When Pull Turns To Shove: A Mathematical Model For Opinion Dynamics

David Sabin-Miller
Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL, USA

Daniel M. Abrams
Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL, USA
Northwestern Institute for Complex Systems, Northwestern University, Evanston, IL, USA and
Department of Physics and Astronomy, Northwestern University, Evanston, IL, 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 differential equations 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. 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.

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 [2–15]. Other researchers have advanced “sociophysics” approaches such as Ising [16, 17], Sznajd [18, 19], and generalized-kinetic models [20–23] which apply techniques from physics to analyze analogous social systems [24]. Complementary to these modeling approaches, theoretical and empirical work from economics and social science has examined the political bias of media entities [25, 26] and their influence on a population [27–31]. All these approaches contribute valuable insight towards 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 worldview.
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.

**Political Spectrum**

Like many prior approaches (e.g. [2][8][10][13][25][28][29][32][34]), we consider a single, finite ideology axis. Extensions to this are possible, as discussed in section S4 of the Supplementary Information (SI), but the unidimensional approximation is supported by empirical results showing that the liberal-conservative dimension captures the great majority of modern legislative behavior [35]. Figure 1(a) shows a one-dimensional projection of political ideology for the U.S. population based on one study [32], though the precise methods of projecting the political landscape onto one axis differ between sources, other recent reports like that of Pew Research [34] 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.

**Opinion Change**

Classic “bounded-confidence” models (e.g., [2][8]), 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 [36] 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., [10][13]), we supplement local-attraction behavior with distal repulsion: individuals who are exposed to ideas which 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. [32], it was found that exposure to 24 tweets per day from a prominent member 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).

**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:

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

shown visually in Fig. 2.

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 con-
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 $\tau$ and current belief score $b$. For this image a repulsion distance of $d = 0.8$ was chosen.

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. [1–7, 9, 10, 13, 15, 17–19, 24]). 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.

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.

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

Toy Models

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 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:

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

This ordinary differential equation (ODE) has fixed points at $b = C$, $b = C \pm d$, and $b = \pm 1$ (due to the imposed domain bounds). The fixed point at $b = C$ 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 don’t include such party-switching dynamics in this initial model.
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" [37]. 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 \( d_{o} \leq d_{i} \). 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 \( \frac{db}{dt} \) is then a simple weighted average of the flow functions in Eq. (2) due to each source:

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

where in general \( f = f(b) \), \( R_i = R(p - b; d_i) \), and \( R_o = R(p - b; d_o) \).

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 section S4 of the SI).

![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.](image)

Adding Personalized Perceptions

Instead of constant political messaging across the whole belief spectrum, we now consider a "perception curve" \( p(b) \) linking individuals’ perceived world to their current beliefs. This reflects the differing "slice" of the world that individuals see as a result of their differing environments and personal biases.

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 worldview.

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 average 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 re-
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.

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.

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 section S5 of the SI).

**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 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 $-1^2$.

If we wish to add variability to the percept instead of the reaction, nonlinear effects become more

---

2 In cases like this with multiple attracting “camps” without significant overlap, the long-term populations of each camp may depend on initial conditions.
important, since $p$ (now properly a probability distribution $\rho(p)$) must be fed through reaction function $R(p - b)$ before its effects are determined. This means 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 $\rho(p)$. Due to the asymmetry of the reaction function across the repulsion boundary—repulsion as modeled above is stronger than attraction—a symmetric distribution of percepts centered at the boundary will have a net repulsive effect. Thus, broadening a perceptual diet effectively narrows the trust region, and there is a critical noise threshold at which we observe noise-induced pitchfork bifurcations of our fixed points—this is discussed in greater detail in section S1 the SI.

