

Planned behavior, perceptual biases, and the dynamics of collective action

Rishi Dadlani ^{1,2,*}, John S. McAlister ^{3,2,*†}, Alice C. Schwarze ⁴, Mari Kawakatsu ^{5,6}, Sarah Iams ⁷,
Tahra L. Eissa ^{8,9,‡} and Nina H. Fefferman ^{10,11,2,‡}

¹Department of Mathematics, [University of Pennsylvania](#), Philadelphia, Pennsylvania 19104, USA

²US NSF Center for Analysis and Prediction of Pandemic Expansion (APPEX), [Arizona State University](#), Phoenix, Arizona 85306, USA

³Program in Applied and Computational Mathematics, [Princeton University](#), Princeton, New Jersey 08544, USA

⁴Department of Mathematics, [Dartmouth College](#), Hanover, New Hampshire 03755, USA

⁵Department of Biology, [University of Pennsylvania](#), Philadelphia, Pennsylvania 19104, USA

⁶Center for Mathematical Biology, [University of Pennsylvania](#), Philadelphia, Pennsylvania 19104, USA

⁷John A. Paulson School of Engineering and Applied Sciences, [Harvard University](#), Cambridge, Massachusetts 02138, USA

⁸Department of Applied Mathematics, [University of Colorado Boulder](#), Boulder, Colorado 80309, USA

⁹Department of Physiology and Biophysics, [University of Colorado Anschutz School of Medicine](#), Aurora CO 80045, USA

¹⁰Department of Ecology and Evolutionary Biology, [University of Tennessee, Knoxville](#), Tennessee 37996, USA

¹¹School of Mathematical and Natural Sciences, [Arizona State University](#), Phoenix, Arizona 85306, USA



(Received 21 October 2025; revised 10 July 2026; accepted 9 September 2026; published 7 October 2026)

Classical models of collective behavior often assume that emergent dynamics result from external and observable interactions among individuals. However, how collective dynamics in human populations depend on the internal psychological processes of individuals remains underexplored. Here, we develop a mathematical model to investigate the effects of internal psychology on dynamics of collective action. Our model is grounded in the theory of planned behavior—a well-established conceptual framework in social psychology that links intrinsic beliefs to behavior. By incorporating temporal biases in social perception and individual differences in decision-making processes, we find that the interplay between internal and external drivers of behavior can produce diverse outcomes, ranging from partial participation in action-taking to rapid or delayed behavioral cascades. Predicting these outcomes is, in general, challenging because distinct outcomes can be preceded by transient dynamics that are qualitatively similar to one another. Nevertheless, in the case of two individuals, we derive a complete, analytical classification of long-term outcomes into three mutually exclusive categories. Our model provides a useful test bed for methods that aim to predict the emergence of collective action, and it lays the groundwork for studying the nuanced dynamics of collective human behavior arising from the interaction between psychological processes and observable actions.

DOI: [10.1103/PhysRevE.114.044308](https://doi.org/10.1103/PhysRevE.114.044308)

I. INTRODUCTION

From pandemic management to climate change, societal challenges often require individuals to act collectively. However, understanding when and how the actions of a few individuals give rise to collective action is challenging, in part because of the intricate interplay between internal and external drivers of social behavior. For example, the decision to vote may be driven by an intrinsic belief in civic duty, but it may also be influenced by observing civic-minded neighbors displaying “I voted” stickers [1]. Similarly, the decision to be vaccinated may be driven by a personal concern about the risks of a pathogen, but it may also be influenced by

discussions with peers who fear manipulation by health authorities [2]. While these examples involve a binary choice between *whether* or not to act, other decisions may involve more than two options. For instance, in stock markets, traders observe each other’s value propositions to decide *when* and *how* to act; this produces a communally constructed understanding of the value of a stock, which in turn influences the traders’ decisions to buy or sell that stock (see Ref. [3] and references therein). Understanding such bidirectional interplay between internal, psychological processes and external, observable behaviors is a fundamental challenge when applying complex systems research to human behavior [4–6].

Classical models of complex systems typically assume that collective dynamics arise from observable interactions among individuals. This assumption is appropriate for phenomena driven by reactions to the physical movements of nearby individuals [7], such as flocking in birds [8] or schooling in fish [9]. Many models of emergent dynamics in human groups—including simple contagion [10], complex contagion [11], social influence [12], and persuasion [13]—also focus on external drivers of behavior, making the assumption that individuals’ decisions to act depend on local exposures

*Joint first authors.

