Falling Through the Cracks: A Dynamical Model for the Formation of In-Groups and Out-Groups

V. Chuqiao Yang¹, Tamara van der Does¹, and Henrik Olsson¹

¹ Santa Fe institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA.

ABSTRACT

Social categorizations are an important aspect of human society, dividing people into “us” and “them” often along attributes such as race or even political ideology. Social categorizations can, on the one hand, lead to a sense of belonging, and on the other hand, fuel conflicts. Thus, it is vital to understand where and how the boundaries for social categories form. We present a dynamical systems model for the formation of two social groups on a continuous attribute. Our model makes the surprising prediction that individuals in the middle of the spectrum are seen as “them” by both social groups, and “us” by neither. We refer to these individuals as inbetweeners. We test the model’s prediction using US 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. Our model suggests that inbetweeners can arise as the result of both groups drawing a more restrictive boundary than the middle of the spectrum, motivated by improving categorization accuracy alone. The model prediction is general and may be extended to social categorization along many other attributes.

Keywords: dynamical system, applied mathematics, categorical perception, social groups

1 Introduction

Dynamical systems models instantiated in simple differential-equations have been able to explain and predict many complex social phenomena [1]. Examples include modeling the extinction of minority languages [2], the decline of religious affiliation [3], the polarization in the US Congress [4], and changes in party memberships in the UK [5]. A remaining challenge in this field is to establish greater connection with the social sciences, such as grounding the models in established empirical findings. In this paper, we use a simple differential-equation based model to explore the formation of social categories. We draw on a number of existing findings in cognitive science and social psychology to motivate key hypotheses of our model and help bridge the gap between fields.

Social categorizations are a complex and important aspect of social interactions. The division between “us” and “them” occurs on many attributes, such as race, gender, sexual orientation, and political ideology [6]. On the one hand, these social categorizations fulfill a sense of community and a positive sense of self [7]. On the other hand, they can fuel social conflicts and impact certain groups’ access to economic and social resources [8–10]. For example, the division between White and Black Americans has led to continuing discrimination and segregation long after the abolition of slavery [11]. Recently, the divisions between Democrats and Republicans have created fear and loathing among US voters [12]. In addition,
individuals can also “fall through the cracks” of social categorizations and thus do not belong to any group. We refer to these individuals as *inbetweeners*. Examples could include mixed-race individuals considered neither truly Black nor White by members of either group, or political independents being considered as “other” by both Democrats or Republicans.

Most empirical and theoretical research on social categories in the social sciences presupposes existing classifications instead of studying the process of category formation. For example, most experimental empirical work on individual classifications of race or gender attributes presents participants with predetermined and forced choices [for a review, see 13]. Similarly, theoretical research on categorization tends to use models that assume a category structure in which individuals can be placed. For example, in the exemplar-based model of social judgment [14], an extension of the widely used generalized context model of categorization [15], the categories are supposed to be known, and an individual must belong to one of the existing categories, thus not accounting for inbetweeners. Overall, these strands of research impose categories and do not study the formation of self-organized social groups. As a result, little is known about how self-organized social categories form, and even less on how inbetweeners are perceived.

In this paper, we go beyond the assumption of fixed social categories and use a dynamical systems approach to derive group-level phenomena from individual-level processes. We present a new model of the formation of two social categories on a continuous attribute based on differential-equations. Motivated by social psychological processes such as individuals’ needs for developing a positive sense of self [7, 16] and forming successful collaborations with others [17], we start with the assumption that people want to correctly categorize themselves and others based on their similarity on the continuous attribute dimension, and thus create a boundary between an in-group and an out-group. The model predicts each group’s boundary based on 1) computing the individual categorization accuracy given their characteristics and the social boundary, and 2) optimizing the collective categorization accuracy of the in-group members. Through this individual and group process, a socially accepted boundary is created. This model predicts that social groups draw a boundary more exclusive than the median of the spectrum, and as a consequence, the individuals in the middle of the spectrum are perceived as out-groups by members of both social groups, becoming the inbetweeners in the system.

## 2 The Mathematical Model

We model the formation of two social groups\(^1\) on a continuous attribute. We will derive two governing equations, one for the boundary position of each group. We show in details the derivation for one group, and that for the other group will be similar. The derivation is achieved in two parts. In the first part (section 2.2), we derive the error in categorization accuracy for each individual as a function of their attribute value and the boundary position. In the second part (section 2.3), the group collectively decides the boundary position, by minimizing the collective error of the in-group members.