RESULTS

When we put together all the modeled effects described above, we find robust realistic distributions at equilibrium (see Fig. 1(b)). We used simple sigmoid perception curves (shown in the inset to Fig. 1(b), details in section S3 of the SI) for the peaks of perceptual diets (standard deviation $\sigma_p = 0.2$), and noisy reactions (standard deviation $\sigma_r = 0.15$), along with simple linear fractional content ratios

\begin{align}
  f_d &= 0.7 + 0.2b \\
  f_r &= 0.7 - 0.2b 
\end{align}

(5) (6)

to emulate a “media bubble” effect. This model’s steady state shows good agreement with real-world distributions of U.S. political ideology from Bail et al. [32] and Pew Research [34]—see Fig. 1(b). 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 can also replicate the experiment of Bail et al. in silico: starting with a population at equilibrium (shown in Fig. 1(b)), and artificially inducing counter-attitudinal Democratic content to Republican experimental subjects (a 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,” causes 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 ($p < 0.01$) [32]. Implementation details can be found in section S3 of the SI.

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. [32] and found similar outcomes.

Our approach allows for modeling of important psychological effects such as self-serving bias (perception curves are increasing functions of $b$) and intergroup bias (repulsion distance depends on source) without requiring a network. This frees future data-gathering efforts from the often challenging task of network tie construction, and allows for easy simulation of very large populations.

In order to approach this topic, we have made considerable simplifications, and it is easy to imagine extensions which might increase the realism of this model (as we discuss in section S4 of the SI). We have intentionally chosen a relatively simple structure which is nonetheless able to capture psychological tendencies for repulsion and tribalism, and couple them to a politicized environment, while preserving mathematical tractability. 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 (see section S5 of the SI). 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.

**ACKNOWLEDGMENTS**

The authors would like to acknowledge NSF support through the Graduate Research Fellowship Program.

DSM proposed the initial model and performed numerical simulations. DMA contributed to modifications of the model. Both authors participated in writing of the manuscript.

The authors have no competing interests.

All data and code is available upon request.

* davidslabmiller@u.northwestern.edu

[1] Richard A. Holley and Thomas M. Liggett, “Ergodic theorems for weakly interacting infinite systems and the voter model,” *The Annals of Probability* **3**, 643–663 (1975).

[2] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch, “Mixing beliefs among interacting agents,” *Advances in Complex Systems* **3**, 87–98 (2000).

[3] Rainer Hegselmann and Ulrich Krause, “Opinion dynamics and bounded confidence models, analysis, and simulation,” *Journal of Artificial Societies and Social Simulation* **5** (2002).

[4] V. Sood and S. Redner, “Voter model on heterogeneous graphs,” *Phys. Rev. Lett.* **94**, 178701 (2005).

[5] Claudio Castellano, Miguel A. Muñoz, and Romualdo Pastor-Satorras, “Nonlinear q-voter model,” *Physical Review E* **80**, 041129 (2009).

[6] Rainer Hegselmann and Ulrich Krause, “Opinion dynamics driven by various ways of averaging,” *Computational Economics* **25**, 381–405 (2005).

[7] Serge Galam and Frans Jacobs, “The role of inflexible minorities in the breaking of democratic opinion dynamics,” *Physica A: Statistical Mechanics and its Applications* **381**, 366–376 (2007).

[8] André CR Martins, “Continuous opinions and discrete actions in opinion dynamics problems,” *International Journal of Modern Physics C* **19**, 617–624 (2008).

[9] Gunjan Verma, Ananthram Swami, and Kevin Chan, “The impact of competing zealots on opinion dynamics,” *Physica A: Statistical Mechanics and its Applications* **395**, 310–331 (2014).

[10] D. Knopoff, “On a mathematical theory of complex systems on networks with application to opinion formation,” *Mathematical Models and Methods in Applied Sciences* **24**, 405–426 (2014).

[11] Zhihong Liu, Jianfeng Ma, Yong Zeng, Li Yang, Qiping Huang, and Hongliang Wu, “On the control of opinion dynamics in social networks,” *Physica A: Statistical Mechanics and its Applications* **409**, 183–198 (2014).

