When pull turns to shove: A continuous-time model for opinion dynamics

David Sabin-Miller[*]

Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois 60208, USA

Daniel M. Abrams

Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois 60208, USA;
Northwestern Institute for Complex Systems, Northwestern University, Evanston, Illinois 60208, USA;
and Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA

Accurate modeling of opinion dynamics has the potential to help us understand polarization and what makes effective political discourse possible or impossible. Here, we use physics-based methods to model the evolution of political opinions within a continuously distributed population. We utilize a network-free system of determining political influence and a local-attraction, distal-repulsion dynamic for reaction to perceived content. Our approach allows for the incorporation of intergroup bias such that messages from trusted in-group sources enjoy greater leeway than out-group ones. We are able to extrapolate these nonlinear microscopic dynamics to macroscopic population distributions by using probabilistic functions representing biased environments. The framework we put forward can reproduce real-world political distributions and experimentally observed dynamics and is amenable to further refinement as more data becomes available.

I. INTRODUCTION

The field of opinion dynamics seeks to understand the evolution of ideas in populations, a complex interdisciplinary endeavor which has attracted a wide variety of approaches from different disciplines. After early mathematical groundwork [1], the growth of network science has led to a boom in models which utilize neighbor-based update rules to examine long-term outcomes for opinion distributions, such as polarization and consensus, e.g., Refs. [2–19]. Other researchers have advanced "sociophysics" approaches, which apply techniques from statistical physics to analyze analogous social systems [20–30]. Complementary to these modeling approaches, theoretical and empirical work from economics and social science has examined the political bias of media entities [31,32] and their influence on a population [33–37]. All these approaches contribute valuable insight toward an understanding of this complex topic, but the disparities between their perspectives make direct cohesion a challenge.

Our model takes a different approach, which we believe achieves the key benefits of previous models while expanding flexibility and retaining the ability to incorporate real-world data as it becomes available. One key structural choice we make is to modularize the process of opinion change by breaking it into two parts: perceptions and reactions.

In our model, individuals perceive a probabilistic mix of politicized experiences which depends on their ideology and party. This might be thought of as the continuum limit of a network approach, where influences are so numerous and varied that interactions are best characterized by a probability distribution rather than explicit neighboring agents. This approach also allows us to encapsulate broader societal influences such as politicized media environments, since individuals' perceptual mix may be constantly changing to reflect their changing world view.

We model individuals' reactions to these perceptions by having their ideology evolve in continuous time. This is governed by ordinary and stochastic differential equations which depend on their current position and their perceptual distribution. Together, these perception and reaction modules capture a feedback loop between individuals' current beliefs, the biased "slice" of the political world they perceive, and how they update those beliefs as a result.

A. Political spectrum

Like many prior approaches (e.g., Refs. [2,3,10,13,31,34,35,38–40]), we consider a single, finite ideology axis. This appears to be supported for the U.S. political environment dominated by two parties: Empirical results show that the liberal-conservative dimension captures the great majority of modern U.S. legislative behavior [41]. However, it's unclear how appropriate this assumption is for other countries with differing political systems: Multiple dimensions may be warranted (see Possible Extensions section).

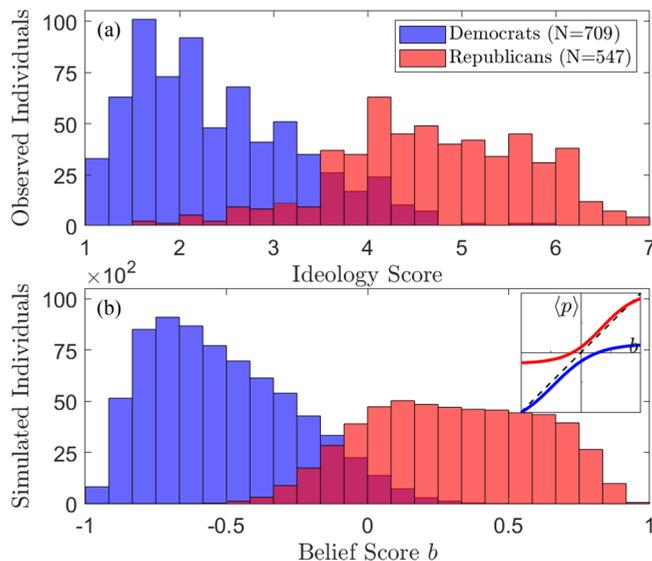[*]davidsabinmiller@u.northwestern.edu

FIG. 1. (a) Empirical ideological distributions by U.S. political party. Average ideological position score from 1 (strongly liberal) to 7 (strongly conservative) on social, economic, and military issues for 1256 U.S. Twitter users. Data from Ref. [38]. (b) Model predictions. Steady state for our simulated population of 70 900 Democrats and 54 700 Republicans, with party perception curves shown in the inset. See Results section for details.

Figure 1(a) shows a one-dimensional projection of political ideology for the U.S. population based on one study [38]; though the precise methods of projecting the political landscape onto one axis differ between sources, other recent reports like that of Pew Research [40] show good qualitative agreement. We will use the term *belief score b* to refer to an individual's ideological position between -1 (extreme liberal) and $+1$ (extreme conservative). We abstract all politically-opinionated information an individual is exposed to (hereafter termed *percepts p*) onto this same axis, so that a percept of $p = +0.5$ is in support of belief score $+0.5$ (conservative), a percept with value $p = 0$ argues for a neutral stance, and so on. Due to the imprecise nature of any measurement on this scale (it's a projection of a highly abstract space that can be quantified in different ways), qualitative results should be robust to small changes in these values.

B. Opinion change

Classic "bounded-confidence" models (e.g., Refs. [2,3]), which allow for individuals to interact only with others who are relatively like-minded, have been used to capture the effect of homophily on interaction. But political issues are contentious and are often brought up between those who disagree and are easily suffused with negative emotional affect rather than agreement or indifference. Repulsion from disliked positions seems to be an important determinant in swing voters: A recent Pew survey [42] found that U.S. independents supporting one of the political parties did so mostly due to negative perceptions of the *other* party. So like some other extensions to bounded-confidence models (e.g., Refs. [10,13]), we supplement local-attraction behavior with distal repulsion: Individuals who are exposed to ideas which
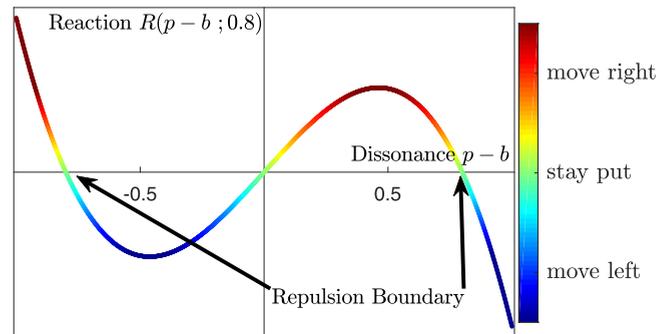

FIG. 2. Example reaction function. Here we show a cubic reaction function, where an individual's reaction depends on dissonance $p - b$. Vertical scale has arbitrary units: The magnitude of this movement depends on time constant τ and current belief score b. For this image a repulsion distance of $d = 0.8$ was chosen.

are too different from their own will not be attracted but rather be *repelled* from the espoused position of the source. There is experimental evidence that this can be a very potent and real source of ideological movement: In recent work from Bail *et al.* [38], it was found that exposure to 24 tweets per day from prominent members of the opposing party can have a significant repulsive effect over the course of a month, even among all other political inputs received by the participants (self-identified politically active Twitter users).

