RESEARCH ARTICLE

Falling through the cracks: Modeling the formation of social category boundaries

Vicky Chuqiao Yang*, Tamara van der Does, Henrik Olsson

Santa Fe Institute, Santa Fe, NM, United States of America

* vicky.chuqiao.yang@gmail.com

Abstract

Social categorizations divide people into "us" and "them", often along continuous attributes such as political ideology or skin color. This division results in both positive consequences, such as a sense of community, and negative ones, such as group conflict. Further, individuals in the middle of the spectrum can fall through the cracks of this categorization process and are seen as out-group by individuals on either side of the spectrum, becoming *inbetweeners*. Here, we propose a quantitative, dynamical-system model that studies the joint influence of cognitive and social processes. We model where two social groups draw the boundaries between "us" and 'them" on a continuous attribute. Our model predicts that both groups tend to draw a more restrictive boundary than the middle of the spectrum. As a result, each group sees the individuals in the middle of the attribute space as an out-group. We test this prediction using U.S. political survey data on how political independents are perceived by registered party members as well as existing experiments on the perception of racially ambiguous faces, and find support.

OPEN ACCESS

Citation: Yang VC, van der Does T, Olsson H (2021) Falling through the cracks: Modeling the formation of social category boundaries. PLoS ONE 16(3): e0247562. https://doi.org/10.1371/journal.pone.0247562

Editor: Luis M. Miller, Spanish National Research Council, SPAIN

Received: September 5, 2020

Accepted: February 9, 2021

Published: March 31, 2021

Data Availability Statement: All data used in our analysis are publicly available from the American National Election Studies (https://electionstudies.org/project/anes-time-series-cumulative-data-file/).

Funding: This research is supported by the Santa Fe Institute Omidyar Fellowship, Suzanne Hurst and Samuel Peters. We also thank the Santa Fe Institute for meeting support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

1 Introduction

Social categorization is a necessary and ubiquitous human social behavior, occurring on many attributes including race, gender, sexual orientation, and political ideology [1]. On the one hand, social categorizing is essential for fulfilling a sense of community and a positive sense of self [2]. On the other hand, it can fuel social conflicts by creating an "us" versus "them" mentality and impacting certain groups' access to economic and social resources [3–6]. For example, the division between White and Black Americans has led to continuing discrimination and segregation long after the abolition of slavery [7]. Recently, the divisions between Democrats and Republicans have created "fear and loathing" among U.S. voters [8].

A common theme in research on social categorization is to investigate the process of categorizing people as belonging to one's in-group or to an out-group. There is a vast literature on various intergroup biases such as in-group favoritism and out-group derogation [9–11]. Most of the research on social categories focuses on the end result of a social categorization process, where the typical assumption is that this process only leads to the perception of two groups [10, 11]. Moreover, most experimental work on individual classifications of race or gender attributes presents participants with pre-determined and forced choices [for a review, see 12].

Competing interests: The authors have declared that no competing interests exist.

Similarly, in theories of social impression formation and social categorization, category representations and group motives are mostly treated as exogenous to the analysis with a fixed category structure in which individuals can be placed [13, 14].

As a result of the categorization process, individuals can also "fall through the cracks" and not belong to any well-established social group. We refer to these individuals as *inbetweeners*. Examples include mixed-race individuals who are considered neither truly Black nor White by members of either group, and political independents considered as "other" by both Democrats and Republicans. With demographic shifts, such as over ten million in the U.S. who identify with two or more races [15] and the increasing gender non-binary population [16], understanding how individuals fall through the cracks of categorization and the subsequent consequences are increasingly important. The existence of inbetweeners can be accommodated in existing models of social categorization by simply assuming that inbetweeners is a separate category. This assumption, however, does not address the question of how the boundaries of other categories are formed and how the inbetweeners category is created.

Social categorization draws on both individual and social-level processes. A model for the formation of boundaries between categories must therefore include both levels. At the individual, or cognitive, level categorization decisions must relate to the distance between individuals. The influence of distance between individuals on categorization is part of many models of social categorization and social judgment [17, 18]. At the social level, this model must take into consideration how other individuals influence the formation of boundaries between categories. The influence of others' beliefs and actions are well established in research on social categorization, social judgments, social learning, and belief formation [19–21]. Adopting categorization beliefs that are not supported by others in one's immediate social environment can be costly, because it triggers disapproval of others, withdrawal of cooperation, open conflict, or even ostracism [22, 23]. In the social categorization literature, however, there are no quantitative models that integrate both the individual and social perspectives to predict how category boundaries are formed.