[12] Shaoli Wang, Libin Rong, and Jianhong Wu, “Bistability and multistability in opinion dynamics models,” *Applied Mathematics and Computation* **289**, 388–395 (2016).

[13] Michela Del Vicario, Antonio Scala, Guido Caldarelli, H Stanley, and Walter Quattrociocchi, “Modeling confirmation bias and polarization,” *Scientific Reports* **7**, 40391 (2016).

[14] Juan Pablo Pinasco, Viktoryia Semeshenko, and Pablo Balenzuela, “Modeling opinion dynamics: Theoretical analysis and continuous approximation,” *Chaos, Solitons & Fractals* **98**, 210–215 (2017).

[15] Mohammad Hadi Afrasiabi, Roch Guérin, and SS Venkatesh, “Opinion formation in Ising networks,” *Online Social Networks and Media* **5**, 1–22 (2018).

[16] Serge Galam, “Rational group decision making: A random field Ising model at $t=0$,” *Physica A: Statistical Mechanics and its Applications* **238**, 66–80 (1997).

[17] Serge Galam, “Contrarian deterministic effects on opinion dynamics: “the hung elections scenario”,” *Physica A: Statistical Mechanics and its Applications* **333**, 453–460 (2004).

[18] Katarzyna Sznajd-Weron and Jozef Sznajd, “Opinion evolution in closed community,” *International Journal of Modern Physics C* **11**, 1157–1165 (2000).

[19] Katarzyna Sznajd-Weron, “Sznajd model and its applications,” *Acta Physica Polonica B* **36**, 2537 (2005).

[20] Giuseppe Toscani et al., “Kinetic models of opinion formation,” *Communications in mathematical sciences* **4**, 481–496 (2006).

[21] Maria Letizia Bertotti and Marcello Delitala, “On a discrete generalized kinetic approach for modelling persuader’s influence in opinion formation processes,” *Mathematical and computer modelling* **48**, 1107–1121 (2008).

[22] Bertram Düring, Peter Markowich, Jan-Frederik Pietschmann, and Marie-Therese Wolfram, “Boltzmann and fokker–planck equations modelling opinion formation in the presence of strong leaders,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **465**, 3687–3708 (2009).

[23] Bertram Düring and Marie-Therese Wolfram, “Opinion dynamics: inhomogeneous boltzmann-type equations modelling opinion leadership and political segregation,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **471**, 20150345 (2015).

[24] Claudio Castellano, Santo Fortunato, and Vittorio Loreto, “Statistical physics of social dynamics,” *Reviews of Modern Physics* **81**, 591–646 (2009).

[25] Tim Groseclose and Jeffrey Milyo, “A measure of media bias,” *The Quarterly Journal of Economics* **120**, 1191–1237 (2005).
[26] Robert M Entman, “Framing bias: Media in the distribution of power,” Journal of Communication 57, 163–173 (2007).
[27] Michael D. Slater, “Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity,” Communication Theory 17, 281–303 (2007).
[28] Stefano DellaVigna and Ethan Kaplan, “The Fox News effect: Media bias and voting,” The Quarterly Journal of Economics 122, 1187–1234 (2007).
[29] Gregory J. Martin and Ali Yurukoglu, “Bias in cable news: Persuasion and polarization,” American Economic Review 107, 2565–99 (2017).
[30] Daron Acemoglu and Asuman Ozdaglar, “Opinion dynamics and learning in social networks,” Dynamic Games and Applications 1, 3–49 (2011).
[31] Kangqi Fan and Witold Pedrycz, “Evolution of public opinions in closed societies influenced by broadcast media,” Physica A: Statistical Mechanics and its Applications 472, 53–66 (2017).
[32] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haochan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky, “Exposure to opposing views on social media can increase political polarization,” Proceedings of the National Academy of Sciences 115, 9216–9221 (2018).
[33] Keith T Poole and Howard Rosenthal, “The polarization of American politics,” The Journal of Politics 46, 1061–1079 (1984).
[34] Pew Research Center, The partisan divide on political values grows even wider, Tech. Rep. (2017).
[35] Christopher Hare and Keith T Poole, “The polarization of contemporary American politics,” Polity 46, 411–429 (2014).
[36] Pew Research Center, Partisanship and Political Animosity in 2016, Tech. Rep. (2016).
[37] Henri Tajfel and John C Turner, “An integrative theory of intergroup conflict,” in Organizational Identity: A Reader (Oxford University Press Oxford, United Kingdom, 1979) Chap. 4, pp. 56–65.
[38] Vanessa Otero, “Media bias chart: Version 5.0,” Accessed at https://www.adfontesmedia.com/ (2020).
Supplementary Information for “When Pull Turns To Shove: A Mathematical Model For Opinion Dynamics”