II. METHODS

The first key component of our model is the *reaction function*. This is a continuous function which relates an individual's shift in ideological belief to the difference between a perceived political opinion (the percept p) and the individual's own belief b; we will refer to this difference $p - b$ as the *dissonance*. A repulsion effect will be modeled through the existence of a *repulsion distance d* such that percepts less dissonant than d will be attractive and percepts more dissonant than d will be repulsive. This parameter d can be allowed to vary depending on the context of the message, which will allow us to model the important effect of intergroup bias: For example, a somewhat challenging position can be repulsive when it comes from a disliked source but attractive when introduced by a member of one's in-group (see "Adding Intergroup Bias" below).

One simple form for a reaction function that satisfies the above conditions employs a cubic dependence on dissonance:

$$R(p - b; d) = (p - b)\left[1 - \frac{(p - b)^2}{d^2}\right], \tag{1}$$

shown visually in Fig. 2. We utilize this cubic reaction function for all our models in this paper, but we expect similar results with any qualitatively similar function used in its place (ideally one inferred from experimental data).

To organically constrain belief dynamics to a bounded domain (in our case, $[-1, 1]$), we temper the above reaction function with a multiplicative factor $(1 - b^2)$. This has the effect of gradually damping motion near the extremes—thus we interpret the ± 1 boundaries of our finite ideology scale to be asymptotic extremes that are only approachable, not

attainable. We also scale the dynamics by a time constant τ which controls the speed of belief change. Then, for an individual j with belief score b_j and repulsion distance d, exposed to percept p (which may depend on many factors), we arrive at the following differential equation for ideological dynamics:

$$\tau \frac{db_j}{dt} = \left(1 - b_j^2\right) \left\{ (p - b_j) \left[1 - \frac{(p - b_j)^2}{d^2} \right] \right\} . \quad (2)$$

A. Perceptual diets

An important question remains: Which individuals are exposed to which messages? The vast majority of work on opinion dynamics has been in a network context, wherein agents update their opinions according to a rule incorporating the positions of some other agent(s) (e.g., Refs. [1–7,9,10,13,15,21–23,30]). Our approach sidesteps the need for constructing explicit influence networks, which are difficult to capture due to the many modalities of human interaction. Instead we suppose that an individual's party affiliation and current political position determine their perceived "slice" of the political world—a probability distribution of political experiences, $\rho(p)$ [43]. This continuum approach allows us to personalize political environments to account for "media bubbles" and other biased environments even without a network and is easily scaled to large populations.

B. Toy models

1. Simplest model

For the simplest concrete implementation of our framework, we might suppose a single-party population is initially distributed across the belief spectrum but is otherwise homogeneous and that every individual perceives the same delta distribution of political content, the constant percept $p = C$. Then upon choosing a repulsion distance d we can exactly determine long-term behavior of the entire group—there will be a single flow function that affects the whole belief spectrum:

$$\tau \frac{db_j}{dt} = \left(1 - b_j^2\right) \left\{ (C - b_j) \left[1 - \frac{(C - b_j)^2}{d^2} \right] \right\} . \quad (3)$$

This ordinary differential equation (ODE) has fixed points at $b_j = C$, $b_j = C \pm d$, and $b_j = \pm 1$ (due to the imposed domain bounds). The fixed point at $b_j = C$ is stable, and stability of the other points alternates.

For example, if we use the cubic reaction function from Fig. 2 above and set $d = 1$, $C = 0.25$, then that party's population experiences differential movement as shown in Fig. 3. Fixed points exist at $\{-1, -0.75, 0.25, 1, 1.25\}$ (though beliefs are constrained to the $[-1, 1]$ domain, so the theoretical fixed point at 1.25 is not meaningful). Given time, all observers between -0.75 and 1 would congregate at 0.25, and all observers starting left of -0.75 would converge to -1. This small segment of the population—the members that are liberal enough to be repelled by the "party line"—might be likely to switch parties in favor of one with more comfortable percepts, though we do not include such party-switching dynamics in this initial model.
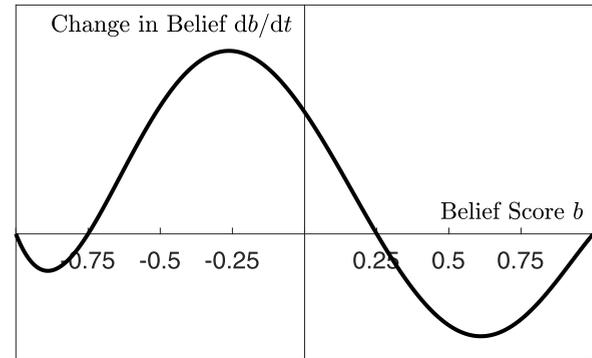

FIG. 3. Flow diagram. Example of differential movement for a population uniformly exposed to a percept with score $+0.25$ assuming repulsion distance 1 [see Eq. (3)]. Vertical axis scaling is arbitrary.

2. Adding intergroup bias

We would also like our modeling framework to accommodate the tendency for individuals to be more receptive to information from those whom they perceive as allies, i.e., part of their "in-group" [44]. For the simplest case, we modify our previous model by adding an "out-group" with its own distinct constant "party line" percept p_o. Now percepts have a party identity attached to them, and we allow individuals to consume a mixed diet of in-group and out-group information, at belief scores of p_i and p_o, respectively. We set repulsion distances d_i and d_o for in-group (e.g., U.S. Republican) and out-group (e.g., U.S. Democrat) messengers, with $d_o \leqslant d_i$. We can set a fixed fraction f for in-party content or allow for a belief-dependent skew $f(b)$ such that, e.g., liberal Republicans view a higher fraction of Democratic content than their conservative party mates. The average flow function db/dt is then a simple weighted average of the flow functions in Eq. (2) due to each source:

$$\tau \frac{db}{dt} = (1 - b^2)[f R_i + (1 - f) R_o] , \quad (4)$$

where in general $f = f(b)$, $R_i = R(p_i - b; d_i)$, and $R_o = R(p_o - b; d_o)$.

To understand the flow in this case, it is informative to consider the purely in-group and purely out-group situations ($f = 1$ or 0, respectively), because all fractional perceptual "diets" are interpolated between them (see Fig. 4). We note that exposure to some out-group content can in some cases *increase* polarization for a small extreme group—for example, in Fig. 4, individuals starting with $b > 0.75$ will on average move *rightward* when exposed to percepts from a 70%/30% combination of in-group and out-group sources, respectively (solid curve), whereas those same individuals would move *leftward* if presented with in-group information alone (dotted curve). This simple example shows how exposure to—and rejection of—opposing content can have a polarizing influence on a population.