In this paper we propose a quantitative, dynamical system model of social categorization that integrates cognitive and social processes. It predicts where category boundaries are placed and the occurrence of inbetweeners. Dynamical system models are useful for studying both formation and evolution processes because they enable tracking the feedback among many variables simultaneously. These models have been successful in explaining and predicting many complex social phenomena [24], such as the extinction of minority languages [25], the decline of religious affiliation [26], the polarization in the U.S. Congress [27], and changes in party memberships in the UK [28]. In this model, the process of creating social category boundaries is influenced by individual-level cognitive processes and social processes. In the cognitive process, individuals want to recognize whether others are similar to them. This consideration is supported by several social-psychological mechanisms such as building successful collaborations [29, 30] as well as forming community and a positive sense of self [2, 31]. However, remembering the exact distance between individuals is costly, therefore individuals will use categories as a summary of others' position. At the social level, we assume that individuals consider the boundary choices of other group members. Those in the same group want to agree on who are in the in-group and who are in the out-group. After we present the mathematical model and its predictions, we present some preliminary empirical validation from the American National Election Studies (ANES) dataset and compare our results with findings from human behavior experiments.

2 The mathematical model

We model the formation of two social groups on a continuous attribute [32] and consider each group to have a boundary that divides the population into in-group and out-group. We will derive two governing equations, one for the boundary position of each group. We show in detail the derivation for one group, which will be similar for the other group. The derivation is achieved in two parts. In the first part (Section 2.1), we consider the individual-level cognitive process and derive the error in categorization for each individual. In the second part (Section 2.2), we consider the group-level social process for agreeing with others in the same group, and derive the boundary position through optimizing the categorization error of the group.

We denote the lower and upper bounds of the continuous attribute value x to be a and b, respectively, and the population distribution on the attribute to be $\rho(x)$. We consider two groups forming on the continuous attribute space. One contains the left extreme of the attribute space, denoted as group 1, with boundary position z_1 that divides the in-group and out-group (see Fig 1 for an illustration of the variables for this group). The other contains the right extreme of the attribute space, denoted as group 2, with boundary position z_2. The set-up for group 2 is symmetrical to that of group 1. We treat the boundary positions z_1 and z_2 as unknowns to be solved in the model.

2.1 Individual-level cognitive process

The central insight from decades of research on categorization is that our cognitive system searches for patterns and structures [33]. The perception and cognitive representations of these patterns and structures can take many forms. In line with prototypical theories of category representations [34, 35], we assume a prototypical representation in the form of the mean position of a group. That is, we assume that all individuals categorized in the same group are perceived to have the group's mean position. For example, all individuals categorized under "Democrat" are perceived to have the mean position of all Democrats. Mathematically, the group positions for the in-group ($g_{\text{in}}(z_1)$) and out-group ($g_{\text{out}}(z_1)$) are defined as the the center of mass of the population distribution in each group,

$$g_1^{\text{in}}(z_1) = \frac{\int_a^{z_1} x\,\rho(x)dx}{\int_a^{z_1} \rho(x)dx}\ , \text{ and } \ g_1^{\text{out}}(z_1) = \frac{\int_{z_1}^{b} x\,\rho(x)dx}{\int_{z_1}^{b} \rho(x)dx}\ . \tag{1}$$

We consider that individuals want to form categories of in-group and out-group because it is less costly than remembering individuals' precise attribute positions [12], and they want the

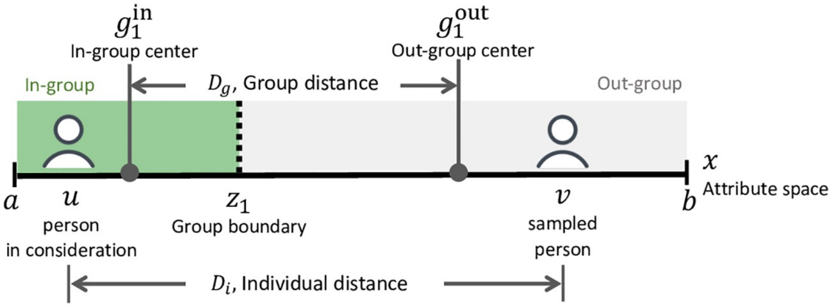

Fig 1. Illustration of the variables in the model. The illustration is presented from the perspective of a member of group 1, interacting with an individual on the other side of the group boundary. The individual categorization error is the difference between individual distance and group distance. Group 2 is not shown in this illustration.

https://doi.org/10.1371/journal.pone.0247562.g001

categorization to reflect the actual differences in attributes as much as possible. Consider a person U in the in-group of group 1, with position u ($u < z_1$) on the attribute space. U interacts with others on the attribute space through random sampling. Let V be another individual on the attribute space (with position v) who interacts with U. U observes the position of V (for example, on the liberal to conservative scale). D_i denotes the distance between the two interacting individuals, $D_i = |u - v|$, and D_g denote the distance between the two individuals' group positions (Fig 1). We define the categorization error for the interaction between U and V to be the squared difference between the individual distance and the group distance, $(D_g - D_i)^2$. This error represents how much the group representation differs from the individual representation. The categorization error for U perceiving *all* sampled individuals is the integral of these errors with respect to v, weighted by the population density ρ,

$$err(u, z_1) = \int_a^b (D_g - D_i)^2 \rho(v) dv \ . \tag{2}$$

The term $D_i = |u - v|$ is calculated for all pairs of individuals. The term D_g varies depending on if v is in the in-group or out-group. If v is in the in-group, both individuals are considered to be in the same group, $D_g = 0$. If v is in the out-group, $D_g = |g_{\text{in}} - g_{\text{out}}|$. Combining with Eq 2, we have,

$$err_1(u, z_1) = \int_a^{z_1} |u - v|^2 \rho(v) dv + \int_{z_1}^b (|g_1^{\text{out}}(z_1) - g_1^{\text{in}}(z_1)| - |u - v|)^2 \rho(v) dv \ . \tag{3}$$