David Sabin-Miller*
Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL, USA

Daniel M. Abrams
Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL, USA
Northwestern Institute for Complex Systems, Northwestern University, Evanston, IL, USA and Department of Physics and Astronomy, Northwestern University, Evanston, IL, USA

S1. 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’s steeper outside the boundary than inside, so repulsion is “stronger” than attraction). Thus, if individuals receive a Gaussian distribution of percepts centered at their “perception curve” value \( p(b) \), this 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 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 \( p(b) \).

The precise effects of this perceptual variety depend on the shape of the perceptual distribution and the choice of reaction function. First we consider Gaussian-distributed percepts centered on the “perception curve” value \( p(b) \), and the cubic reaction 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. An approximate version of this integral, with infinite limits of integration (so that many terms drop out), is:

\[
\tau \left\langle \frac{db}{dt} \right\rangle = (1 − b^2) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_p} e^{-\frac{(x−\mu)^2}{2\sigma_p^2}} \left[ x \left( 1 − \frac{x^2}{d^2} \right) \right] dx
\]

\[
= (1 − b^2) \mu \left[ \frac{(d^2 − 3\sigma_p^2)}{d^2} − \frac{\mu^2}{d^2} \right].
\]  

(S7)

We see that \( d^2 − 3\sigma_p^2 \) plays the role that \( d^2 \) played before, setting the non-origin zeros of the cubic; when \( \mu^2 \) is greater than this value, the average movement of the individual is away from the peak percept. There is also a critical variance \( \sigma_c^2 = d^2/3 \) at which the system undergoes a pitchfork bifurcation. For \( \sigma_p > \sigma_c \) the bracketed term in Eq. (S7) is always negative, i.e., the net change in belief is always contrary to the average media perceived. See Figure S7.

These distribution widths are not unrealistically large; as seen in Fig. S7, 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...
FIG. S7: **Effect of perception distribution width on reaction.** Net change in belief $\frac{db}{dt}$ versus expected value of dissonance $\mu$ for varying levels of perception distribution width, from Eq. (S7) with $d = 0.8$. The critical standard deviation for $d = 0.8$ is $\sigma_c = 0.8/\sqrt{3} \approx 0.46$.

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. S8. 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. S8: **Reaction map for normally distributed diets.** Average movement caused by normally 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).
A. Bounded Percepts: Truncated Gaussian

In deriving Eq. [S7] 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 \(\mu\) substitution, and \(x\) represents percept value instead of dissonance):

\[
\tau \langle \frac{db}{dt} \rangle = (1 - b^2) \int_{-1}^{1} \left[ \frac{1}{\sqrt{2\pi\sigma_p}} e^{-\frac{(x-p)^2}{2\sigma_p^2}} \right] \left\{ (x - b) \left[ 1 - \frac{(x - b)^2}{d^2} \right] \right\} dx
\]

\[
= A \frac{(1 - b^2)}{d^2\sqrt{2\pi}} \left\{ [1 - d^2 + 2\sigma_p^2 + p^2 - 3pb + 2b^2] \left[ e^{-\frac{-(1+p)^2}{2\sigma_p^2}} - e^{-\frac{-d^2+2b^2}{2\sigma_p^2}} \right] + [p - 3b] \left[ e^{-\frac{-(1+p)^2}{2\sigma_p^2}} + e^{-\frac{d^2+2b^2}{2\sigma_p^2}} \right] \right\}
\]