†Contact author: jm9097@princeton.edu

‡Joint senior authors.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](#) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

to the actions of their neighbors. While useful, these models do not fully capture how social influence can shape behavior. In particular, there is a need to complement observable interactions between individuals with psychological processes within individuals that govern (1) how the actions of others are perceived, (2) how those perceptions shape an individual's intention to act, and (3) whether or not that intention results in an action.

The *theory of planned behavior* from social psychology offers a conceptual framework that links these psychological dynamics to behavior. This well-established framework posits that an individual's intention to take an action is shaped by three key components: (1) *internal attitude*, defined as the individual's intrinsic belief about whether to take the action or not; (2) *subjective social norm*, defined as the behavioral norm perceived from the observed behavior of others; and (3) *perceived behavioral control*, defined as the perceived degree of control over an intended behavior (or its outcome) [14,15]. This theory provides a basis for studying how group outcomes depend on factors that shape intrinsic dynamics, including perceptual biases and individual differences in the integration of external information. While recent studies on belief formation have examined the role of biases (e.g., temporal biases [16] or bounded confidence [17–20] where biases are shifted toward like-minded individuals) and individual variation (e.g., behavioral diversity in networks where opinions and links coevolve [21,22]), their impact on action-taking dynamics remains underexplored.

Here we draw on the theory of planned behavior to study how intrinsic dynamics affect the frequency and timing of action-taking in social groups. We develop a mathematical model of collective action-taking that incorporates the three components of the theory of planned behavior: internal attitude, subjective social norm, and perceived behavioral control. To examine how temporally biased perceptions affect collective dynamics, we also introduce an immediacy bias, i.e., a bias toward the most recent observations [23,24]. We show that temporal biases in observations and heterogeneity in internal processes lead to diverse collective outcomes, including partial or delayed cascades of action-taking at the population level and modulations in the frequency of action-taking at the individual level. In the simplest case of two individuals, we derive a complete, analytical classification of the collective outcomes. Rooted in psychological theory, our model produces dynamics that are richer and more realistic than in traditional models of behavior that focus exclusively on external drivers of behavior, and it lays the groundwork for studying the nuanced dynamics of collective human behavior arising from the interplay between simple action rules and perceptual biases.

II. MODEL

Drawing on the theory of planned behavior, we develop a piecewise continuous, deterministic model of collective action-taking behavior. Our model incorporates variables that capture individuals' internal attitudes, perceived social norms, and perceived behavioral control, as well as parameters that capture the effects of immediacy (i.e., the most recent observations) on behavior. We will first introduce a general model

TABLE I. Summary of key parameters and variables.

Symbol	Name	Constraint
N	Population size	$N \in \mathbb{Z}_+$
M	Number of actions	$M \in \mathbb{Z}_+$
$x_{i,j}$	Behavioral intention	$x_{i,j} \in (-1, 1)$
$y_{i,j}$	Nudge effect	$y_{i,j} \in [0, 1]$
$z_{i,j}$	Action count	$z_{i,j} \in \mathbb{Z}_{\geq 0}$
$\alpha_{i,j}$	Internal attitude	$\alpha_{i,j} \in (-1, 1)$
$\beta_{i,j}$	Perceived social norm	$\beta_{i,j} \in [0, 1]$
γ_i	Perceived behavioral control	$\gamma_i \in [0, 1]$
σ_A	Strength of internal attitudes	$\sigma_A \geq 0$
σ_S	Strength of perceived social norms	$\sigma_S \geq 0$
σ_C	Strength of perceived behavioral control	$\sigma_C \geq 0$
r	Immediacy parameter	$r \geq 0$
μ_S	Social norm threshold	$\mu_S \in [0, 1]$
μ_C	Baseline perceived control	$\mu_C \geq 0$

with M possible actions. We will subsequently focus on the special case of $M = 1$ to demonstrate the rich dynamics that can arise even from this simple implementation of our model.