### 2.1 Model setup

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 \(\rho(x)\). For the group containing the left extreme of the attribute, we denote the boundary position that divides the in-group and out-group to be \(z\) (see Figure 1 for

\(^1\)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.
an illustration of the variables). Individuals with attribute values smaller than $z$ are the in-group, while the out-group is comprised of individuals with values greater than $z$. We denote the similarity between two positions $x_1$ and $x_2$ to be $s(x_1, x_2)$. Similarity is a concept in cognitive science that describes how similar two attribute values are to human perception. Empirically, similarity is a nonlinear decreasing function with the distance between $x_1$ and $x_2$. In this paper, we use the functional form of

$$s(x_1, x_2) = \exp(-c|x_1 - x_2|),$$

where $c$ is the sensitivity parameter—larger $c$ means individuals are more sensitive to differences in attribute space. Variants of this functional form is used in many categorization and social judgment models [e.g., 14] and it is motivated by evidence reviewed in [15, 18]. In appendix A, we show that our major predictions are robust when using one-norm distance as an alternative measure of similarity.

**Figure 1.** Illustration of the variables in the model. The illustration is presented from the perspective of a member of the group on the left, interacting with an individual on the other side of the group boundary. The individual categorization error is the difference between the perceived similarity (derived from perceived distance) and the actual similarity (derived from actual distance).

### 2.2 Derive the individual categorization error.

As described previously, we derive the error in categorization accuracy for each individual as a function of their attribute value and the boundary position. The central insight from decades of research on categorization is that our cognitive system searches for patterns and structures [19]. The perception and cognitive representations of these patterns and structures can take many forms. In line with prototypical theories of category representations [20, 21], we assume a simple 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.

The mathematical representation of this idea is as follows. Let $g_{\text{in}}(z)$ and $g_{\text{out}}(z)$ denote the group positions for the in-group and the out-group, respectively, defined as the center of mass of the population distribution in each group,

$$g_{\text{in}}(z) = \frac{\int_{a}^{z} x \rho(x) dx}{\int_{a}^{z} \rho(x) dx}, \quad \text{and} \quad g_{\text{out}}(z) = \frac{\int_{z}^{b} x \rho(x) dx}{\int_{z}^{b} \rho(x) dx}.$$
Consider a person in the in-group, with position $u$ ($u < z$) on the attribute space. Let $v$ denote a person randomly sampled from the population. The term $s(u, v)$ represents the actual similarity between the individual at $u$ and the sampled person at $v$. The perceived similarity of the two individuals under categorical perception is the similarity between the two group positions: if $v$ is in-group, $s(g_{in}(z), g_{in}(z))$; and if $v$ is out-group, $s(g_{in}(z), g_{out}(z))$ (see figure 1 for an illustration).

We consider that one of the purposes of social categorization is to tell people apart accurately, so that individuals can categorize people different from them as out-group, and people similar as in-group. Several social psychological processes support this hypothesis. Following a Game Theory argument, individuals may want to correctly categorize other people in order to be able to maximize their chances for successful collaborations [17, 22]. Beyond collaborations, individuals may want to correctly determine other people’s similarity levels out of a desire to create a community and a positive sense of self, as described by classic Social Identity Theory [7, 16]. In line with this hypothesis, we define the categorization error between a person at $u$ and an encountered person at $v$ to be the squared difference between the actual similarity and the perceived similarity: if $v$ is in-group, $[s(g_{in}(z), g_{in}(z)) - s(u, v)]^2$, and if $v$ is out-group, $[s(g_{in}(z), g_{out}(z)) - s(u, v)]^2$. The categorization error for the individual at $u$ perceiving all sampled individuals is the integral of these errors with respect to $v$, weighted by the population density $\rho$,

$$err(u, z) = \int_a^z [s(g_{in}(z), g_{in}(z)) - s(u, v)]^2 \rho(v) dv + \int_z^b [s(g_{in}(z), g_{out}(z)) - s(u, v)]^2 \rho(v) dv.$$  \hspace{1cm} (3)

The first term in equation (3) represents the error when the sampled person is in-group ($v < z$). The second term represent the case when the sampled person is out-group ($v > z$, the case illustrated in figure 1). In the first term, the expression can be simplified since $s(g_{in}(z), g_{in}(z)) = 1$. The individual error is a function of both the individual position $u$ and the boundary position $z$.