Note that we assume that this "tribal" bias only affects the *reaction* to content, not its subjectively perceived ideological score p. However, the inclusion of such an additional bias effect is reasonable and may be handled with a slight increase to model complexity (see Possible Extensions section).
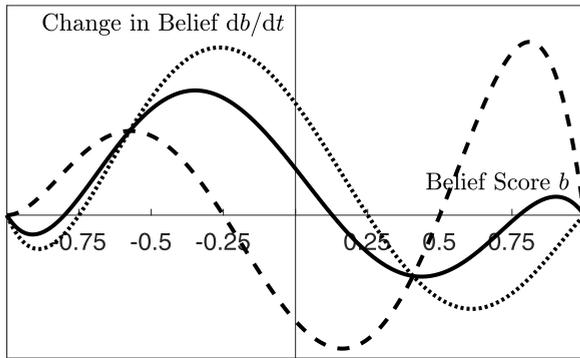
FIG. 4. Flow with different messengers. The flow functions for in-group (dotted) and out-group (dashed) messages of $p_i = +0.25$ and $p_o = -0.25$ with repulsion distances of 1 and 0.75, respectively. The solid curve is the net flow if individuals are exposed to 70% in-group and 30% out-group percepts. Vertical axis scaling is arbitrary.
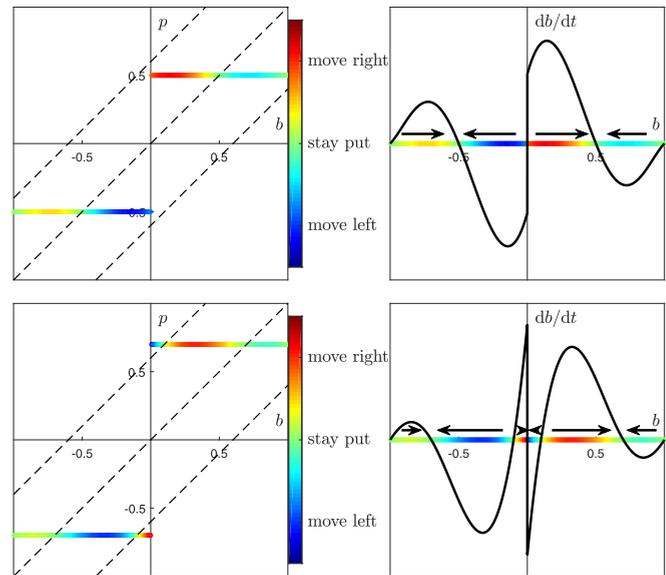

FIG. 5. Two-take world. Graphical analysis of step-function perception curves. Left panels: perception curves color coded for the movement induced, along with dashed $p = b$ line and repulsion boundaries. Right panels: Projection of that flow-velocity color onto the belief axis, compared with the exact population flow calculated from Eq. (2) (black curve). Top row: When perception curves lie within the trust region, we see two attractors at the "party line" belief values. Bottom row: With more extreme "party lines," centrists are repelled by either party position, creating a stable central attractor.

3. Adding personalized perceptions

Putting aside the in-group/out-group balance for a moment, we might expand our simplest model in a different way: by linking individuals' perceptions to their current beliefs via a "perception curve" $p(b)$, which indicates what content they see as a function of position. This reflects the differing "slices" of the political world that individuals see as a result of the differing environments and personal biases that accompany their ideologies.

In our simplest model, where $p = C$, the perception curve is a horizontal line in b vs p space; individuals at all b values perceive the same thing. In a hypothetical "perfectly targeted" world, the perception curve would be the 45° line $p = b$, and nobody would change belief because each person would perceive content perfectly in line with their current world view.

Luckily, we don't need to privilege one such curve in particular—a graphical analysis method lets us combine *any* perception curve with the reaction function and read off a (qualitative) flow for each segment of the population. To do this, we plot the perception curve $p(b)$ and overlay the 45° line for reference—any time the perception curve intersects it, the individuals at that belief score are stationary, since their perceptions are in agreement with their current beliefs. If the perception curve is slightly above the 45° line, individuals with those beliefs are perceiving something slightly more conservative than their own views, and move right. Similarly, people move left wherever the perception curve is slightly below the 45° line.

We also overlay the repulsion boundaries at distance d above and below that $p = b$ line. If the perception curve exits the resulting "trust band" over some b interval, that segment of the population is repelled and moves the opposite direction from what would be expected based on small deviations from the 45° line.

It is then straightforward to determine the qualitative behavior of the whole population given any perception curve $p(b)$ by visually examining intersections of the perception curve with the 45° lines, as in Fig. 5 (left panels). With a closed form expression for $p(b)$, we can use Eq. (2) to obtain an exact flow function (black curves on right panels of Fig. 5)

and confirm our qualitative analysis. If multiple parties are present, this analysis is performed separately for each, and the resulting reactions are combined as in Eq. (4).

The real benefit of this graphical approach is its generality; one can draw any perception curve one would like and simply read off the fixed points and stability. Whenever the perception curve crosses the diagonal with slope less than one, that crossing becomes a stable fixed point. Whenever it crosses with a slope greater than one, the crossing becomes an unstable fixed point instead. If the perception curve crosses a repulsion boundary, shallow crossings create unstable points and steep crossings create stable ones.

While the choice of perception curve entails a large degree of modeling freedom, based on our graphical analysis reasoning we know our model's qualitative predictions aren't particularly sensitive to the choice. Ideally, real-world data could (and should) be used to construct such a curve (e.g., by evaluating the partisan positions of news sources and other political influences experienced by individuals across the political spectrum), though we leave this for future work (see Future Work subsection below).

4. Adding heterogeneity

To move toward a more realistic scenario, we must allow for heterogeneity of both environments and individuals. We can introduce random variation in two distinct components of the model: perceptions (so individuals are exposed to a range of different inputs rather than a single determined value) and the reaction function (so otherwise identical individuals
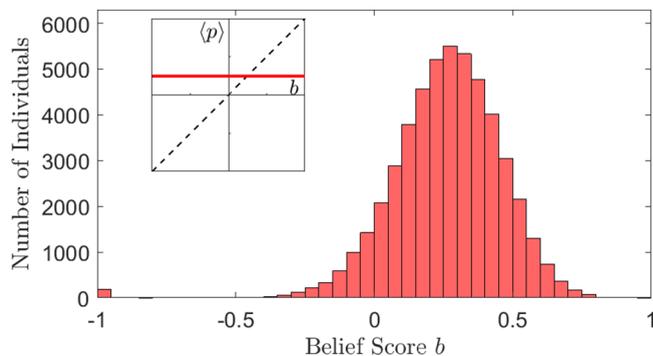
FIG. 6. Simulated population distribution. The stable population state induced by flow function Eq. (3) with added Gaussian noise ($\tau = 1$, $dt = 0.001$, $\sigma = 0.25$, $N = 50\,000$). Population was initialized to match Republicans from Bail [38]. Inset: Perception curve, the constant $C = 0.25$. All variation is in reaction.

can react differently to the same percept). For the latter, we add Gaussian noise to the reaction function R, which causes the stable fixed points from our prior analysis to expand into finite-width stable distributions; these may be estimated easily and accurately by Euler-Maruyama numerical integration of our now-stochastic differential equation (SDE). For example, with the conditions in Fig. 6, the main body of the party congregates around the primary attractor at 0.25, and a small group is repelled to -1 [45].

If we wish to add variability to the percept instead of the reaction, nonlinear effects become more important, since p—now properly a probability distribution $\rho(p)$—must be fed through reaction function $R(p - b)$ before its effects are determined. We suppose that percepts occur on a significantly faster time scale than opinion change itself, so that the net effect of the perceptual diet is a weighted average over all possible percepts, which for smooth percept distributions becomes an integral of $R(p - b)$ against $\rho(p)$. Regardless of the shape of perception distribution, we can still use our graphical analysis technique to solve for net opinion drift, though the repulsion boundaries may warp, as we show in Appendix A.