A. M -action model

We consider a well-mixed population of N individuals, each of whom can take one of M actions at a given time t . An individual is characterized by their intention to act: *behavioral intention* $x_{i,j}(t) \in [-1, 1]$ captures how strongly individual i intends to take action j at time t . Individuals take actions whenever their behavioral intentions are sufficiently strong, as described below. Taking actions influences others' intentions to act: the *nudge effect* $y_{i,j}(t) \in [0, 1]$ captures how strongly individual i 's actions up to time t nudge other individuals to take action j . The population is thus described by $2NM$ dynamical variables, $x_{i,j}$ and $y_{i,j}$, with $i \in \{1, \dots, N\}$ and $j \in \{1, \dots, M\}$ (see Table I for a summary of key variables and parameters).

To track the number of actions taken over time, we also define *action count* $z_{ij}(t) \in \mathbb{Z}_{\geq 0}$ as the number of times individual i has taken action j up to time t . We initialize the action count as zero for all actions, i.e., $z_{i,j}(0) = 0$ for all individuals i and actions j .

1. Dynamics of behavioral intentions and nudge effects

The theory of planned behavior postulates that behavioral intentions are influenced by three interacting types of behavioral perceptions within an individual: their internal attitude, their perceived social norm, and their perceived behavioral control [14] (Fig. 1). Based on this, we model the dynamics of the behavioral intentions $x_{i,j}(t)$ using the following system of ordinary differential equations (ODEs):

$$\dot{x}_{i,j} = [f_A(\alpha_{i,j}) + f_S(\beta_{i,j})]f_C(\gamma_i)g(x_{i,j}) \quad (1)$$

for $i \in \{1, \dots, N\}$ and $j \in \{1, \dots, M\}$, where the functions f_A , f_S , and f_C , described below, capture the effects of internal attitudes, perceived social norms, and perceived behavioral control, respectively. We implement $f_C(\gamma_i)$ as a multiplicative factor because, while perceived behavioral control can

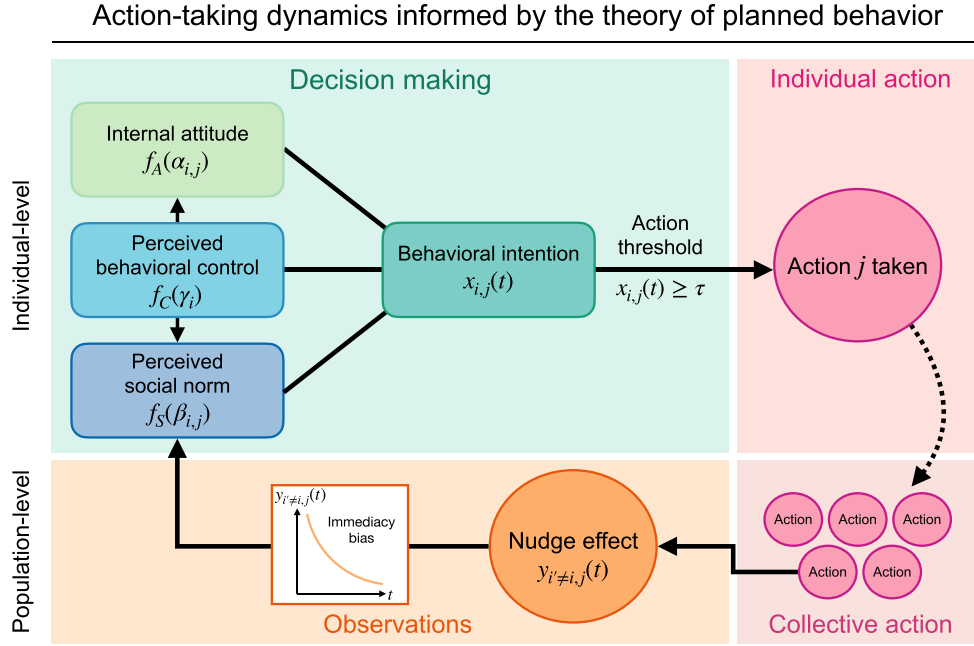


FIG. 1. Model of collective action-taking informed by the theory of planned behavior. Observed actions modulate an individual's internal perceptions of social norms, which, together with internal attitude and perceived behavioral control, shape the individual's behavioral intention. When this intention reaches a threshold, the individual takes an action, which can then be observed by other members of the population and nudge them to act. The impact of others' actions on an individual is weighted in time by an immediacy bias, which prioritizes more recent observations.

amplify the effects of intrinsic motivations (internal attitude) or social drive (perceived social norm), it alone cannot boost behavioral intentions [i.e., if $f_A(\alpha_{i,j}) + f_S(\beta_{i,j}) = 0$, then $\dot{x}_{i,j} = 0$ regardless of the value of $f_C(\gamma_i)$]. The function $g(x_{i,j})$ governs the intrinsic growth of behavioral intentions:

$$g(x_{i,j}) = (1 - x_{i,j})(1 + x_{i,j}). \quad (2)$$

That is, starting from any initial value $x_{i,j,0} \in (-1, 1]$, the behavioral intention $x_{i,j}$ will saturate at $x_{i,j} = 1$ in the absence of effects due to f_A , f_S , and f_C [i.e., when $\dot{x}_{i,j} = g(x_{i,j})$].