### 2.3 Derive group categorization error and group boundary.

We present here the second part of the derivation, where the group collectively decides the boundary position. Based on the insights from Game Theory and Social Identity theory described above [7, 16, 17, 22], we assume individuals negotiate to decide on the group boundary, by minimizing the average collective categorization error of the in-group. This assumes that only the in-group utility matters for deciding the location of the group boundary. The average collective in-group error is

$$Err(z) = \frac{1}{\int_a^b \rho(x) dx} \int_a^z err(u, z) \rho(u) du.$$  \hspace{1cm} (4)

Note that because equation (4) calculates the average collective error, the model does not impose any preferences on group size.

Finally, we let the group dynamically adjust its boundary positions to minimize the collective error,

$$\frac{dz}{dt} = -k \frac{dErr(z)}{dz},$$  \hspace{1cm} (5)

where $t$ is time, and $k$ is a constant that sets the time scale of the system. The intuitive understanding of equation (5) is that the category boundary evolves towards the direction that reduces the in-group’s collective categorization error. Equation (5) concludes the derivation of the governing equation for the
group boundary for the group on the left side of the attribute space. By symmetry, the same process determines the group boundary for the group on the right, leading to a similar governing differential equation.

3 Results

3.1 Uniform attribute distribution

![Figure 2](image-url)

**Figure 2.** Stable fixed points of boundary positions for both groups. (a) Solutions in the case of $c = 1$. Green and red show the region seen as in-group by groups 1 and 2, respectively. Light grey shows the regions seen as out-groups by either group. Individuals between the two boundaries are seen as out-groups by members of both social groups (inbetweeners). (b) Solutions as a function of the sensitivity parameter, $c$. The dark grey region represents the inbetweeners. Both panels are for uniform attribute distribution $\rho(x)$ defined for $x$ between 0 and 1.

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. We choose the uniform distribution here because of its mathematical simplicity and will explore more general distributions in the next subsection. We solve $dErr(z)/dz = 0$ numerically (as there is no closed-form expression for the solution), and find that equation (4) has one stable fixed point, suggesting there is a stable boundary position for each social group.

A surprising finding is that the stable boundary position for the two groups do not coincide, and both groups’ preferred boundaries are more exclusive than the median of the attribute space, 0.5. For example,
for \( c = 1 \), the solution for the stable fixed point is \( z^* = 0.30 \). This result means that individuals with positions between 0 and 0.30 are considered in-groups by group 1 (the group including the left extreme of the attribute), and those between 0.30 and 1 are considered out-groups. By symmetry, group 2 (the group including the right extreme of the attribute) has preferred boundary at \( 1 - z^* = 0.70 \), meaning individuals between 0.70 and 1 are considered in-groups, and those between 0 and 0.7 out-groups. This result leads to individuals between 0.3 and 0.7 to be perceived as out-groups by both groups, whom we refer to as inbetweeners (see panel (a) of figure 2 for an illustration). Panel (b) of figure 2 shows solutions of boundary positions of the two groups (\( z^* \) and \( 1 - z^* \)) as a function of \( c \). As \( c \) increases, meaning people more sensitive to differences in the attribute space, the social group boundaries become more exclusive. For all \( c > 0 \), the fixed point satisfies \( z^* < 0.5 \), meaning inbetweeners occur for all values of \( c \).

### 3.2 Beta attribute distribution

![Figure 3](image)

**Figure 3.** Solutions of the fixed points of boundary positions (\( z \)) for both groups as a function of \( c \), for Beta distribution \( \rho(x) \) defined for \( x \) between 0 and 1. Panels (a) and (b) show the distribution \( \rho(x) \) that generated the results in panels (c) and (d), respectively. The shape parameters used are (a): \( \alpha = 2, \beta = 2 \); (b): \( \alpha = 2, \beta = 4 \).

The occurrence of inbetweeners is not unique to the uniform attribute distribution. Here we show results obtained considering the attribution distribution \( \rho(x) \) to be a Beta distribution. The Beta distribution is parameterized by two positive shape parameters, \( \alpha \) and \( \beta \), with probability density function \( f(x) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha, \beta) \), 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 attribution distribution is known to be unimodal, such as political ideology of the US public, measured by positions on public policy issues [23].