5. Full model

Taking all these effects together, the model has the following structure:

$$R(p - b; d) = (p - b)\left[1 - \frac{(p - b)^2}{d^2}\right] \tag{5a}$$

$$v_{\text{in}} = \int_{-1}^{1} R(p - b; d_{\text{in}})\, \rho_{\text{in}}(p; b, \sigma_p)\, dp \tag{5b}$$

$$v_{\text{out}} = \int_{-1}^{1} R(p - b; d_{\text{out}})\, \rho_{\text{out}}(p; b, \sigma_p)\, dp \tag{5c}$$

$$\tau\, db = (1 - b^2)\{[f v_{\text{in}} + (1 - f) v_{\text{out}}] dt + \sigma_r dW\} \tag{5d}$$

with ρ_{in} and ρ_{out} as the perceptual distributions for in-group and out-group content, respectively (with standard deviation σ_p), and v_{in} and v_{out} the opinion drift due to those influences. This framework allows examination of the long-term impact of a probabilistic content-generating environment—which differentiates based on party and ideology—on a population

which reacts to what they experience in a nonlinear manner heavily influenced by identity.

III. RESULTS

The full model, with some reasonable parameter assumptions, exhibited equilibrium distributions which agree well with real-world observations (see Fig. 1). To represent a "media bubble" effect, we scaled in-group fraction linearly and symmetrically from 0.5 (for $b = +1$ Democrats and $b = -1$ Republicans) to 0.9 (for $b = -1$ Democrats and $b = +1$ Republicans):

$$f_D(b) = 0.7 + 0.2b, \tag{6a}$$

$$f_R(b) = 0.7 - 0.2b. \tag{6b}$$

Equations (5a), (5b), (5c), and (5d) determined population movement over time, simulated using Euler-Maruyama numerical integration. Perceptual diets were beta distributions bounded by $[-1, 1]$, with standard deviation $\sigma_p = 0.2$ and peak (mode) given by sigmoid perception curves p_D (for Democrats) and p_R (for Republicans):

$$p_D(b) = 0.7 \tanh\left[\frac{1.00}{0.7}(b + 0.46)\right] - 0.55, \tag{7a}$$

$$p_R(b) = 0.6 \tanh\left[\frac{1.05}{0.6}(b - 0.35)\right] + 0.42. \tag{7b}$$

For details on these beta distributions and the implementation of equations (5b) and (5c), see Appendix B. Reaction parameter values were $d_i = 1.3$, $d_o = 0.8$, $\sigma_r = 0.15$. For finding equilibria, the time constant $\tau = 1$ was used, and populations of both parties were initialized as uniformly distributed on $[-1, 1]$.

For easy comparison with real data, Fig. 1(b) shows a simulation of one hundred times Bail *et al.*'s experimental population: 70 900 Democrats and 54 700 Republicans. In this comparison, we must note that our belief scale is not identical to theirs; ± 1 on our scale are asymptotically extreme, whereas 1 and 7 on Bail *et al.*'s scale are attainable and signify strong agreement on all surveyed issues.

We also replicated the experiment of Bail *et al.* in silico: Starting with a population at equilibrium [shown in Fig. 1(b)], and artificially inducing counterattitudinal Democratic content to Republican experimental subjects (a beta distribution peaked at $p = -0.75$, weighted as if it consisted of 24 percepts on top of a presumed diet of 100 percepts per day) over the course of 30 "days," caused the mean position of those subjects to shift rightwards by a little less than half its natural standard deviation (from 0.30 to 0.42, stdev $\sigma \approx 0.3$). This matches the findings of Bail *et al.*, who found average rightward movement of 0.6 points on a 1-to-7 scale, which represented between 0.11 and 0.59 standard deviations ($p < 0.01$) [38]. Further implementation details can be found in Appendix B.

IV. POSSIBLE EXTENSIONS

In order to keep the number of tuneable parameters and functional forms to a minimum in the absence of much constraining data, we have made many simplifying assumptions.

As such, it is easy to imagine extensions which might increase the realism of this model and which would ideally be implemented when data is available to constrain them.

One simple extension is the addition of more groups/parties, such as independent/unaligned individuals and messages. This would require another perception curve per group, and a three- or more-way fractional content breakdown instead of the single in-group fraction $f(b)$ as our analysis used. Different levels of out-group trust would be represented by different repulsion distances for each type of source.

Additional affiliations beyond party, such as religion, race, regional identity, etc., could also be added to the model. However, the number of identity combinations rises exponentially with each additional affiliation type, and each could affect the perception curve(s)—since affiliations can change the environment individuals are exposed to—and intergroup trust levels, so we submit that this option should be approached with caution.

The repulsion distance d, representing "trust," "credulity," or "benefit of the doubt," need not be binary or universal across the population. Especially in the presence of multiple identity markers, one might let d for each interaction depend on each group identity of the individual and of the messenger to allow for more intricate intergroup prejudices. One could also add noise to d values to model individual variation in level of credulity towards other groups or let d evolve dynamically over time or as a function of political position. However, given the difficulty of measuring intergroup trust levels, we chose to avoid overfitting by using only two d values (in-group and out-group) in our simulations.

One might also reasonably suggest that the political content representing each party could depend on an observer's own party, not just their ideological position b. For instance, a liberal Democrat might perceive different Republican content than a liberal Republican. This would require additional perception curves, rather than re-using the same curve for all observers—each observer/messenger identity pair would require its own curve. This would allow for the possibility that individuals perceive a more extreme version of the other party as they become more extreme themselves, in other words a *negatively* sloped perception curve for out-group content. We note that the large dissonance numbers likely in this case would require a reaction function in which repulsion saturates (i.e., not our cubic assumption) to avoid dwarfing the effect of attraction.

As some other models have attempted [16,19], one might also consider multiple ideological dimensions: Instead of a scalar belief value b, an n-dimensional vector $\mathbf{b}$ would represent an individual's beliefs with respect to each of n issue axes. Percepts would engage with one or more of these issues. Lacking relevant data, we do not put forward assumptions on how reaction dynamics might be coupled; one might assume that dynamics along each axis would be largely independent of one another, since position on one issue rarely affects position on another directly. However, it is possible that the dynamics along multiple axes would be coupled by tribalism; being repelled from a message might drive an individual closer to the opposing camp on more axes than just the one being engaged with, as the individual identifies more strongly with the whole opposing party.

Finally, one might add mechanisms by which the perception curves can change over time. Time dependence could be introduced to investigate hypotheses about the impact of changing or time-correlated media environments, or perception curves might evolve in response to the population state. The latter option would provide a form of indirect coupling between modeled individuals, allowing it to be self-contained over long time scales, but would require conjecture about how environment-creating entities perceive the population state and strategize in response.

V. DISCUSSION

We have put forward a modeling framework for individual political opinion drift which separates perceived content and the reaction of the viewer to that content, in order to separately model perceptual filtering, the shift from attraction to repulsion for dissonant content, and the effect of intergroup bias. We have presented toy models to elucidate each effect on its own in the absence of noise and introduced a graphical analysis technique for qualitative analysis of behavior under general belief-dependent perception curves. With the inclusion of additive noise, analytically determined fixed points widen into stable distributions.

With all these effects included and some simple parameter assumptions, we showed that population distributions matching recent survey data emerge naturally. Furthermore, we were able to simulate the experiment of Bail *et al.* [38] and found similar dynamics under counterattitudinal perturbation.

Further work

A paucity of available data has forced us to make assumptions on functional forms and parameter values. While these are reasonable placeholders, they can be modified or replaced as empirical data become available; it isn't hard to imagine experiments which might elucidate qualitative and quantitative effects of interest. For example, to refine the reaction function, further experiments like that of Bail *et al.* [38] might investigate the impact of political opinions on individuals and how the messenger's apparent identity affects the reception of dissonant ideas.

The lack of reliance on network structure means that data collection can focus on averages and distributions rather than influence-network properties and tie reconstruction. Perceptual diets might be estimated from the top down, by assigning each media outlet or other notable source of political influence an ideology score (as others have done, e.g., Refs. [31,32,35,46]), and surveys or viewership data could determine which content is consumed in what proportion by each part of the ideological spectrum. Alternatively, self-report of political influences and their positions could produce estimates of perceptual diets which also account for interpretation bias—the same content might be interpreted differently by different observers.