With respect to the nudge effects, we implement an immediacy bias, which captures the cognitive tendency to give more weight to information encountered most recently [23]. Thus, in our model, nudge effects $y_{i,j}$ decay over time; for simplicity, we assume that this decay follows an exponential function:

$$\dot{y}_{i,j} = -ry_{i,j}, \quad (3)$$

where $r \geq 0$ is the *immediacy* parameter. When immediacy bias is absent (i.e., $r = 0$), nudge effects remain constant between actions; when immediacy bias is strong (i.e., large r), nudge effects decay quickly.

2. Internal attitudes

The *internal attitude* $\alpha_{i,j} \in (-1, 1)$ captures individual i 's intrinsic propensity toward taking action j (Fig. 1). A larger value of $\alpha_{i,j}$ means that i will take action j more readily even in the absence of social influence. The effect of internal attitudes $\alpha_{i,j}$ on behavioral intentions $x_{i,j}$ is determined by the function f_A . According to the theory of planned behavior, internal attitudes drive self-reinforcement in behavioral

intentions [15]. To reflect this, we assume that f_A satisfies $f_A(0) = 0$ and $f'_A(\alpha_{i,j}) > 0$ for all $\alpha_{i,j} \in (-1, 1)$; that is, in the absence of social influence (perceived control or perceived social norm), individual i 's intention $x_{i,j}$ will grow if i 's intrinsic propensity to take action j is positive ($\alpha_{i,j} > 0$), and $x_{i,j}$ will decay if $\alpha_{i,j} < 0$.

3. Perceived social norm

Nudge effects shape the social norms that individuals perceive toward an action, which, in turn, affects their intentions to act (Fig. 1). To operationalize this, we define the time-dependent *perceived social norm* $\beta_{i,j} \in [0, 1]$ as the average nudge effect toward action j experienced by an individual i :

$$\beta_{i,j} = \frac{1}{N-1} \sum_{\substack{i'=1, \\ i' \neq i}}^N y_{i',j}(t). \quad (4)$$

The effect of the perceived social norm on behavioral intentions is governed by the function $f_S(\beta_{i,j})$, which we assume is increasing in $\beta_{i,j}$; that is, the stronger the social expectations to take action j , the faster the growth of intention $x_{i,j}$.

4. Perceived behavioral control

Finally, *perceived behavioral control* $\gamma_i \in [0, 1]$ measures how easy an individual i believes it will be to take actions. We operationalize this measure as the average level of activity observed by individual i among all others in the population:

$$\gamma_i = \frac{1}{(N-1)M} \sum_{\substack{i'=1, \\ i' \neq i}}^N \sum_{j'=1}^M y_{i',j'}. \quad (5)$$

That is, the perceived behavioral control γ_i for individual i is the observed nudge effects averaged over all other individuals and over all possible actions.

5. Action threshold and immediacy

Individuals take actions whenever their behavioral intentions are sufficiently strong: specifically, individual i takes action j if $x_{i,j} \geq \tau$, where τ is the *action threshold* $\tau \in \mathbb{R}$. When i takes action j at time t , the variables $x_{i,j}$, $y_{i,j}$, and $z_{i,j}$ undergo a discontinuous change:

$$\begin{aligned} x_{i,j}(t+dt) &= 0, \\ y_{i,j}(t+dt) &= 1, \\ z_{i,j}(t+dt) &= z_{i,j}(t) + 1. \end{aligned} \quad (6)$$