Panels (a) and (b) in figure 3 show two examples of the Beta distribution as attribute distribution $\rho(y)$, one symmetrical and one asymmetrical. We solve for the fixed points of equation (5) and their stability for given $c$ values and the shape parameters $\alpha$ and $\beta$. In panels (c) and (d) of figure 3, we show numerical solutions for the fixed points boundary positions and their stability given the two attribute distribution in panels (a) and (b), respectively. In both panels of figure 3, the inbetweeners appear. Note that in the upper regions of panel (d), only one social group forms. As $c$ increases, the group boundaries, marked by the stable fixed points, become more exclusive, similar to the results in the uniform distribution case (shown in figure 2). Different from the uniform distribution results, the dynamical system has a stable and an unstable fixed point for $c$ smaller than a critical value. As $c$ exceeds this critical value, a saddle-node bifurcation occurs, and the system has no fixed points for $c$ greater than the critical value. This behavior predicts that when people are too sensitive to small differences among people, categorization along this attribute ceases to exist.

3.3 Validate model’s prediction with empirical findings

Our model predicts that those in the middle of the attribute space are seen as out-group by both social groups. In this section, we first perform analysis on empirical data from a US political survey to verify our prediction, then summarize existing experimental studies on the perception of racially ambiguous individuals to show qualitative support.

3.3.1 Compare model’s prediction with political surveys.

![Figure 4](image_url)

**Figure 4.** The mean thermometer values (reflecting feeling favorably or unfavorably) towards both political parties and political independents reported by registered party members from ANES data. 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.

To test the prediction of inbetweeners, we use the American National Election Studies (ANES) dataset.
This dataset is a large nationally representative survey of political attitudes among the US public, reported on the individual level. In our analysis, we utilize the self-reported party registration as well as a set of thermometer questions that evaluate whether people feel favorably or unfavorably toward a number of social groups (such as 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 appendix B for 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. The data we utilize contain 1923 individuals in total.

We use registered Democrat and Republican party members to represent the two groups on opposite sides of a continuous spectrum. In order to measure whether individuals are seen as in- or out-group, we compare thermometer scores given to individuals clearly in the in-group (e.g., Republicans for Republicans), in the out-group (e.g., Republicans for Democrats), and politically independents. We want to test if independents are perceived by both parties as part of the out-group (low thermometer score similar to the other party), in-group (high thermometer score similar to own party), or somewhere in between (neither high or low, between other and own parties). If feelings towards independents are similar to feelings towards other party, then the data supports our models’ prediction.

Figure 4 shows the mean thermometer values for political independents, Democrats, and Republicans, reported by registered members of both parties. For both registered Democrats and Republicans, political independents are perceived similarly to members of the other party, while members of 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 those for political independents are indistinguishable ($p = 0.42$). The same test shows that for registered Democrats, the mean value for Republicans is slightly higher than that of the political independents ($p < 0.001$). The own party is perceived significantly more favorably than both the other party and the independents ($p < 0.001$).

### 3.3.2 Perception of racially ambiguous individuals in existing empirical studies.

Our model’s prediction is also in agreement with experimental 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, and confirmed in a number of empirical studies [13, 24]. First, using established racial categories, perceivers in the U.S. tend to categorize racially ambiguous individuals as the out-group [25, 26], which was replicated in South Africa [27] and Italy [28]. Second, using memory tests, an experimental study [29] found that racially ambiguous faces are perceived as out-groups by mono-racial individuals. Finally, using open-ended categorization, a recent study pointed out that perceivers use a third category (in this case, Hispanic or Middle Eastern) for racially-ambiguous individuals (who were mixed Black and White) [30]. These studies show that racial groups tend to draw boundaries that exclude individuals of mixed races, supporting our model’s prediction of inbetweeners.

### 4 Discussion

Using a simple differential-equation-based model, we find that inbetweeners, those in the middle of the attribute space and excluded by both social groups, can arise as the result of social groups drawing boundaries more restrictive than the median of the attribute spectrum, driven solely by a desire to improve
categorization accuracy. Our theoretical finding is supported by empirical analysis of feelings towards politically independent individuals reported by registered Democrats and Republicans in the ANES data, as well as by previous empirical findings surrounding the in-group over-exclusion effect in racial categorization [13, 24].