The first term in Eq (3) represents the error when the sampled person is in the in-group ($v < z_1$). The second term represents the case when the sampled person is in the out-group ($v > z_1$, the case illustrated in Fig 1). The expression for the boundary of group 2, z_2 will be similar to the derivation process above. The domain of integration for the in-group will be changed from between a and z_1 to between z_2 and b. The domain for the out-group will be changed from between z_1 and b to between a and z_2.

Motivated by previous research [36, 37], we considered an alternative formulation using similarity instead of distance. The model reaches the same main conclusion, though more mathematically involved, as shown in S1A Appendix.

2.2 Group-level social process

In the social process, we consider individuals to be motivated to form a consistent category boundary with others in the same group. This, we assume, is a consequence of social learning and conformism where people are motivated by accuracy and affiliation goals [38]. In our implementation, we approximate this process by assuming that individuals are concerned with agreeing with other in-group members [39] and strive to minimize the average collective in-group categorization error. Individuals observe other members' categorizations and update their boundary position in the direction of the other members' boundaries. Individuals do this while taking the individual categorization error (of Eq 4) into account. This process repeats until the group arrives at one boundary position. The average collective in-group error for group 1 is,

$$Err_1(z_1) = \frac{1}{\int_a^{z_1} \rho(x) dx} \int_a^{z_1} err_1(u, z_1) \rho(u) du \ . \tag{4}$$

Note that Eq (4) does not impose any preferences on group size. Moreover, this formulation assumes that individuals weigh all other members of the in-group as equally important. Finally, we consider that the group dynamically adjusts its boundary position to minimize the collective error,

$$\frac{dz_1}{dt} = -k\frac{dErr_1(z_1)}{dz_1} \ , \tag{5}$$

where t is time, and k is a constant that sets the time scale of the system. The intuitive understanding of Eq (5) is that the category boundary evolves towards the direction that reduces the in-group's collective categorization error. A similar process occurs for group 2, where the domain of integration in Eq 5 is replaced by between z_2 and b, and $err_1(u, z_1)$ is replaced by $err_2(u, z_2)$. The social process above can also be formulated as optimizations on the individual level, though with more complexity (S1B Appendix).

3 Results

3.1 Model predictions

We first present the results in the case where the attribute distribution $\rho(x)$ is a uniform distribution between 0 and 1 to demonstrate the behavior of the model. With the uniform $\rho(x)$, the individual-level categorization error for members of group 1 is,

$$err_1(u, z_1) = u^2 - uz_1 + z_1^2/2 - z_1/4 + 1/12. \tag{6}$$

With this, we can analytically calculate the collective error,

$$Err_1(z_1) = \frac{1}{3}z_1^2 - \frac{1}{4}z_1 + \frac{1}{12} \ . \tag{7}$$

Eq (5) has one stable fixed point, $z_1^* = 3/8 = 0.375$, meaning the boundary position for group 1 stabilizes at 0.375: this group considers those with attribute value $x < 0.375$ as in-group, and those with attribute value $x > 0.375$ as out-group. A same set of equations can be derived for group 2 (individuals on the right side of the spectrum). By symmetry, the preferred group boundary of group 2 is $z_2^* = 0.625$. This leads to individuals between 0.375 and 0.625 to be considered out-group by both social groups, which we refer to as inbetweeners (see Fig 2(a)).

The occurrence of inbetweeners is not unique to the uniform attribute distribution. We now present results obtained considering the attribution distribution $\rho(x)$ as a Beta distribution. The Beta distribution is parameterized by two positive shape parameters, α and β, with probability density function (PDF) $f_{\text{beta}}(x, \alpha, \beta) = x^{\alpha-1}(1 - x)^{\beta-1}/B(\alpha, \beta)$, where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$, and $\Gamma(\cdot)$ is the Gamma function. The distribution is defined for x in the interval [0, 1]. We choose the Beta distribution because by adjusting the shape parameters we can produce a wide variety of unimodal distributions, both symmetrical and skewed. A number of real-world attribute distributions are known to be unimodal, such as political ideology of the U.S. public measured by positions on public policy issues [40]. Panels (b) and (c) in Fig 2 show two examples of the Beta distribution as attribute distribution $\rho(x)$, one symmetrical and one asymmetrical. In both cases, inbetweeners appear, though the location and size of the region vary with the distribution. We have also analyzed the results for a bi-modal attribute distribution. We construct bi-modal distributions by summing two skewed Beta distributions that are symmetrical to each other. The PDF is $f_{\text{bimodal}}(x, \alpha, \beta) = 1/2[f_{\text{beta}}(x, \alpha, \beta) + f_{\text{beta}}(x, \beta, \alpha)]$. Panel (d) in Fig 2 shows the results for a bi-modal distribution with shape parameters $\alpha = 2$ and $\beta = 7$.