B. One-action model

Although the M -action model is highly flexible and can accommodate realistic social settings, our analysis will focus on the one-action model, which produces rich dynamics even in a simple scenario of interest—namely, with a single action ($M = 1$) and linear functions f_A , f_S , and f_C —while remaining sufficiently simple to build intuition and, in the case of two individuals ($N = 2$), sufficiently tractable to yield analytical results, as we will show. We implement a one-action model, described by the 2N dynamical variables $x_i(t)$ and $y_i(t)$ for $i \in \{1 \dots, N\}$, using the following functions:

$$\begin{aligned} f_A(\alpha_i) &= \sigma_A \alpha_i, \\ f_S(\beta_i) &= \sigma_S (\beta_i - \mu_S), \\ f_C(\gamma_i) &= \sigma_C (\gamma_i + \mu_C). \end{aligned} \quad (7)$$

Parameters σ_A , σ_S , and σ_C tune the strength of the effects of internal attitudes, perceived social norms, and perceived behavioral control on behavioral intentions. The parameter μ_S indicates the level of collective action-taking that leads to positive perceived social norms. For example, when $r = 0$, the value of μ_S indicates the fraction of individuals that need to take an action before perceived social norms increase individuals' intentions to take an action. The parameter μ_C indicates the base level of perceived behavioral control when no one has taken an action. When $\mu_C = 0$, a stable fixed point of the population is $x_i = y_i = 0$ for all individuals i . In other words, unless the dynamical system starts from an initial state that is sufficiently far from that trivial fixed point, all individuals in the population remain inactive for all time. These functional forms are chosen for their simplicity but also because they reasonably capture the expected effects of internal attitudes, social norms, and perceived behavioral control. Although these functional forms could include nonlinearities, these additions would require more parameters that are less relevant to our focus on the theory of planned behavior.

When $M = 1$, individuals have a binary choice between taking an action or remaining inactive at each time t , and perceived social norm becomes identical to perceived behavioral control for each individual i and at each time t , i.e., $\beta_i = \gamma_i$, since both quantities reduce to the average nudge

effect experienced by i . Thus, the dynamics of x_i and y_i are governed by

$$\begin{aligned} \dot{x}_i &= [\sigma_A \alpha_i + \sigma_S (\gamma_i - \mu_S)] \sigma_C (\gamma_i + \mu_C) (1 - x_i) (1 + x_i), \\ \dot{y}_i &= -r y_i, \end{aligned} \quad (8)$$

with discontinuous updates

$$\begin{aligned} x_{i^*}(t+dt) &= 0, \\ y_{i^*}(t+dt) &= 1, \\ z_{i^*}(t+dt) &= z_{i^*}(t) + 1, \end{aligned} \quad (10)$$

whenever i^* takes action (i.e., $x_{i^*}(t) \geq \tau$).

III. RESULTS

A. Emergence and dynamics of action-taking cascades

Our overarching goal is to understand how perceptual nuances embedded in the model—social norm, behavioral control, and immediacy bias—affect the emergence and dynamics of action-taking cascades. To study these two aspects of behavioral cascades, we must first understand which individuals, if any, in a given population take actions. We therefore begin by asking two questions: (1) How does variation among the intrinsic action-taking propensities in a population, as captured by internal attitudes α_i , determine who takes actions in the long run? (2) How frequently will individuals act?

1. Variation in internal attitudes produces distinct collective outcomes

To explore how variation in internal attitudes impacts long-term action taking, we first numerically simulate the one-action model in populations of $N = 4$ individuals with different sets of internal attitudes α_i . We find three qualitatively distinct collective outcomes:

- (1) the “no action” scenario, where no individual takes action [e.g., Fig. 2(a)];
- (2) the “partial action” scenario, where some but not all individuals act repeatedly over time while the rest remain inactive [e.g., Fig. 2(b)]; and
- (3) the “full action” scenario, where all individuals take action [e.g., Fig. 2(c)].

To build intuition for these outcomes, suppose that $x(0) = y(0) = 0$ for all individuals. If individual i with the smallest α_i acts at time t , then all other individuals must have also acted no later than individual i (see Lemma A2 in the Appendix). Thus, a sufficient and necessary condition for the “no action” outcome is that no individual is capable of initiating the first action. In the absence of nudge effects, individual i will never act if $\sigma_A \alpha_i - \sigma_S \mu_S \leq 0$. This observation leads to the following result, which is proved as Corollary A4 in the Appendix:

Corollary 1. A system falls under the “no action” classification if and only if the individual i with the largest α_i satisfies $\sigma_A \alpha_i - \sigma_S \mu_S \leq 0$ and $x_j(0) < \tau$ for all individuals j .