Besides these extensions, social categorization can involve other subtle individual-level processes in practice. Empirical studies find that when individuals are seen as ambiguous on one attribute dimension, the perceiver may switch to other dimensions to make its categorization [13]. For example, individuals use cues on facial characteristics and clothing to categorize gender ambiguous people [31, 32]. Moreover, the characteristics of the perceiver can influence the categorization process. Individuals who want to retain their social status, who have a weak identity, or do not often interact with other groups tend to have a stricter definition of in-group [13, 28]. Besides subtle individual-level processes, the cultural and historical background also shape categories. For example, a salient categorization dimension in the US and South Africa is race, due to a history of segregation, slavery, and mistreatment. However, the salient categorization dimensions in Europe are different, such as religion or language [33]. Social boundaries are not only different between cultures, but also evolve. In the US, the growth in immigration from Latin America and Asia since 1965 has redefined the white-black dichotomous categorization towards a tri-racial system, separating Whites, honorary Whites (Asian-Americans, and some Latinos), and Blacks [34]. This is one example when the changing distribution of attributes leads to changing group categories.

The dynamical systems framework we utilized would be appropriate, in future research, to explore how cultural changes and shifting attribute distributions lead to changes in social category boundaries.

Only a small volume of empirical studies and even fewer theoretical studies account for individuals who do not belong to any social categories. Our model provides a rare theoretical result on how inbetweeners can arise in social categorization. Although this work dominantly uses data on political ideology and attitudes, we think the process modeled can be generalized to many other social categories, and our result of inbetweeners may be extended more generally to individuals in the middle region of attribute spaces, such as those who are mixed race, gender non-binary, or in interdisciplinary scientific fields. With demographic shifts (more than ten million individuals in the US identify with two or more races in 2017 [35]), the increasing visibility of gender non-binary individuals [36], and the promise of new interdisciplinary sciences, understanding how these individuals are perceived becomes increasingly important. However, data gathering about these individuals is minimal, especially in the census and large-scale population surveys, making it hard to further develop scientific theories about them. We encourage more data-gathering efforts on studying inbetweeners that are not accepted by established social categories.

Much significance of social categories is not in the category themselves, but in how these categories affect how individuals are perceived and treated. Our model suggests that individuals with characteristics in the middle of the attribute space “fall through the cracks” and are seen as out-group by social groups at both extremes. We believe this result can possibly affect many social processes. One speculative example is the greater perceived polarization in the US public. Among the US public, the policy position distribution has remained dominantly moderate; however, the perception of mass polarization increased in the last a few decades due to individuals increasingly sorted into political groups and greater antipathy between these groups [12, 23, 37]. A possible explanation of this paradox is through the political independents perceived as out-group of both parties, which is predicted by our model and reflected in the data. Our hypothesis is that motivated by the need for belonging and community, individuals holding moderate political positions can be motivated to identify with one of the two increasingly polarized political parties,
despite misalignment on issue positions, leading to increased sorting. Empirically testing how political independents are perceived, and how this relate to sorting and perceived mass polarization, can be an important piece of future research. One difficulty in testing this hypothesis is the lack of data. As we noted previously, the attitude surveys on political independents were only included in the 1980 and 1984 ANES Survey and since discontinued. We encourage gathering more empirical data on how political independents are perceived over time to enable testing of this hypothesis.

Our work comes from the perspective of the applied mathematics community, and dynamical systems in particular. This community has seen increasing interest and success in using its quantitative tools to address phenomena in human society. In responding to the call for incorporating interdisciplinary methods to study human social behavior [38], our work offers insight into the process of social categorization through synthesizing quantitative tools in dynamical system with empirical findings in cognitive science and sociology. We hope our work serves as an example of bridging the gap between the communities in applied mathematics and those in the social sciences and will inspire further effort from the applied mathematics community connecting with existing findings in the social sciences.

**Data accessibility.** All data used in our analysis are publicly available from the web link provided in appendix B.

**Authors’ contributions.** All authors designed the research. VCY developed the mathematical model, performed numerical simulations and data analysis. All authors interpreted the results and contributed to writing the paper.

**Competing interests.** We declare we have no competing interests.

**Funding.** We thank the Santa Fe Institute for meeting support. VCY acknowledges funding from the Santa Fe Institute Omidyar Fellowship.

**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.

### Appendices

#### A Distance-based model

In the main text, we choose similarity as the fundamental quantity because of the abundant empirical support in cognitive science. Here we present an alternative model and show that the qualitative result of inbetweeners is robust using another similarity metric, the one-norm distance.

