{material payoff cognitive dissonance disapproval by peers}}_{A_3(x - \tilde{x})^2} - \underbrace{A_1(x - y)^2}_{A_2(x - \tilde{y})^2} - \underbrace{A_2(x - \tilde{y})^2}_{A_3(x - \tilde{x})^2}, \qquad (1)$$

where:

- x is the individual's action and \tilde{x} his *empirical expectation* (the action he believes that peers will adopt): these actions determine his expected *material payoff*;
- y is the individual's *personal norm* (the action he views as "the right thing to do" absent material well-being considerations and influence from peers or authorities): cognitive dissonance arises if the preference weight A_1 exceeds 0 and the action x differs from the personal norm y;
- \tilde{y} is the *normative expectation* (the action that the individual believes that peers

think he should do): discomfort from *peer disapproval* arises if the preference weight A_2 exceeds 0 and x differs from \tilde{y} ;

- the individual suffers if the preference weight A_3 exceeds 0 and their action does not conform with that of their peers;
- the individual suffers if the preference weight A_4 exceeds 0 and their action differs from G, the action prescribed by an external authority.

This framework provides a formal description of associated psychological mechanisms at multiple levels of analysis [10, 11]. That is, each component of the preferences above demarcates a particular function: (i) maximizing material well-being, (ii) respecting one's personal norm, (iii) avoiding peer disapproval, (iv) conforming with others' behavior, and (v) obeying authorities. Gathering them into one mathematical function further shows how each purpose may be carried out algorithmically. Specifically, an individual decides on "the right thing to do" (y), and forms beliefs about others' behaviors (\tilde{x}) and others' expectations (\tilde{y}) , using the available information. Among the feasible behaviors, they then select the behavior which best reconciles the multiple, possibly conflicting, purposes (i.e., the x that maximizes the value of the function).

The values of the preference weights A_1 , A_2 , A_3 , and A_4 define the extent to which each factor matters for the individual. The preference weights may differ across individuals, but they are taken to be constant over time for each individual. By contrast, the personal norm and the beliefs may change as new information appears. For example, the personal norm may change from "why should I bother washing my hands?" to "I should wash my hands" upon learning about the social benefits of hand hygiene. Empirical and normative expectations depend on observations of others' behaviors and discourse. With the preferences described by the above utility function, norm change can thus result from new information becoming available: such new information can generate changes in an individual's personal norm, triggering a behavioral change, which in turn affects others' behaviors through the peer effects, and so on, until a new social norm (the common and commonly accepted behavior) emerges. Evolutionary logic provides insights on the ultimate drivers of the personal norm (y), our susceptibility to social influence $(\tilde{x}, \tilde{y}, \text{ and } G)$, and their associated preference weights A_1 , A_2 , A_3 , and A_4 .

3 Evolutionary foundations

3.1 Personal norms

Theoretical research shows that evolutionary forces may shape a personal norm, and the weight attached to deviations from it (A_1) . Humans should have developed a propensity to evaluate actions in the light of what their reproductive success would be if these actions were also adopted by others, and to care about deviations from this norm [12, 13]. This universalization process is reminiscent of Kantian ethics ("live your life as though every act were to become a universal law"), anchored in reproductive success. Personal norms such as "You shall not murder" or "You shall not steal" may thus have evolved a long time ago.

Evolutionary models show that the ultimate driver of such universalization ethics is the propensity for individuals with a recent common ancestor to interact [14, 15]. This propensity would have resulted from patterns of human group formation, and intergroup migration and interactions. It is viewed as forming part of the "environment of human evolutionary adaptation" [16].

Importantly, if fitness is correlated with material well-being, evolution should have led to such universalization ethics not only with reproductive success as criterion, but also with material well-being as criterion [14]. Moreover, these models encompass both biological and cultural transmission of traits.

These findings thus suggest that in the function above: (a) y would be the action that would maximize the material payoff, if that action was universalized; and (b) A_1 should be positive.

3.2 Susceptibility to social influence

The last three components in the function u specify distinct forms of social influence on decision-making. Substantial efforts have been dedicated to unraveling the evolutionary roots of susceptibility to such influences.

The current consensus is that our ancestors embarked on an evolutionary path marked by an increased reliance on cooperation [17]. In particular, the unstable environment of the Pleistocene should have played a crucial role in shaping the significance of culture and social learning as engines of cooperation in human evolution. Fluctuating conditions would have favored social learning over both genetic adaptation and individual learning, thanks to the induced flexibility in responding to environmental changes.