Overall, we believe this model shows considerable promise in replicating dynamics and distributions from the real world. We have intentionally chosen a relatively simple structure which is nonetheless able to capture important psychological tendencies for repulsion and tribalism and couple them to a politicized environment, while preserving mathematical

tractability. However, as outlined above, the realism can be significantly increased with extensions which augment the framework.

As data are collected to inform the base model and extensions, the validity and predictive power will grow. In this way, we hope that this framework will offer a lens with which to better understand individual and population-level opinion dynamics and the feedback effects that arise due to the modern reality of personalized political environments. We hope this endeavor leads to a new sort of data-driven political modeling to better understand human behavior, polarization, and strategies for effective political dialogue.

All data and code are available [47].

ACKNOWLEDGMENTS

APPENDIX A: PERCEPTION DISTRIBUTIONS

As we mention in the "adding heterogeneity" section of the main paper, if our model is to have any claim at accurately modeling the political lives of real people, it must allow individuals to consume not just a single, constant percept $p(b)$ but rather a whole distribution of content, $\rho(p; b, \sigma_p)$. In this case, instead of using the single p value to determine an individual's reaction, we calculate their weighted-average reaction by integrating the probability distribution of percepts they might receive multiplied by the reaction those percepts would cause. We note that our cubic reaction function is asymmetric across the repulsion boundary (it is steeper outside the boundary than inside, so repulsion is "stronger" than attraction). Thus, if individuals receive a distribution of percepts centered at their "perception curve" value $p(b)$, a symmetric widening of their experiences has the *asymmetric* effect of shifting the system's fixed points: Since it takes fewer repulsive events than attractive ones to maintain net-zero movement, the new fixed point occurs when the center of the perceptual distribution is still in the trust region. In other words, the repulsion boundary is effectively narrowed with regard to the peak percept value $p(b)$.

In Appendix A we will focus on opinion drift caused by perceptions from a single group, neglecting reaction noise (i.e. $\tau \, db/dt = (1 - b^2)v$, a simplification of equation Eq. (5), but dynamics in a full model are a simple linear combination of two such sources and a reaction-noise term.

The precise effects of perceptual variety depend on the shape of the perceptual distribution and the choice of reaction function. We now consider three options in order of increasing realism.

1. Gaussian distributed percepts

First we'll consider Gaussian-distributed percepts centered on the "perception curve" value $p(b)$, and the cubic reaction
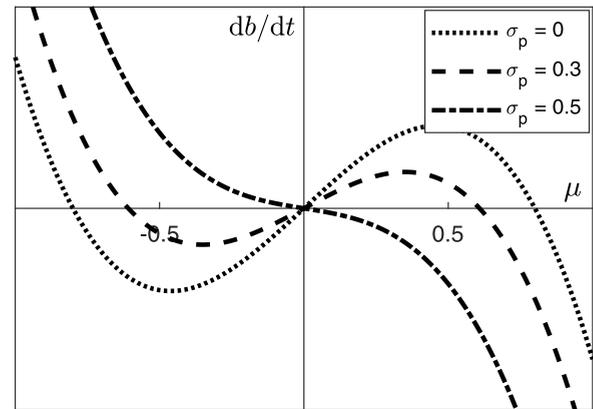

FIG. 7. Effect of perception distribution width on reaction. Net change in belief db/dt versus expected value of dissonance μ for varying levels of perception distribution width, from Eq. (A1) with $d = 0.8$. The critical standard deviation for $d = 0.8$ is $\sigma_c = 0.8/\sqrt{3} \approx 0.46$.

function from Eq. (1) of the main text, $(p - b)[1 - (p - b)^2/d^2]$. These choices are convenient in that the integral for average belief change is analytically tractable. For clarity, we change variables to "average dissonance" $\mu = p(b) - b$, and let x be the dummy variable of integration for possible dissonance. If we allow percepts outside of $[-1, 1]$ in this way, the integral is quite clean:

$$\tau \frac{db}{dt} = (1 - b^2) \int_{-\infty}^{+\infty} \underbrace{x \left(1 - \frac{x^2}{d^2}\right)}_{R(x;d)} \underbrace{\left[\frac{1}{\sqrt{2\pi}\,\sigma_p} e^{-\frac{(x-\mu)^2}{2\sigma_p^2}}\right]}_{\rho(x;b,\sigma_p)} dx$$

$$= (1 - b^2)\,\mu \left[\frac{(d^2 - 3\sigma_p^2)}{d^2} - \frac{\mu^2}{d^2}\right]. \tag{A1}$$

This is a cubic in μ with zeros representing the three fixed points: at $\mu = 0$ and at $\mu = \pm\sqrt{d^2 - 3\sigma_p^2}$. So as the parameter σ_p grows, we see a pitchfork bifurcation as the nonorigin zeros disappear: For $\sigma_p > \sigma_c = d/\sqrt{3}$, the bracketed term in Eq. (A1) is always negative. This means the net movement of the individual is *away* from the average percept they see. See Fig. 7.

These distribution widths are not unrealistically large; as seen in Fig. 7, for a repulsion distance of 0.8 the standard deviation needs only be 0.46 for the overall effect of a content distribution to be repulsive (i.e., causing movement away from that distribution's mean). Thus, especially for out-group content with a naturally narrower repulsion distance, viewing a wider distribution of that content can actually cause repulsion, since the extreme percepts will repel the viewer more than the moderate percepts will attract them.

To visualize the effects of normally distributed perceptual distributions $\rho(p; b, \sigma_p)$ replacing deterministic percepts $p(b)$, we can examine density plots for the net movement for all combinations of b and p (repulsion distance $d = 0.8$): see Fig. 8. This is the space that our graphical analysis technique utilizes: If we establish a perception curve $p(b)$, the values of this map that the curve crosses are the realized average movement for each part of the population.
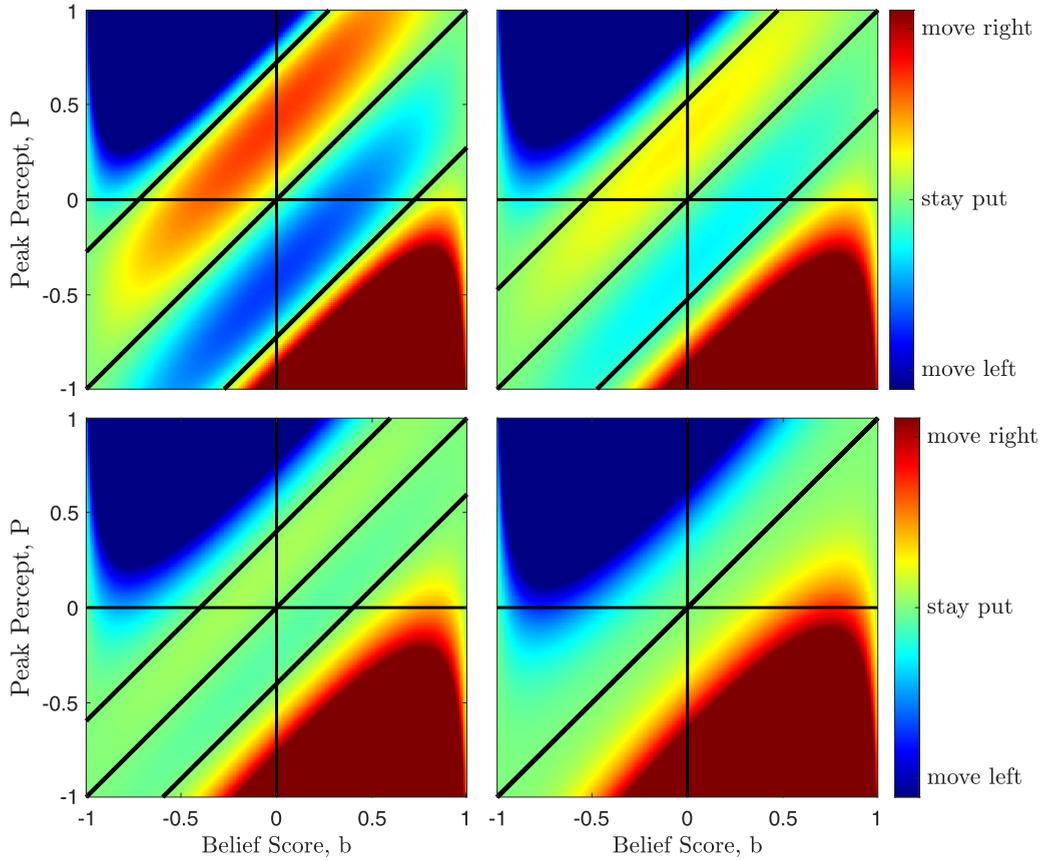
FIG. 8. Reaction map for normally distributed diets. Net movement caused by normally distributed perceptual diets with peak $P = p(b)$, for individuals at belief score b, and repulsion distance $d = 0.8$. Results shown for $\sigma_p = 0.2$ (top left), 0.3 (top right), 0.4 (bottom left), and 0.5 (bottom right).