Similar to equation (3) of the main text, the categorization error for an individual at position $u$ (considering group on the left, $u < x$), is

$$
err(u, z) = \int_a^z (|g_{in}(z) - g_{in}(v)| - |u - v|)^2 \rho(v) dv + \int_z^b (|g_{out}(z) - g_{in}(z)| - |u - v|)^2 \rho(v) dv.
$$

(A1)

Simplify equation (A1), using the fact that $g_{out} > g_{in}$, we have

$$
err(u, z) = \int_a^z (u - v)^2 \rho(v) dv + \int_z^b (g_{out}(z) - g_{in}(z) - (v - u))^2 \rho(v) dv.
$$

(A2)
The collective average in-group error for the group on the left is

\[ Err(z) = \frac{1}{\int_a^z \rho(x) dx} \int_a^z err(u,z) \rho(u) du . \]  

(A3)

The dynamical system for the boundary position is,

\[ \frac{dz}{dt} = -k \frac{dErr(z)}{dz} . \]  

(A4)

For an uniform distribution \( \rho(x) \), defined on \( x \) in the interval \([0,1] \), we can analytically calculate the collective error,

\[ Err(z) = \frac{1}{3} z^2 - \frac{1}{4} z + \frac{1}{12} . \]  

(A5)

Equation (A4) has one stable fixed point, \( z^* = 3/8 = 0.375 \), meaning the boundary position for the group on the left side of the spectrum sets its group boundary at 0.375: this group considers those with attribute value \( x < 0.375 \) as in-group, and \( x > 0.375 \) as out-group. A same set of equations can be derived for the group on the right. By symmetry, the preferred group boundary of the group on the right is \( 1 - 0.375 = 0.625 \). This leads to individuals between 0.375 and 0.625 being considered out-group by both social groups. Results from this distance-based model show that the prediction of outsiders is not unique to the choice of similarity in our model. The prediction of inbetweeners is a more fundamental result arising from groups setting more restrictive boundaries than the median of the spectrum in the categorization process.

**B American National Election Studies Data**

The American National Election Studies data used in this paper is the cumulative data file of 1940-2016, May 31, 2018 version. It was downloaded from [https://electionstudies.org/project/anes-time-series-cumulative-data-file/](https://electionstudies.org/project/anes-time-series-cumulative-data-file/) on Oct 4, 2018.

The phrasing of the thermometer questions is as follows: “We’d also like to get your feelings about some groups in American society. When I read the name of a group, we’d like you to rate it with what we call a feeling thermometer. Ratings between 50 and 100 degrees mean that you feel favorably and warm toward the group; ratings between 0 and 50 degrees mean that you don’t feel favorably towards the group and that you don’t care too much for that group. If you don’t feel particularly warm or cold toward a group you would rate them at 50 degrees. If we come to a group you don’t know much about, just tell me and we’ll move on to the next one. Using the thermometer, how would you rate the following”. The three groups used in our analysis are “Democrats”, “Republicans”, and “political independents”.

**References**

1. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* 81, 591 (2009).
2. Abrams, D. M. & Strogatz, S. H. Linguistics: Modelling the dynamics of language death. *Nature* **424**, 900 (2003).

3. Abrams, D. M., Yaple, H. A. & Wiener, R. J. Dynamics of social group competition: Modeling the decline of religious affiliation. *Phys. Rev. Lett.* **107**, 088701 (2011).

4. Lu, X., Gao, J. & Szymanski, B. K. The evolution of polarization in the legislative branch of government. *The Royal Soc. Interface* **16**, 20190010 (2019).

5. Jeffs, R. A., Hayward, J., Roach, P. A. & Wyburn, J. Activist model of political party growth. *Phys. A: Stat. Mech. its Appl.* **442**, 359–372 (2016).

6. Lamont, M. & Molnár, V. The study of boundaries in the Social Sciences. *Annu. Rev. Sociol.* **28**, 167–195 (2002).

7. 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.

8. Ashmore, R. D., Deaux, K. & McLaughlin-Volpe, T. An Organizing Framework for Collective Identity: Articulation and Significance of Multidimensionality. *Psychol. Bull.* **130**, 80–114 (2004).

9. 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).

10. Roccas, S., Sagiv, L., Schwartz, S., Halevy, N. & Eidelson, R. Toward a unifying model of identification with groups: Integrating theoretical perspectives. *Pers. Soc. Psychol. Rev.* **12**, 280–306 (2008).