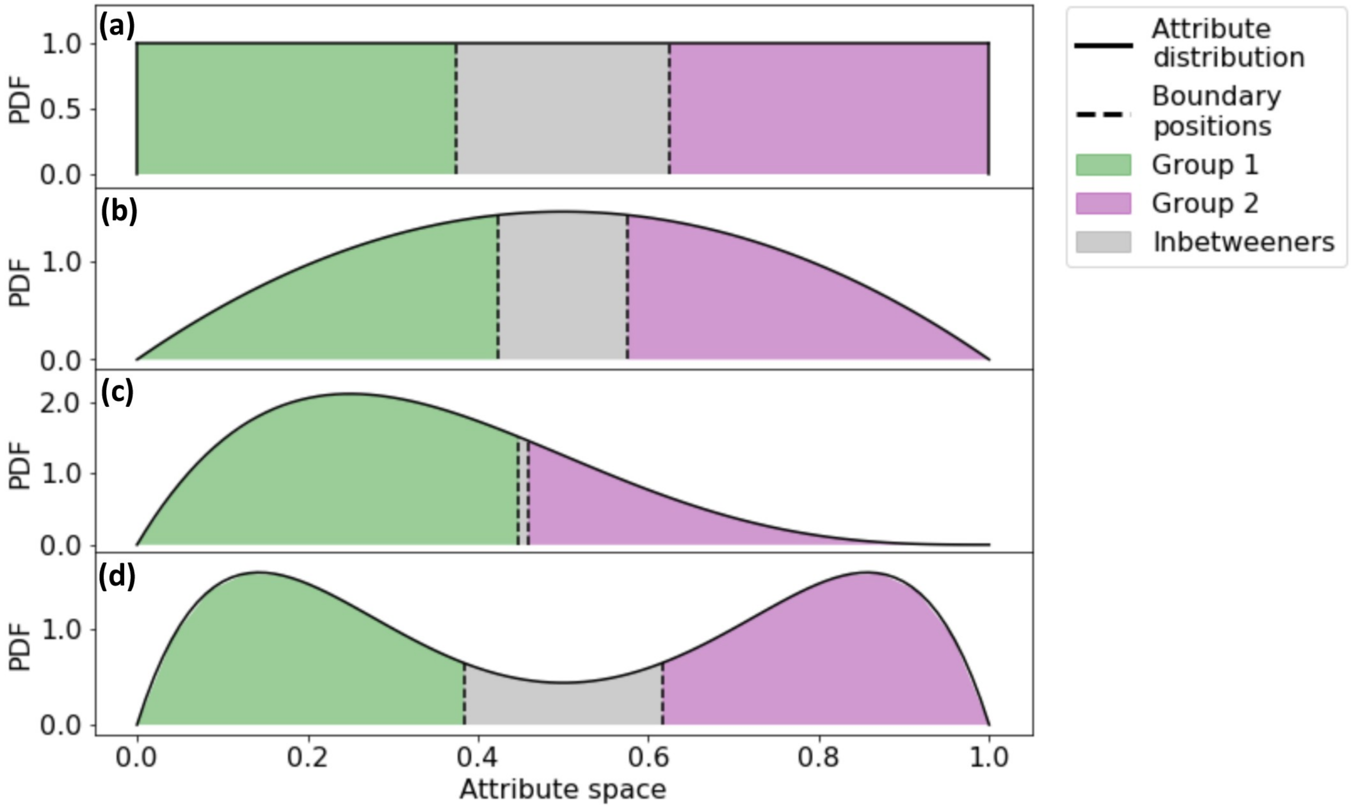

Fig 2. Stable fixed points of boundary positions for both groups: (a) for a uniform attribute distribution, (b) for the symmetrical Beta distribution with shape parameters $\alpha = 2$ and $\beta = 2$, (c) for the asymmetrical Beta distribution with $\alpha = 2$ and $\beta = 4$, and (d) for a bi-modal distribution which is the sum of two Beta distributions, whose shape parameters are $\alpha = 2, \beta = 7$, and $\alpha = 7, \beta = 2$.

https://doi.org/10.1371/journal.pone.0247562.g002

3.2 Validation with empirical findings

Our model predicts that those in the middle of the attribute space are seen as out-group by the two social groups as an outcome of the categorization process, becoming inbetweeners. While this paper's main focus is a contribution to theory, we look at data for preliminary validation of the model's predictions.

We test our predictions using the American National Election Studies (ANES) dataset. We focus on how Democrats and Republicans perceive those in the middle of the liberal-conservative attribute space. Since political independents tend to self-identify as being in the middle of the liberal-conservative spectrum, and Democrats and Republicans tend to select positions on either side (see S1D Appendix), we use political independents as an approximation for those in the middle of the attribute space. We want to test if independents are perceived by both parties as part of the in-group (as favorably as one's own party), as the out-group (as unfavorably as the other party), or somewhere in between. If the perception of independents is similar to that of the other party, then the data supports our models' prediction. We draw on research in social psychology [14, 41] to argue that negative feelings are strongly driven by an out-group categorization. Even though feelings towards others are also driven by the difficulty in categorizing them [42, for a review], we present here established categories thus removing any cognitive categorization work. We use feelings as an approximation for the categorization process and propose in the discussion other ways to test this model.