\[
+ \frac{(1 - b^2)}{2d^2} \left\{ d^2 - 3\sigma_p^2 - [p - b]^2 \right\} \left[ \text{erf} \left( \frac{1 + p}{\sqrt{2\sigma_p}} \right) + \text{erf} \left( \frac{-1 + p}{\sqrt{2\sigma_p}} \right) \right].
\]

(S8)

where \(A\) is a normalization factor depending on \(b\) and \(\sigma_p\) needed to make the truncated Gaussian integrate to 1:

\[
A = \frac{1}{\sqrt{2\pi\sigma_p}} \int_{-1}^{1} e^{-\frac{(x-p)^2}{2\sigma_p^2}} dx = \frac{2}{\text{erf} \left( \frac{1 + p}{\sqrt{2\sigma_p}} \right) - \text{erf} \left( \frac{-1 + p}{\sqrt{2\sigma_p}} \right)}
\]

Fig. [S9] shows the reaction map for these truncated-normal diets, computed analytically at each \(b\) and \(p\) combination.

**FIG. S9:** Reaction map for truncated-normal diets. Expected 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).
B. Bounded Percepts: Beta Distributions

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

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

(S9)

where $\alpha$ and $\beta$ are parameters of the distribution and $\Gamma$ is the gamma function.

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

$$\text{mode} = p = \frac{\alpha - \beta}{\alpha + \beta - 2} \quad \text{(S10)}$$

$$\text{variance} = \sigma_p^2 = \frac{4\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad \text{(S11)}$$

for $\alpha, \beta > 1$ in terms of $p$ and $\sigma$. Examples are shown in Fig. S10.

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. S11 we can see the repulsion boundaries bending and bifurcating as $\sigma_p$ increases.

Computing a reaction map like S8, S9 or S11 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.

S2. STOCHASTIC DIFFERENTIAL EQUATION DETAILS

For the “realistic” simulation shown in Fig. 1 of the main text, we used beta-distributed perceptual diets with constant standard deviation $\sigma_p$. Party perception curves $p_D(b)$ and $p_R(b)$ determine the peaks of these distributions, so that individuals see in-group and out-group content distributions $\rho_{in}(p; b, \sigma_p)$ and $\rho_{out}(p; b, \sigma_p)$. The average effect of each distribution is the integral against the reaction such content would cause (i.e., Eqs. S13 and S14). Each group’s effect on the observer is weighted by its content fraction, then the stochastic reaction-noise term is added before all movement is edge-damped, leaving us with Eq. S15.
FIG. S11: 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).

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

\[ v_{in} = \int_{-1}^{1} R(p - b; d_{in}) \rho_{in}(p; b, \sigma_p) dp \]

\[ v_{out} = \int_{-1}^{1} R(p - b; d_{out}) \rho_{out}(p; b, \sigma_p) dp \]

\[ \tau \ dB = (1 - b^2) \left\{ [f v_{in} + (1 - f) v_{out}] \ dt + \sigma_r dW \right\} \]

This can be made computationally feasible by discretizing the $b$ and $p$ domains (e.g., to the nearest hundredth) and computing the integrals $v_{in}$ and $v_{out}$ at each possible combination—as was done for Figs. S9 and S11 which show $v_{out}$ for different $\sigma_p$ values. Then in iteration, we simply reference the nearest-case pre-computed value (nearest-neighbor interpolation) rather than computing each individual’s weighting integral at each time-step.

**S3. IMPLEMENTATION DETAILS**

Our “realistic” simulation shown in Fig. 1b of the main text initialized the population’s starting beliefs to uniform random values in $[-1,1]$ for both parties. The sigmoid perception curves (shown in Fig. 1b inset) used were:

Republicans: $p_R(b) = 0.6 \tanh[1.05/0.6(b - 0.35)] + 0.42$,

Democrats: $p_D(b) = 0.7 \tanh[1.00/0.7(b + 0.46)] - 0.55$.