11. Fox, C. & Guglielmo, T. A. Defining America’s racial boundaries: Blacks, Mexicans, and European immigrants, 1890-1945. *Am. J. Sociol.* **118**, 327–379 (2012).

12. Iyengar, S. & Westwood, S. J. Fear and loathing across party lines: New evidence on group polarization. *Am. J. Polit. Sci.* **59**, 690–707 (2015).

13. Bodenhausen, G. V., Kang, S. K. & Peery, D. Social Categorization and the Perception of Social Groups. In *The Sage Handbook of Social Cognition*, 318–336 (Sage Publications London, 2012).

14. Smith, E. R. & Zarate, M. A. Exemplar-based model of social judgment. *Psychol. Rev.* **99**, 3–21 (1992).

15. Nosofsky, R. M. Choice, similarity, and the context theory of classification. *J. Exp. Psychol. Learn. Mem. Cogn.* **10**, 104–114 (1984).

16. Cerulo, K. A. Identity Construction: New Issues, New Directions. *Annu. Rev. Sociol.* **23**, 385–409 (1997).

17. Smaldino, P. E., Pickett, C. L., Sherman, J. & Schank, J. An Agent-Based Model of Social Identity Dynamics. *J. Artif. Soc. Soc. Simul.* **15**, 1–17 (2012).

18. Shepard, R. Toward a universal law of generalization for psychological science. *Science* **237**, 1317–1323 (1987).
19. Rosch, E. & Mervis, C. B. Family resemblances: Studies in the internal structure of categories. *Cogn. Psychol.* 7, 573–605 (1975).

20. Posner, M. I. & Keele, S. W. On the genesis of abstract ideas. *J. Exp. Psychol.* 77, 353–363 (1968).

21. Rosch, E. H. Natural categories. *Cogn. Psychol.* 4, 328–350 (1973).

22. Smaldino, P. E. Social identity and cooperation in cultural evolution. *Behav. Process.* 0–1 (2018).

23. Fiorina, M. P. & Abrams, S. J. Political polarization in the American public. *Annu. Rev. Polit. Sci.* 11, 563–588 (2008).

24. Leyens, J. P. & Yzerbyt, V. Y. The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. *Eur. J. Soc. Psychol.* 22, 549–569 (1992).

25. Gaither, S. E., Pauker, K., Slepian, M. L. & Sommers, S. R. Social belonging motivates categorization of racially ambiguous faces. *Soc. Cogn.* 34, 97–118 (2016).

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

27. Pettigrew, T. F., Allport, G. W. & Barnett, E. O. Binocular resolution and perception of race in South Africa. *Br. J. Psychol.* 49, 265–278 (1958).

28. Castano, E., Yzerbyt, V., Bourguignon, D. & Seron, E. Who may enter? The impact of in-group identification on in-group/out-group categorization. *J. Exp. Soc. Psychol.* 38, 315–322 (2002).

29. Pauker, K. *et al.* Not so black and white: Memory for ambiguous group members. *J. Pers. Soc. Psychol.* 96, 795 (2009).

30. Nicolas, G., Skinner, A. L. & Dickter, C. L. Other than the sum: Hispanic and Middle Eastern categorizations of Black–White mixed-race faces. *Soc. Psychol. Pers. Sci.* 10, 532–541 (2019).

31. Freeman, J. B., Rule, N. O., Adams, R. B. & Ambady, N. The neural basis of categorical face perception: Graded representations of face gender in fusiform and orbitofrontal cortices. *Cereb. Cortex* 20, 1314–1322 (2010).

32. Macrae, C. N. & Martin, D. A boy primed Sue: Feature-based processing and person construal. *Eur. J. Soc. Psychol.* 37, 793–805 (2007).

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

34. Bonilla-Silva, E. From bi-racial to tri-racial: Towards a new system of racial stratification in the USA. *Ethn. Racial Stud.* 27, 931–950 (2004).

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

36. Schilt, K. & Lagos, D. The Development of Transgender Studies in Sociology. *Annu. Rev. Sociol.* 43, 425–443 (2017).
37. Hill, S. J. & Tausanovitch, C. A disconnect in representation? Comparison of trends in congressional and public polarization. *The J. Polit.* **77**, 1058–1075 (2015).

38. Gintis, H. Towards the unity of the human behavioral sciences. *Polit. Philos. & Econ.* **3**, 37–57 (2004).