2. Truncated Gaussian Percepts

In deriving Eq. (A1) we approximated by integrating over the entire real line for dissonance when it should be constrained to the range allowed by percepts in $[-1, 1]$—that is, from dissonance $x = -1 - b$ to $1 - b$. That makes the result somewhat more complicated (note: lacking symmetry around b, we don't utilize the μ substitution, and x represents percept value instead of dissonance):

$$
\tau \frac{db}{dt} = (1 - b^2) \int_{-1}^{1} \left[\frac{A}{\sqrt{2\pi}\,\sigma_p} e^{-\frac{(x-P)^2}{2\sigma_p^2}} \right]
$$

$$
\times \left\{ (x - b) \left[1 - \frac{(x-b)^2}{d^2} \right] \right\} dx
$$

$$
= A \frac{(1 - b^2)\sigma_p}{d^2 \sqrt{2\pi}} \left\{ \left[1 - d^2 + 2\sigma_p^2 + P^2 - 3Pb + 2b^2 \right] \right.
$$

$$
\times \left[e^{\frac{-(-1+P)^2}{2\sigma_p^2}} - e^{\frac{-(1+P)^2}{2\sigma_p^2}} \right]
$$

$$
\left. + [P - 3b] \left[e^{\frac{-(-1+P)^2}{2\sigma_p^2}} + e^{\frac{-(1+P)^2}{2\sigma_p^2}} \right] \right.
$$

$$
+ \frac{(1 - b^2)(P - b)}{2d^2} \left\{ d^2 - 3\sigma_p^2 - [P - b]^2 \right\}
$$

$$
\times \left[\operatorname{erf}\left(\frac{1 + P}{\sqrt{2}\sigma_p} \right) + \operatorname{erf}\left(\frac{-1 + P}{\sqrt{2}\sigma_p} \right) \right], \tag{A2}
$$

using shorthand $P = p(b)$ for compactness. We also note that σ_p in this case is the standard deviation of the full Gaussian, not the truncated one. A is a normalization factor depending on b and σ_p needed to make the truncated Gaussian integrate to 1:

$$
A = \frac{1}{\frac{1}{\sqrt{2\pi}\sigma_p} \int_{-1}^{1} e^{\frac{-(x-P)^2}{2\sigma_p^2}} \, dx} = \frac{2}{\left[\operatorname{erf}\left(\frac{1+P}{\sqrt{2}\sigma_p} \right) - \operatorname{erf}\left(\frac{-1+P}{\sqrt{2}\sigma_p} \right) \right]}.
$$

Figure 9 shows the reaction map for these truncated-normal diets, computed analytically at each b and P combination. Very small values (near zero mean movement) are colored black, showing areas of relative indifference where reaction noise will dominate.

3. Beta Distributed Percepts

For our simulations, we bounded perceptual diets in a more natural way, by utilizing beta distributions stretched to fit $[-1, 1]$. These distributions approach zero at the boundaries
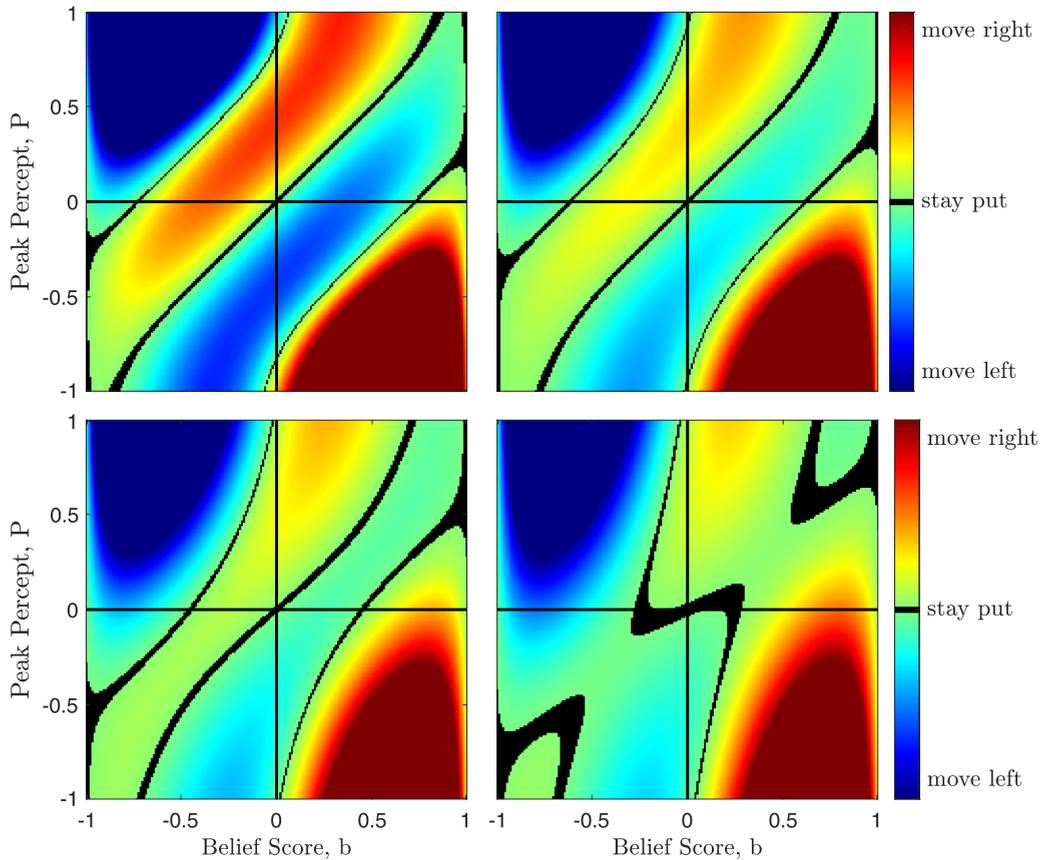
FIG. 9. Reaction map for truncated-normal diets. Net movement caused by truncated-normal perceptual diets with peak P (vertical axis), for individuals at belief score b (horizontal axis), with repulsion distance $d = 0.8$. Results shown for $\sigma_p = 0.2$ (top left), 0.3 (top right), 0.4 (bottom left), and 0.5 (bottom right).

of our domain, fitting our asymptotic-extremes interpretation of this axis. The beta distribution with our endpoints has the equation

$$\text{Beta}_{[-1,1]}(x; \alpha, \beta) = 4 \frac{(1+x)^{\alpha-1}(1-x)^{\beta-1}}{2^{\alpha+\beta}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}, \tag{A3}$$

where α and β are parameters of the distribution and Γ is the gamma function.