Early models showed the evolutionary viability of social learning by way of copying behaviors that are common (conformist bias), or exhibited by more successful (payoff bias) or prestigious (prestige bias) individuals [18, 19]. This suggests that when forming their normative expectations (\tilde{y}), and when comparing their behavior to that of peers, individuals should be expected to pay more attention to more successful or prestigious individuals. Recent work extends these findings while offering interesting nuances [20, 21, 22].

Self-domestication is also believed to have played a key role in human evolution [23]. This process, potentially driven by partner choice or the elimination of aggressive individuals, would have led to increased compliance with group norms [24, 25]. Moreover, humans have an extended childhood and significant involvement of non-parents in child-rearing. These factors likely fostered docility in children and, later, conformity in adults, i.e., positive preference weights A_3 and A_4 [26, 27].

Apart from social learning based on observation, humans have also developed intentional teaching, a form of learning that would have facilitated the rapid spread of innovations; those adopting the most beneficial innovations would have been more fit. In a model of the co-evolution of vertical social learning (i.e. from the previous generation) and horizontal social learning (i.e. from the current generation) based on prestige or conformity, the evolutionarily favored type of social learning was found to depend on how environmental conditions change over the course of an individual's lifetime [28].

Furthermore, self-inflicted punishment for deviating from normative expectations (\tilde{y}) can evolve [29]. Arguably, this can be interpreted as caring about peer disapproval, i.e., to imply a positive preference weight A_2 . Relatedly, norms and punishment of norm violators can co-evolve [30], and populations may include both over-socialized and under-socialized individuals, where the former are willing to make large material sacrifices, while the latter are unwilling, to adopt the desired behavior and to punish norm violators.

Overall, these studies explain how different forms of susceptibility to social influence could have become deeply ingrained in human nature. They suggest that when forming their beliefs, humans may be expected to pay attention to a variety of informational cues, including the behavior of peers and authorities, as postulated in the conceptual framework described above.

4 Opportunities for further integration into psychological research

The formalization of norm psychology described above can help psychologists move beyond verbal theories to develop precise predictions about normative behavior, identify potential intervention targets for behavioral change, and measure beliefs and preferences. For example, recent experimental designs detect the use of universalization ethics in social dilemmas and enable distinguishing between pro-social concerns and universalization ethics [31, 32, 33]. Further, experimental evidence shows that language as well as charisma of authorities matter for norm perception and norm compliance [34, 35]. Given their extensive training in experimental methods and measurement validation, psychologists are well-positioned to contribute to developing methods to test predictions of formal models and precisely measure theoretical parameters.

Different subfields of psychology may be differentially interested in specific parameters of the utility function u. For example, the beliefs (\tilde{x} and \tilde{y}) can be conceptualized as capturing an individual's "construal" of the situation, which is a central focus of social psychology [36]. Additionally, if between-person variation in preference weights (i.e., A_1 , A_2 , A_3 , and A_4) is stable across time, they may be especially relevant to personality psychologists because they dictate the extent to which different forms of normative influence drive an individual's behavior, and thus can generate betweenperson differences in behavior even when beliefs are identical. A recent replication of the famous Asch conformity experiment offers indirect evidence that varying parameters of the utility function underpinning (non-)normative behavior (e.g., material payoffs) and personality traits such as openness, influence conformity rates [37]. Future research could measure individual differences in belief parameters and preference weights more directly to examine their stability across time and situations.

Psychological accounts of social emotions may also complement the above conceptual framework. Theoretical and empirical analyses suggest that social emotions, such as pride, shame, guilt, gratitude, and anger are strongly linked to social value [38, 39, 40]. To the extent that the costs of nonconformity and peer disapproval are realized as social value changes and associated problems (e.g., ostracism, reduced access to resources), the utility function components may describe parts of the computational architecture of social emotions. For example, the discrepancy between a behavior (x)and normative expectations (\tilde{y}) may be experienced as feelings of prospective shame or guilt by an individual when weighing the utility of an action; and peers may feel the discrepancy between an individual's action and their own personal norms as anger. Other emotions, such as admiration, may function to identify and tag individuals whose actions or preferences become integrated within individual beliefs about actions prescribed by authority (G). These and other connections between computational approaches to emotions, psychological game theory [41, 42], and formal theories of norm psychology should be explored in future research.