Parameter values were:

\[ d_i = 1.3, \ d_o = 0.8, \ \sigma_p = 0.2, \ \sigma_r = 0.15 . \]
In-group fraction scaled 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):

\[
f_D(b) = 0.7 + 0.2b, \\
f_R(b) = 0.7 - 0.2b.
\]

Equations (S12), (S13), (S14), (S15) determined population movement over time, utilizing Euler-Maruyama numerical integration. For finding equilibria, the time constant \(\tau = 1\) was used.

For simulation of Bail’s experiment, 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 distribution. This extra out-group effect was weighted as if it were 24 additional percepts on top of a 100-percept daily diet, i.e., with weight \(f_{bot} = 24/124\). So Eq. (S15) becomes

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

Under this assumption, the time constant \(\tau = 30\) caused movement in agreement with Bail et al. [32]: 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).

**S4. SOME ADDITIONAL EXTENSIONS**

One simple extension is the addition of more groups/parties, such as independent/unaligned individuals and messages. This would require another perception curve, and a three- or more-way fractional content breakdown instead of the single in-group fraction \(f(b)\) as our analysis used.

Additional affiliations such as religion, race, regional identity, etc., may be added to the model without change to the framework. These affiliations would appear as labels that affect the perception curve(s)—since affiliations can change what individuals are exposed to—and inter-group bias, since identity can affect how one reacts to others’ opinions and identity. In particular, the repulsion distance \(d\), representing “trust,” “credulity,” or “benefit of the doubt,” can depend on each affiliation of the individual and of the messenger to allow for more intricate inter-group prejudices than just in-party and out-of-party trust levels. One could of course also add noise to \(d\) values to model individual variation in level of credulity towards other groups. However, given the difficulty of measuring inter-group trust levels, we chose to avoid over-fitting by only utilizing party affiliation in our simulations.

One might also generalize this model to multiple 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. Individuals might also be attracted or repelled from an entire message rather than each position espoused; if a percept engages with three issues and its position on one would be repulsive but on the other two it would be attractive, does the individual shift each position independently, or form a combined judgment of the percept as a whole, and attract or repel from it accordingly?

One might also suggest that individuals perceive a more extreme version of the other party as they become more extreme themselves. This would require additional perception curves for out-group content rather than re-using the same curve for both parties—in particular, the curves that determine out-group content might be *decreasing* with belief instead of increasing. The large dissonance numbers in this case would likely require a different reaction function—one in which repulsion saturates—since repulsion quickly dwarfs attraction under our cubic function. As an additional bonus, these new cross-party perception curves
could incorporate partisan interpretation bias, such that even the same statement could be given a different subjective ideological rating ($p$ value) when individuals believe it comes from an out-party source.

Another potential extension would be the addition of mechanisms by which the perception curves can change. Time-dependence could be introduced to investigate hypotheses about changing 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.

S5. FURTHER WORK

The modular structure of our model is amenable to the incorporation of further data, replacing idealized functions and parameters with empirical distributions from surveys and experiments. For example, to refine the reaction function, further experiments like that of Bail et al. [32] might investigate the impact of political opinions on individuals, and how the messenger’s apparent identity affects the reception of dissonant ideas. As for perceptual environments, the non-network approach 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: each media outlet (or other notable source of political influence) might be assigned an ideology score (as others have done, see, e.g., [25, 26, 29, 38]), and surveys or viewership data could determine which content is consumed in what proportion by each part of the ideological spectrum. Alternately, 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.

Regardless of the approach, any such data-driven realization of this framework will possess greater validity and predictive power as more data is collected. We hope that this modeling framework will afford a better understanding of individual and population-level opinion dynamics, and the feedback effects due to personalized political environments.