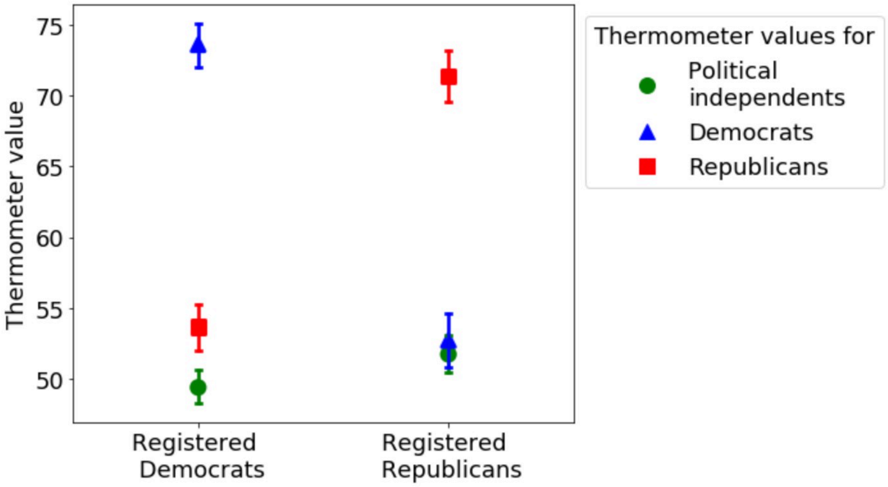

Fig 3. The mean thermometer values (reflecting feeling favorably or unfavorably) towards both political parties and political independents reported by registered party members (ANES). The error bars are 95% confidence interval of the mean. For both registered Democrats and Republicans, political independents are perceived similarly compared to members of the other party, while members of own party are perceived more favorably.

https://doi.org/10.1371/journal.pone.0247562.g003

The ANES dataset is a nationally representative survey of political attitudes in the U.S. public. We use registered Democrat and Republican party members to represent the two groups on opposite sides of a continuous spectrum, as measured by self-reported party registration. We use a set of thermometer questions to measure attitudes towards Democrats, Republicans, and political independents. In each thermometer question, participants are asked to report a number between 0 and 100—if they feel favorably about a group, a number greater than 50, and if they feel unfavorably about them, a number lower than 50 (see S1C Appendix for data source and questionnaire details). We use data from years 1980 and 1984, because the thermometer questions about political independents were only asked in these two years' surveys (N = 1,923).

Fig 3 shows the mean thermometer values towards political independents, Democrats, and Republicans, reported by registered members of both parties. For both Democrats and Republicans, political independents are perceived similarly to members of the other party, while members of one's own party are perceived a lot more favorably. We perform a two-sided t-test and show that for Republican party members, the mean of thermometer values for Democrats and for political independents are indistinguishable ($p = 0.42$). The same test shows that for registered Democrats, the mean value for feelings toward Republicans is slightly higher than that of the political independents ($p < 0.001$). One's own party is perceived significantly more favorably than both the other party and the independents ($p < 0.001$).

Beyond this ANES empirical test, our model's prediction is also in agreement with previous empirical studies on racial categorizations: in-group members tend to categorize ambiguous individuals as out-group, a process known as the in-group over-exclusion effect [12, 43]. The following four types of experimental studies have confirmed this phenomenon. First, using established racial categories, perceivers in the U.S. tend to categorize racially ambiguous individuals as the out-group [44, 45], which was replicated in South Africa [46] and Italy [47]. Second, using memory tests, an experimental study [48] finds that racially ambiguous faces are perceived as out-groups by mono-racial individuals. Third, using open-ended categorization, a recent study points out that perceivers tend to use a third category (in this case, Hispanic or Middle Eastern) for racially-ambiguous individuals (who were mixed Black and White) [49].

Finally, studies measuring feelings towards bi-racial individuals find that they are on average rated more negatively through the process of categorization [50, 51]. Taken together, these studies suggest that racial groups tend to draw boundaries that exclude individuals of mixed races, supporting our model's prediction of inbetweeners.

4 Discussion

We propose a dynamical system model that integrates cognitive and social processes to arrive at social categorizations. The model predicts that social groups tend to draw boundaries that are more restrictive than the median of the attribute spectrum. As a consequence, those in the middle of the attribute space are excluded by both social groups, becoming inbetweeners. Our theoretical finding is supported by empirical analysis of attitudes towards politically independent individuals by registered Democrats and Republicans, as well as by previous empirical findings surrounding the in-group over-exclusion effect in racial categorization [12, 43]. The prediction of the existence of inbetweeners is unique to our model. Our model provides a rare theoretical result on how inbetweeners can arise through the process of social categorization. Although this work dominantly uses data on political ideology and racial categorization, our results can be extended more generally to individuals in the middle region of attribute spaces, such as those who are gender non-binary or in interdisciplinary scientific fields.