The conceptual framework offered here can also guide research on cultural differences in norms and norm enforcement, which remains "mixed and fragmented" [43]. Recent findings indicate that changes in social distancing, handwashing, and other norms during the COVID pandemic varied significantly across countries [44, 45]; our framework suggests that this could be partially driven by differences in the preference weights A_1 , A_2 , A_3 , and A_4 , which in turn could be due to differences in environmental conditions in the evolutionary past. Further, the psychological mechanisms underpinning punishment can be expected to respond to the material payoff of enacting punishment, normative expectations about punishment, empirical expectations, and authority proscriptions. Thus, future cross-cultural research can identify and systematically test socio-ecological factors that may contribute to differences in preferences and beliefs to better understand cultural differences in norms, enforcement, and rates of change. Relatedly, future research could examine whether and why different groups within populations differ in norm-driving preferences; for example, experimental evidence indicates that men are more sensitive to peer approval than women, which may suggest that evolution has shaped male and female preferences differently [46].

5 Concluding remarks

Evolutionary theory provides a foundation for the view that norms result from individual decisions based on preferences and beliefs, and suggests promising paths for collaborative research between scholars in economics, psychology, and evolutionary sciences [43, 47]. Indeed, social norms lend themselves naturally to interdisciplinary research [48, 49]. Such research can not only satisfy our wish to understand the world, but also inform the design of policies aimed at influencing social norms [50, 51, 52, 53].

References

- Chandra Sripada and Stephen Stitch. A framework for the psychology of norms. In P. Carruthers, S. Laurence, and S. Stitch, editors, *Innateness and the Structure of the Mind, Vol. II*, pages 280–302. Oxford University Press, Oxford, 2007.
- [2] Cristina Bicchieri. The Grammar of Society. Cambridge University Press, Cambridge, 2006.
- [3] Cristina Bicchieri. Norms in the Wild. Oxford University Press, Oxford, 2016.
- [4] Giovanna D'Adda, Martin Dufwenberg, Francesco Passarelli, and Guido Tabellini.

Social norms with private values: Theory and experiments. *Games and Economic Behavior*, 124(2020):288–304, 2020.

- [5] Luise Görges and Daniele Nosenzo. Measuring social norms in economics: Why it is important and how it is done. *Analyse und Kritik*, 42(2):285–312, 2020.
- [6] Aron Szekely, Francesca Lipari, Alberto Antonioni, Mario Paolucci, Angel Sánchez, Luca Tummolini, and Giulia Andrighetto. Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications*, 12(1):1–7, 2021.
- [7] Arthur Schram, Jin Di Zheng, and Tatyana Zhuravleva. Corruption: A crosscountry comparison of contagion and conformism. *Journal of Economic Behavior Organization*, 193:497–518, 2022.
- [8] Sergey Gavrilets, Denis Tverskoi, and Angel Sánchez. Modelling social norms: an integration of the norm-utility approach with beliefs dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 379(1897), 2024.
- [9] Sergey Gavrilets. Coevolution of actions, personal norms and beliefs about others in social dilemmas. *Evolutionary Human Sciences*, 3:1–22, 2021.
- [10] David Marr. Vision A Computational Investigation Into the Human Representation and Processing of Visual Information. MIT Press, Cambridge MA, 1982.
- [11] Laith Al-shawaf. Psychological review levels of analysis and explanatory progress in psychology: Integrating frameworks from biology and cognitive levels of analysis and explanatory progress in psychology. *Psychological Review*, In Press.
- [12] Ingela Alger and Jörgen W. Weibull. Homo moralis—preference evolution under incomplete information and assortative matching. Econometrica, 81(6), 2013.
- [13] Ingela Alger. Evolutionarily stable preferences. Philosophical Transactions of the Royal Society B: Biological Sciences, 378(1876), 2023.

- [14] Ingela Alger, Jörgen W. Weibull, and Laurent Lehmann. Evolution of preferences in structured populations: Genes, guns, and culture. *Journal of Economic Theory*, 185:104951, 2020.
- [15] Ingela Alger and Laurent Lehmann. Evolution of semi-Kantian preferences in two-player assortative interactions with complete and incomplete information and plasticity. *Dynamic Games and Applications*, 13(4):1288–1319, 2023.
- [16] Carel P. van Schaik. The Primate Origin of Human Behavior. Wiley-Blackwell, New Jersey, 2016.
- [17] Peter J. Richerson, Sergey Gavrilets, and Frans B.M. De Waal. Modern theories of human evolution foreshadowed by Darwin's Descent of Man. *Science*, 372(6544), 2021.
- [18] Kenichi Aoki and Marcus W. Feldman. Evolution of learning strategies in temporally and spatially variable environments: A review of theory. *Theoretical Population Biology*, 91:3–19, 2014.
- [19] Rachel L. Kendal, Neeltje J. Boogert, Luke Rendell, Kevin N. Laland, Mike Webster, and Patricia L. Jones. Social learning strategies: Bridge-building between fields. *Trends in Cognitive Sciences*, 22(7):651–665, 2018.
- [20] Traci Hong, Joshua Cabrera, and Christopher E. Beaudoin. Disentangling realworld and virtual-world social norms: The persuasive elements and social psychological effects of a serious game. *Telematics and Informatics Reports*, 9:100038, 2023.
- [21] Marcel Montrey and Thomas R. Shultz. The evolution of high-fidelity social learning. Proceedings of the Royal Society B: Biological Sciences, 287(1928), 2020.
- [22] Charles Perreault and Robert Boyd. Evolution of social learning with payoff and content bias. *Games*, 13(1), 2022.