We can construct a distribution to have any desired mode (peak) $P = p(b)$ and standard deviation σ_p by solving the implicit equations

$$\text{mode} = P = \frac{\alpha - \beta}{\alpha + \beta - 2} \tag{A4}$$

$$\text{variance} = \sigma_p^2 = \frac{4\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{A5}$$

for $\alpha, \beta > 1$ in terms of P and σ_p. Examples are shown in Fig. 10. Unfortunately, when using these beta distributions, the weighting integrals with our cubic reaction function aren't possible to evaluate in closed form. However, we may numerically compute these integrals for a finite grid of P and b values at any chosen standard deviation to visualize the reaction space. In Fig. 11, we can see the repulsion boundaries bending and bifurcating as σ_p increases.

Computing a reaction map like in Figs. 8, 9, or 11 allows us to use our graphical analysis technique with *any* perception curve, to get a sense of average population drift for the whole political spectrum. For our "realistic" simulation shown in Fig. 1 of the main text, we used beta-distributed perceptual diets. To compute in-group and out-group drift values v_{in} and v_{out} in a computationally feasible way, we discretized the b
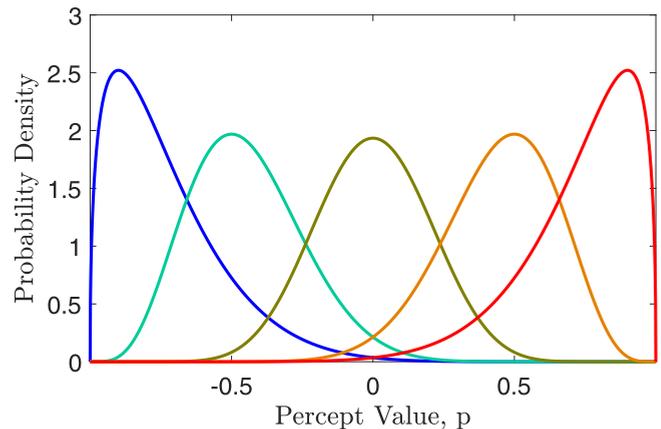

FIG. 10. Beta-distributed diets. Examples of beta distributions with peaks at $P = -0.9$ (blue), -0.5 (teal), 0 (green), 0.5 (orange), and 0.9 (red). All have the same standard deviation, $\sigma_p = 0.2$.
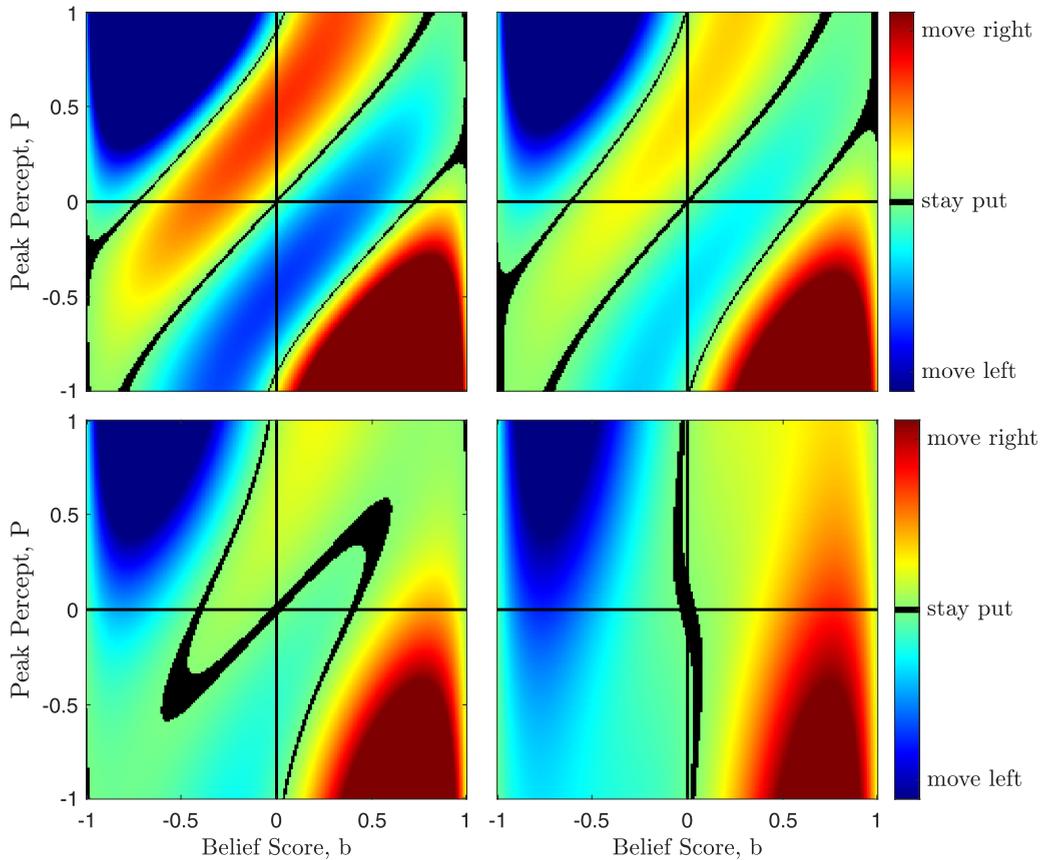
FIG. 11. Reaction map for beta-distributed diets. Average movement caused by beta-distributed perceptual diets with peak P, for individuals at belief score b, and repulsion distance $d = 0.8$. Results shown for $\sigma_p = 0.2$ (top left), 0.3 (top right), 0.4 (bottom left), and 0.5 (bottom right).

and P domains to the nearest hundredth and computed the integrals v_{in} and v_{out} at each possible combination—as was done for Figs. 9 and 11, which show v_{out} for different σ_p values. Then in iteration, we used nearest-neighbor interpolation with this map rather than computing each individual's weighting integral at each time step.

APPENDIX B: BAIL *et al.* IN SILICO DETAILS

For simulation of Bail's experiment [38], the population was initialized at its equilibrium, but in addition to v_{in} and v_{out} there was a third influence v_{bot} based on an out-group distribution peaked at value $P = -0.75$ shown to Republicans and $P = 0.3$ shown to Democrats, to roughly match the other

party's equilibrium distribution. This extra out-group effect was weighted as if it consisted of 24 additional percepts on top of a 100-percept daily diet, i.e., with weight $f_{\text{bot}} = 24/124$. So our full SDE becomes

$$\tau \, db = (1 - b^2)\{[(1 - f_{\text{bot}})(f v_{\text{in}} + (1 - f) v_{\text{out}})$$
$$+ f_{\text{bot}} v_{\text{bot}}] dt + \sigma_r dW\}. \tag{B1}$$

Under this assumption, the time constant $\tau = 30$ caused movement in agreement with Bail *et al.* [38]: slight leftward movement of Democrat mean from -0.51 to -0.53 (about 6% of its natural standard deviation), but significant rightward movement of the Republican mean from 0.30 to 0.42 (about 40% of its natural standard deviation).