Our model is intentionally parsimonious, aiming to capture key cognitive and social processes. We show here that a simple model can capture the main aspects of social category boundary formation. We do not attempt to model the influence of the myriad of motivational factors investigated in the social categorization literature, such as self-image maintenance [52] or motivated reasoning with stereotypes [53]. We also leave out many complex cognitive and social processes which can influence social categorization, such as individuals' previous experiences and implicit biases towards members across the attribute spectrum [12, 47, 51] or culturally-based group hierarchies [47, 54]. Social boundaries also evolve over time. For example, the shifting demographics in the U.S. since the 1960s have extended racial category labels beyond the dichotomy of Black and White [55]. The dynamical systems framework we propose can be used, in future research, to explore how demographic and cultural shifts lead to changes in social category boundaries. In this manuscript, we study the rise of inbetweeners as a result of categorization. It is possible for this result to have further downstream consequences, such as forming a new group with other inbetweeners. It would be useful for future research to study how inbetweeners form new groups.

Our empirical validation shows initial support for the presented model, however the thermometer measures can only approximate how individuals categorize each other as in- or out-group. Affect (feelings toward others) is the result of many categorization processes beyond boundary formation, such as the difficulty of categorizing the other [14, 42]. Future research can benefit from investigating the perception of inbetweeners in an experimental setting, where the distribution of individuals in the attribute space is known and the outcome variable focuses on actual categorization and not affect. This could be achieved by asking individuals across the political spectrum to mark others as in- or out-group based on their policy views.

Much significance of social categories is not in the categories themselves, but in how these categories affect how individuals are perceived and treated. Our model's prediction that individuals with characteristics in the middle of the attribute space "fall through the cracks" may affect many social processes. One speculative example is the disconnect between issue polarization and social polarization in the U.S. public. Previous empirical research has found that identification with political parties and antipathy towards the opposing party increased disproportionally compared to opinion on issues [56–59]. Our model provides a possible explanation that political independents are perceived as out-group by both political parties. Motivated by the need for belonging and community, individuals holding moderate positions might decide to instead identify with one of the two polarized parties, despite misalignment on issue positions. Empirically testing how political independents are perceived and how this relates to social polarization can be an important direction for future research.

Supporting information

S1 Appendix.
(PDF)

Acknowledgments

We thank the 2018 Fall JSMF-SFI Postdoctoral Conference for allowing a research jam on the topic of categorical perception, where this project started. We thank Mirta Galesic and Sidney Redner for helpful feedback on the manuscript.

Author Contributions

Conceptualization: Vicky Chuqiao Yang, Tamara van der Does, Henrik Olsson.

Data curation: Vicky Chuqiao Yang.

Formal analysis: Vicky Chuqiao Yang.

Investigation: Vicky Chuqiao Yang, Tamara van der Does, Henrik Olsson.

Methodology: Vicky Chuqiao Yang, Tamara van der Does.

Project administration: Vicky Chuqiao Yang.

Visualization: Vicky Chuqiao Yang.

Writing – original draft: Vicky Chuqiao Yang, Tamara van der Does.

Writing – review & editing: Vicky Chuqiao Yang, Tamara van der Does, Henrik Olsson.

References

1. Lamont M. & Molnár V. The study of boundaries in the Social Sciences. *Annual Review of Sociology* 28, 167–195 (2002). https://doi.org/10.1146/annurev.soc.28.110601.141107
2. Tajfel H. & Turner J. C. The Social Identity Theory of Intergroup Behavior. In Worchel S. & Austin W. (eds.) *Psychology of Intergroup Relations* (Nelson-Hall, Chicago, IL, 1986), 2nd edn.
3. Bremmer I. *Us vs. Them: The Failure of Globalism* (Penguin, 2018).
4. Ashmore R. D., Deaux K. & McLaughlin-Volpe T. An Organizing Framework for Collective Identity: Articulation and Significance of Multidimensionality. *Psychological Bulletin* 130, 80–114 (2004). https://doi.org/10.1037/0033-2909.130.1.80 PMID: 14717651
5. Thoits P. A. & Virshup L. K. Me's and We's. In Ashmore R. D. & Jussim L. J. (eds.) *Self and Identity: Fundamental Issues*, 106–133 (Oxford University Press, Oxford, 1997).
6. Roccas S., Sagiv L., Schwartz S., Halevy N. & Eidelson R. Toward a unifying model of identification with groups: Integrating theoretical perspectives. *Personality and Social Psychology Review* 12, 280–306 (2008). https://doi.org/10.1177/1088868308319225 PMID: 18641386
7. Fox C. & Guglielmo T. A. Defining America's racial boundaries: Blacks, Mexicans, and European immigrants, 1890-1945. *American Journal of Sociology* 118, 327–379 (2012). https://doi.org/10.1086/666383
8. Iyengar S. & Westwood S. J. Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science* 59, 690–707 (2015). https://doi.org/10.1111/ajps.12152
9. Brewer M. B.The psychology of prejudice: Ingroup love and outgroup hate?*Journal of Social Issues* 55, 429–444 (1999). https://doi.org/10.1111/0022-4537.00126