- [23] Richard W. Wrangham. Hypotheses for the evolution of reduced reactive aggression in the context of human self-domestication. *Frontiers in Psychology*, 10 (AUG):1–11, 2019.
- [24] Erik O. Kimbrough, Gordon M. Myers, and Arthur J. Robson. Infanticide and human self domestication. *Frontiers in Psychology*, 12(May):1–5, 2021.
- [25] Marcelo R. Sánchez-Villagra and Carel P. van Schaik. Evaluating the selfdomestication hypothesis of human evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, 28(3):133–143, 2019.
- [26] Thomas J. Bouchard. Authoritarianism, religiousness, and conservatism: Is "obedience to authority" the explanation for their clustering, universality and evolution? In E. Voland and W. Schiefenhövel, editors, *The Biological Evolution of Religious Mind and Behavior*, pages 165–180. Springer, Berlin, 2009.
- [27] Tian Chen Zeng, Joey T. Cheng, and Joseph Henrich. Dominance in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1845): 20200451, 2022.
- [28] Bram Kuijper, Olof Leimar, Peter Hammerstein, John M. McNamara, and Sasha R.X. Dall. The evolution of social learning as phenotypic cue integration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1828), 2021.
- [29] Erol Akçay and Jeremy Van Cleve. Internalizing cooperative norms in groupstructured populations. In Sarah F. Brosnan and Walter Wilczynski, editors, Cooperation and Conflict: The Interaction of Opposites in Shaping Social Behavior, pages 26–44. Cambridge University Press, 2021.
- [30] Sergey Gavrilets and Peter J. Richerson. Collective action and the evolution of social norm internalization. Proceedings of the National Academy of Sciences of the United States of America, 114(23):6068–6073, 2017.

- [31] Sydney Levine, Max Kleiman-Weiner, Laura Schulz, Joshua Tenenbaum, and Fiery Cushman. The logic of universalization guides moral judgment. *Proceedings* of the National Academy of Sciences of the United States of America, 117(42): 26158–26169, 2020.
- [32] Karen Huang, Regan M. Bernhard, Netta Barak-Corren, Max H. Bazerman, and Joshua D. Greene. Veil-of-ignorance reasoning mitigates self-serving bias in resource allocation during the COVID-19 crisis. *Judgment and Decision Making*, 16 (1):1–19, 2021.
- [33] Boris van Leeuwen and Ingela Alger. Estimating social preferences and Kantian morality in strategic interactions. *Journal of Political Economy Microeconomics*, In Press.
- [34] Jinyi Kuang and Cristina Bicchieri. Language matters: how normative expressions shape norm perception and affect norm compliance. *Philosophical Transactions* of the Royal Society B: Biological Sciences, 379(1897), 2024.
- [35] John Antonakis, Giovanna D'Adda, Roberto A. Weber, and Christian Zehnder.
 "Just words? Just speeches?" On the economic value of charismatic leadership. Management Science, 68(9):6355–6381, 2022.
- [36] Timothy D. Wilson. What is social psychology? The construal principle. Psychological Review, 129(4):873–889, 2022.
- [37] Axel Franzen and Sebastian Mader. The power of social influence: A replication and extension of the Asch experiment. *PLoS ONE*, 18(11 November):1–14, 2023.
- [38] Mitchell Landers, Daniel Sznycer, and Patrick Durkee. Are self-conscious emotions about the self? Testing competing theories of shame and guilt across two disparate cultures. *Emotion*, 2024.
- [39] Alexie Leroux, Sébastien Hétu, and Daniel Sznycer. The shame system operates with high precision. *Evolutionary Psychology*, 21(3):1–11, 2023.