[1] R. A. Holley and T. M. Liggett, Ergodic theorems for weakly interacting infinite systems and the voter model, Ann. Probab. **3**, 643 (1975).

[2] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, Mixing beliefs among interacting agents, Adv. Complex Syst. **3**, 87 (2000).

[3] R. Hegselmann and U. Krause, Opinion dynamics and bounded confidence models, analysis, and simulation, J. Artif. Soc. Soc. Simul. **5**, 1 (2002).

[4] V. Sood and S. Redner, Voter Model on Heterogeneous Graphs, Phys. Rev. Lett. **94**, 178701 (2005).

[5] C. Castellano, M. A. Muñoz, and R. Pastor-Satorras, Nonlinear q-voter model, Phys. Rev. E **80**, 041129 (2009).

[6] R. Hegselmann and U. Krause, Opinion dynamics driven by various ways of averaging, Comput. Econ. **25**, 381 (2005).

[7] S. Galam and F. Jacobs, The role of inflexible minorities in the breaking of democratic opinion dynamics, Physica A **381**, 366 (2007).

[8] A. C. R. Martins, Continuous opinions and discrete actions in opinion dynamics problems, Int. J. Mod. Phys. C **19**, 617 (2008).

[9] G. Verma, A. Swami, and K. Chan, The impact of competing zealots on opinion dynamics, Physica A **395**, 310 (2014).

[10] D. Knopoff, On a mathematical theory of complex systems on networks with application to opinion formation, Math. Models Methods Appl. Sci. **24**, 405 (2014).

[11] Z. Liu, J. Ma, Y. Zeng, L. Yang, Q. Huang, and H. Wu, On the control of opinion dynamics in social networks, Physica A **409**, 183 (2014).

[12] S. Wang, L. Rong, and J. Wu, Bistability and multistability in opinion dynamics models, Appl. Math. Comput. **289**, 388 (2016).

[13] M. Del Vicario, A. Scala, G. Caldarelli, H Stanley, and W. Quattrociocchi, Modeling confirmation bias and polarization, Sci. Rep. **7**, 40391 (2016).

[14] J. P. Pinasco, V. Semeshenko, and P. Balenzuela, Modeling opinion dynamics: Theoretical analysis and continuous approximation, Chaos, Solitons Fractals **98**, 210 (2017).

[15] M. H. Afrasiabi, R. Guérin, and S. S. Venkatesh, Opinion formation in Ising networks, Online Social Networks and Media **5**, 1 (2018).

[16] M. Ye, M. H. Trinh, Y.-H. Lim, B. D. O. Anderson, and H.-S. Ahn, Continuous-time opinion dynamics on multiple interdependent topics, Automatica **115**, 108884 (2020).

[17] T. F. A. Alves, G. A. Alves, F. W. S. Lima, and A. Macedo-Filho, Phase diagram of a continuous opinion dynamics on Barabasi-Albert networks, J. Stat. Mech.: Theory Exp. (2020) 033203.

[18] A. Kowalska-Styczeń and K. Malarz, Noise induced unanimity and disorder in opinion formation, PLOS One **15**, e0235313 (2020).

[19] M. V. Maciel and A. C. R. Martins, Ideologically motivated biases in a multiple issues opinion model, Physica A **553**, 124293 (2020).

[20] S. Galam, Rational group decision making: A random field Ising model at t = 0, Physica A **238**, 66 (1997).

[21] S. Galam, Contrarian deterministic effects on opinion dynamics: "the hung elections scenario," Physica A **333**, 453 (2004).

[22] K. Sznajd-Weron and J. Sznajd, Opinion evolution in closed community, Int. J. Mod. Phys. C **11**, 1157 (2000).

[23] K. Sznajd-Weron, Sznajd model and its applications, Acta Phys. Pol. B **36**, 2537 (2005).

[24] G. Toscani, Kinetic models of opinion formation, Commun. Math. Sci. **4**, 481 (2006).

[25] M. L. Bertotti and M. Delitala, On a discrete generalized kinetic approach for modelling persuader's influence in opinion formation processes, Math. Comput. Model. **48**, 1107 (2008).

[26] B. Düring, P. Markowich, J.-F. Pietschmann, and M.-T. Wolfram, Boltzmann and Fokker–Planck equations modelling opinion formation in the presence of strong leaders, Proc. R. Soc. A **465**, 3687 (2009).

[27] B. Düring and M.-T. Wolfram, Opinion dynamics: Inhomogeneous Boltzmann-type equations modelling opinion leadership and political segregation, Proc. R. Soc. A **471**, 20150345 (2015).

[28] S. Galam, Y. Gefen, and Y. Shapir, Sociophysics: A new approach of sociological collective behaviour. i. mean-behaviour description of a strike, J. Math. Sociol. **9**, 1 (1982).

[29] S. Galam, Majority rule, hierarchical structures, and democratic totalitarianism: A statistical approach, J. Math. Psychol. **30**, 426 (1986).

[30] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. **81**, 591 (2009).

[31] T. Groseclose and J. Milyo, A measure of media bias, Q. J. Econ. **120**, 1191 (2005).

[32] R. M. Entman, Framing bias: Media in the distribution of power, J. Commun. **57**, 163 (2007).

[33] M. D. Slater, Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity, Commun. Theory **17**, 281 (2007).

[34] S. DellaVigna and E. Kaplan, The Fox News effect: Media bias and voting, Q. J. Econ. **122**, 1187 (2007).

[35] G. J. Martin and A. Yurukoglu, Bias in cable news: Persuasion and polarization, Am. Econ. Rev. **107**, 2565 (2017).

[36] D. Acemoglu and A. Ozdaglar, Opinion dynamics and learning in social networks, Dynamic Games and Applications **1**, 3 (2011).

[37] K. Fan and W. Pedrycz, Evolution of public opinions in closed societies influenced by broadcast media, Physica A **472**, 53 (2017).

[38] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. B. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, Exposure to opposing views on social media can increase political polarization, Proc. Natl. Acad. Sci. **115**, 9216 (2018).

[39] K. T. Poole and H. Rosenthal, The polarization of American politics, J. Politics **46**, 1061 (1984).

[40] Pew Research Center, The partisan divide on political values grows even wider, Tech. Rep. (2017), https://www.pewresearch.org/politics/2017/10/05/the-partisan-divide-on-political-values-grows-even-wider/.

[41] C. Hare and K. T. Poole, The polarization of contemporary American politics, Polity **46**, 411 (2014).

[42] Pew Research Center, Partisanship and political animosity in 2016, Tech. Rep. (2016), https://www.pewresearch.org/politics/2016/06/22/partisanship-and-political-animosity-in-2016/.

[43] Note that this is not necessarily inconsistent with a network approach; we believe a dynamical interaction network dependent on affiliation and political position could lead to a similar model.

[44] H. Tajfel and J. C. Turner, An integrative theory of intergroup conflict, in *Organizational Identity: A Reader*, edited by M. J. Hatch and M. Schultz (Oxford University Press, Oxford, 1979), Chap. 4, pp. 56–65.

[45] In cases like this with multiple attracting "camps" without significant overlap, the long-term populations of each camp may depend on initial conditions.

[46] V. Otero, Media bias chart: Version 5.0, Accessed at https://www.adfontesmedia.com/ (2020).

[47] D. Sabin-Miller and D. M. Abrams, Data and Code for "When Pull Turns to Shove: A Continuous-Time Model for Opinion Dynamics", https://doi.org/10.21985/n2-4msy-vj08 (2020).