10. Hewstone M., Rubin M. & Willis H. Intergroup bias. *Annual Review of Psychology* 53, 575–604 (2002). https://doi.org/10.1146/annurev.psych.53.100901.135109 PMID: 11752497

11. Dovidio J. F. & Gaertner S. L. Intergroup bias. *Handbook of Social Psychology* (2010).

12. Bodenhausen G. V., Kang S. K. & Peery D. Social categorization and the perception of social groups. *The Sage Handbook of Social Cognition* 318–336 (2012).

13. Brewer M. B.A dual process model of impression formation. In Srull T. K. & Wyer R. S. Jr (eds.) *Advances in Social Cognition*, vol. 1 (Erlbaum, Hillsdale, NJ, 1988).

14. Fiske S. T. & Neuberg S. L. A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in Experimental Social Psychology*, vol. 23, 1–74 (Elsevier, 1990).

15. US Census Bureau. American Community Survey (2017).

16. Schilt K. & Lagos D. The Development of Transgender Studies in Sociology. *Annual Review of Sociology* 43, 425–443 (2017). https://doi.org/10.1146/annurev-soc-060116-053348

17. Galesic M., Olsson H. & Rieskamp J. A sampling model of social judgment. *Psychological Review* 125, 363 (2018). https://doi.org/10.1037/rev0000096 PMID: 29733664

18. Smith E. R. & Zarate M. A. Exemplar-based model of social judgment. *Psychological Review* 99, 3–21 (1992). https://doi.org/10.1037/0033-295X.99.1.3

19. Cialdini R. B. & Trost M. R. Social influence: Social norms, conformity and compliance. In Gilbert D. T., Fiske S. T. & Lindzey G. (eds.) *The Handbook of Social Psychology*, vol. 2, 151–192 (Boston: McGraw-Hill, 1998), 4th edn.

20. Festinger L.A theory of social comparison processes. *Human Relations* 7, 117–140 (1954). https://doi.org/10.1177/001872675400700202

21. Ajzen I.et al. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50, 179–211 (1991). https://doi.org/10.1016/0749-5978(91)90020-T

22. Williams K. D.Ostracism. *Annu. Rev. Psychol.* 58, 425–452 (2007). https://doi.org/10.1146/annurev.psych.58.110405.085641 PMID: 16968209

23. Feinberg M., Willer R. & Schultz M. Gossip and ostracism promote cooperation in groups. *Psychological Science* 25, 656–664 (2014). https://doi.org/10.1177/0956797613510184 PMID: 24463551

24. Castellano C., Fortunato S. & Loreto V. Statistical physics of social dynamics. *Reviews of Modern Physics* 81, 591 (2009). https://doi.org/10.1103/RevModPhys.81.591

25. Abrams D. M. & Strogatz S. H. Linguistics: Modelling the dynamics of language death. *Nature* 424, 900 (2003). https://doi.org/10.1038/424900a PMID: 12931177

26. Abrams D. M., Yaple H. A. & Wiener R. J. Dynamics of social group competition: Modeling the decline of religious affiliation. *Physical Review Letters* 107, 088701 (2011). https://doi.org/10.1103/PhysRevLett.107.088701 PMID: 21929211

27. Lu X., Gao J. & Szymanski B. K. The evolution of polarization in the legislative branch of government. *Journal of the Royal Society Interface* 16, 20190010 (2019). https://doi.org/10.1098/rsif.2019.0010 PMID: 31311437

28. Jeffs R. A., Hayward J., Roach P. A. & Wyburn J. Activist model of political party growth. *Physica A: Statistical Mechanics and its Applications* 442, 359–372 (2016). https://doi.org/10.1016/j.physa.2015.09.002

29. Smaldino P. E., Pickett C. L., Sherman J. & Schank J. An Agent-Based Model of Social Identity Dynamics. *Journal of Artificial Societies and Social Simulation* 15, 1–17 (2012). https://doi.org/10.18564/jasss.2030

30. Smaldino P. E.Social identity and cooperation in cultural evolution. *Behavioural Processes* 161, 108–116 (2019). https://doi.org/10.1016/j.beproc.2017.11.015 PMID: 29223462

31. Cerulo K. A.Identity Construction: New Issues, New Directions. *Annual Review of Sociology* 23, 385–409 (1997). https://doi.org/10.1146/annurev.soc.23.1.385

32. For simplicity, we present the model for two groups, which already leads to complex model behavior. The modeling framework is readily extendable to *n* groups.