- [40] Daniel Sznycer, Aaron Sell, and Debra Lieberman. Forms and functions of the social emotions. Current Directions in Psychological Science, 30(4):292–299, 2021.
- [41] Pierpaolo Battigalli and Martin Dufwenberg. Belief-dependent motivations and psychological game theory. *Journal of Economic Literature*, 60(3):833–882, 2022.
- [42] Charles Bellemare, Alexander Sebald, and Sigrid Suetens. Guilt aversion in economics and psychology. *Journal of Economic Psychology*, 73(452):52–59, 2019.
- [43] Catherine Molho, Francesca De Petrillo, Zachary H. Garfield, and Sam Slewe. Cross-societal variation in norm enforcement systems. *Philosophical Transactions* of the Royal Society B: Biological Sciences, 379(1897):17–20, 2024.
- [44] Giulia Andrighetto, Aron Szekely, Andrea Guido, Michele Gelfand, Jered Abernathy, Gizem Arikan, Zeynep Aycan, Shweta Bankar, Davide Barrera, and Dana Basnight-Brown et al. Changes in social norms during the early stages of the COVID-19 pandemic across 43 countries. *Nature Communications*, 15(1), 2024.
- [45] Eva Vriens, Giulia Andrighetto, and Luca Tummolini. Risk, sanctions and norm change: the formation and decay of social distancing norms. *Philosophical Trans*actions of the Royal Society B: Biological Sciences, 379(1897), 2024.
- [46] Giovanni Di Bartolomeo, Martin Dufwenberg, Stefano Papa, and Laura Razzolini. Guilt aversion : Eve versus Adam. Working Paper, University of Arizona, 2024.
- [47] Cecilia Heyes. Rethinking Norm Psychology. Perspectives on Psychological Science, 19(1):12–38, 2024.
- [48] Cristina Bicchieri, Eugen Dimant, Michele Gelfand, and Silvia Sonderegger. Social norms and behavior change: The interdisciplinary research frontier. *Journal of Economic Behavior and Organization*, 205:A4–A7, 2023.
- [49] Michele J. Gelfand, Sergey Gavrilets, and Nathan Nunn. Norm dynamics: Interdisciplinary perspectives on social norm emergence, persistence, and change. *Annual Review of Psychology*, 75:341–378, 2024.

- [50] S. Nageeb Ali and Roland Bénabou. Image versus information: Changing societal norms and optimal privacy. American Economic Journal: Microeconomics, 12(3): 116–164, 2020.
- [51] Leonardo Bursztyn, Alessandra L. González, and David Yanagizawa-Drott. Misperceived social norms: women working outside the home in Saudi Arabia. American Economic Review, 110(10):2997–3029, 2020.
- [52] Charles Efferson, Sonja Vogt, and Ernst Fehr. The promise and the peril of using social influence to reverse harmful traditions. *Nature Human Behaviour*, 4(1): 55–68, 2020.
- [53] Charles Efferson, Sönke Ehret, Lukas Von Flüe, and Sonja Vogt. When norm change hurts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 379(1897):0–1, 2024.

6 Further information on references of particular interest

[8] ** This paper reviews theoretical approaches for modelling the origin, persistence and change of social norms developed over the last 40 years.

[11] ** This review integrates Tinbergen's Four Questions and Marr's Three Levels of Explanation, emphasizing the development of formal models at the computational and algorithmic levels of analysis in order to strengthen our understanding of psychological phenomena, such as norm psychology.

[33] ** This article proposes an experimental protocol and statistical methods to estimate preferences that combine material self-interest, Kantian ethics, and other-regard; it reports estimates of said preferences based on decisions in three economic games (sequential prisoner's dilemma games, trust games, and ultimatum bargaining games).
[37] * This replication of the famous Asch conformity experiment found that monetary incentives reduce but do not negate the effect of social influence, and that the personality trait openness is associated with lower susceptibility to social influence.

[45] * Based on data on empirical and normative expectations about social distancing and of sanctioning, this article provides evidence on how social distancing and sanctioning norms evolved as risk associated with COVID-19 changed.

[49] ** This article provides an interdisciplinary review of the emerging field of norm dynamics by integrating research across the social sciences through a cultural-evolutionary lens.

7 Funding sources

I.A. and S.G. acknowledge funding from IAST funding from the French National Research Agency (ANR) under grant ANR-17-EURE-0010 (Investissements d'Avenir program). I.A. further acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 789111 - ERC EvolvingEconomics). S.G. was supported by the U.S. Army Research Office grants W911NF-18-1-0138, the Air Force Office of Scientific Research grant FA9550-21-1-0217, and the John Templeton Foundation.