33. Rosch E. & Mervis C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7, 573–605 (1975). https://doi.org/10.1016/0010-0285(75)90024-9

34. Posner M. I. & Keele S. W. On the genesis of abstract ideas. *Journal of Experimental Psychology* 77, 353–363 (1968). https://doi.org/10.1037/h0025953 PMID: 5665566

35. Rosch E. H.Natural categories. *Cognitive Psychology* 4, 328–350 (1973). https://doi.org/10.1016/0010-0285(73)90017-0

36. Nosofsky R. M.Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology*: *Learning*, *Memory*, *and Cognition* 10, 104–114 (1984). PMID: 6242730

37. Shepard R.Toward a universal law of generalization for psychological science. *Science* 237, 1317–1323 (1987). https://doi.org/10.1126/science.3629243 PMID: 3629243

38. Cialdini R. B. & Goldstein N. J. Social influence: Compliance and conformity. *Annual Review of Psychology* 55, 591–621 (2004). https://doi.org/10.1146/annurev.psych.55.090902.142015 PMID: 14744228

39. Soll J. B. & Larrick R. P. Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology*: *Learning*, *Memory*, *and Cognition* 35, 780 (2009). PMID: 19379049

40. Fiorina M. P. & Abrams S. J. Political polarization in the American public. *Annual Review of Political Science* 11, 563–588 (2008). https://doi.org/10.1146/annurev.polisci.11.053106.153836

41. Fiske S. T. Schema-triggered affect: Applications to social perception. In *Affect and Cognition: 17th Annual Carnegie Mellon Symposium on Cognition*, 55–78 (Hillsdale: Lawrence Erlbaum, 1982).

42. Lick D. J. & Johnson K. L. The interpersonal consequences of processing ease: Fluency as a metacognitive foundation for prejudice. *Current Directions in Psychological Science* 24, 143–148 (2015). https://doi.org/10.1177/0963721414558116

43. Leyens J. P. & Yzerbyt V. Y. The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. *European Journal of Social Psychology* 22, 549–569 (1992). https://doi.org/10.1002/ejsp.2420220604

44. Gaither S. E., Pauker K., Slepian M. L. & Sommers S. R. Social belonging motivates categorization of racially ambiguous faces. *Social Cognition* 34, 97–118 (2016). https://doi.org/10.1521/soco.2016.34.2.97

45. Peery D. & Bodenhausen G. V. Black + White = Black: Hypodescent in reflexive categorization of racially ambiguous faces. *Psychological Science* 19, 973–977 (2008).

46. Pettigrew T. F., Allport G. W. & Barnett E. O. Binocular resolution and perception of race in South Africa. *British Journal of Psychology* 49, 265–278 (1958). https://doi.org/10.1111/j.2044-8295.1958.tb00665.x PMID: 13596569

47. Castano E., Yzerbyt V., Bourguignon D. & Seron E. Who may enter? The impact of in-group identification on in-group/out-group categorization. *Journal of Experimental Social Psychology* 38, 315–322 (2002). https://doi.org/10.1006/jesp.2001.1512

48. Pauker K.et al. Not so black and white: Memory for ambiguous group members. *Journal of Personality and Social Psychology* 96, 795 (2009). https://doi.org/10.1037/a0013265 PMID: 19309203

49. Nicolas G., Skinner A. L. & Dickter C. L. Other than the sum: Hispanic and Middle Eastern categorizations of Black–White mixed-race faces. *Social Psychological and Personality Science* 10, 532–541 (2019). https://doi.org/10.1177/1948550618769591

50. Halberstadt J. & Winkielman P. Easy on the eyes, or hard to categorize: Classification difficulty decreases the appeal of facial blends. *Journal of Experimental Social Psychology* 50, 175–183 (2014). https://doi.org/10.1016/j.jesp.2013.08.004

51. Freeman J. B., Pauker K. & Sanchez D. T. A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science* 27, 502–517, (2016). https://doi.org/10.1177/0956797615627418 PMID: 26976082

52. Fein S. & Spencer S. J. Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology* 73, 31 (1997). https://doi.org/10.1037/0022-3514.73.1.31

53. Kundra Z. & Sinclair L. Motivated reasoning with stereotypes: Activation, application, and inhibition. *Psychological Inquiry* 10, 12–22 (1999). https://doi.org/10.1207/s15327965pli1001_2

54. Wimmer A.*Ethnic Boundary Making*: *nstitutions*, *Power*, *Networks* (Oxford University Press, Oxford, UK, 2013).

55. Bonilla-Silva E.From bi-racial to tri-racial: Towards a new system of racial stratification in the USA. *Ethnic and Racial Studies* 27, 931–950 (2004). https://doi.org/10.1080/0141987042000268530

56. Mason L."I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science* 59, 128–145 (2015). https://doi.org/10.1111/ajps.12089

57. Fiorina M. P. & Abrams S. J. Political polarization in the American public. *Annual Review of Political Science* 11, 563–588 (2008). https://doi.org/10.1146/annurev.polisci.11.053106.153836

58. Hetherington M. J.Review article: Putting polarization in perspective. *British Journal of Political Science* 39, 413–448 (2009). https://doi.org/10.1017/S0007123408000501

59. Hill S. J. & Tausanovitch C. A disconnect in representation? Comparison of trends in congressional and public polarization. *The Journal of Politics* 77, 1058–1075 (2015). https://doi.org/10.1086/682398