In our example “no action” scenario [Fig. 2(a)], individuals 1–3 have positive internal attitudes α_i , but all values of α_i fall short of the social-norm baseline

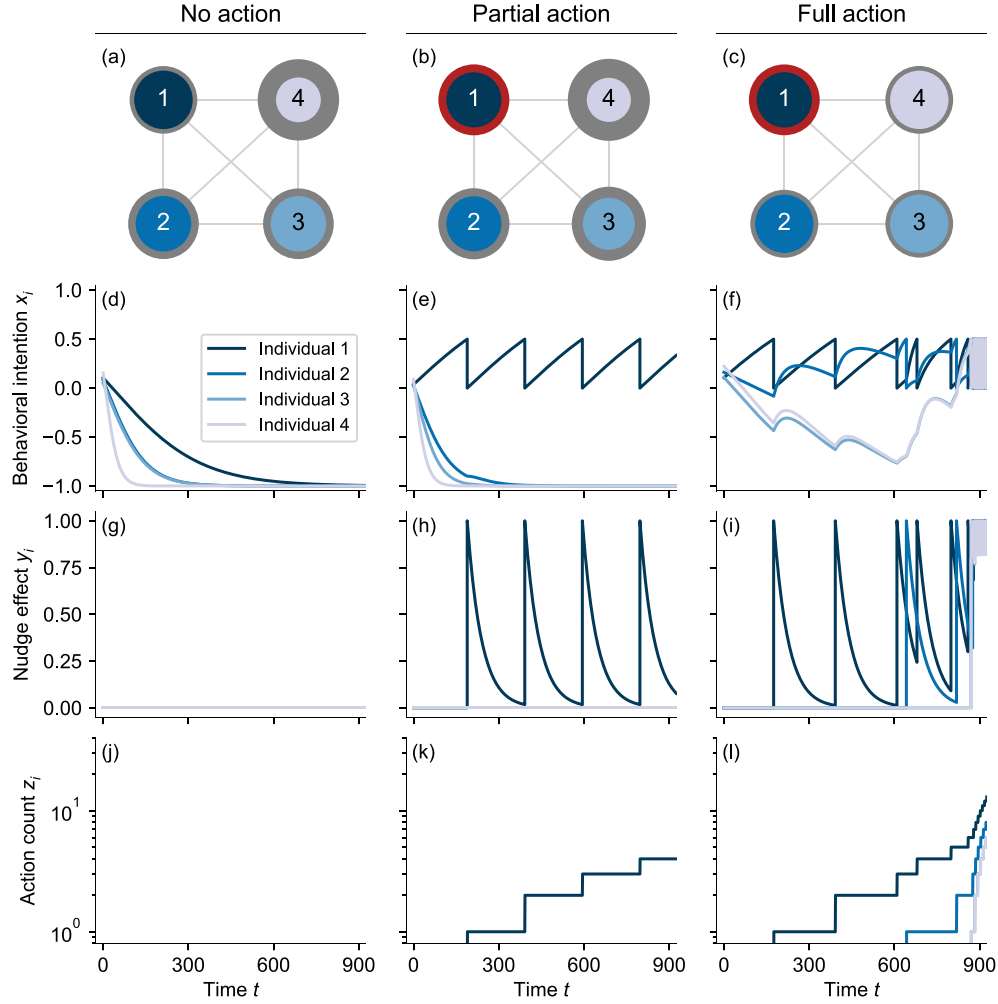


FIG. 2. Collective outcome depends on internal attitudes. We performed numerical simulations of the one-action model with $N = 4$ individuals and with different distributions of internal attitudes α_i . Columns show three qualitatively distinct outcomes: “no action” (left column), “partial action” (middle column), and “full action” (right column). (a)–(c) Circles with red (gray) edges indicate individuals whose internal attitude is larger than (smaller than) the baseline activation threshold $\sigma_S \mu_S / \sigma_A$ for an inactive population. Edge widths denote the magnitude of the difference between their internal attitude and baseline threshold $|\alpha_i - \sigma_S \mu_S / \sigma_A|$. (d)–(l) Behavioral intentions x_i (d)–(f), nudge effects y_i (g)–(i), and action counts z_i (j)–(l) of individuals $i \in \{1, 2, 3, 4\}$, indicated by the colors. Parameters: $\sigma_A = 1$, $\sigma_S = 1$, $\sigma_C = 0.05$, $\mu_S = \mu_C = 0.5$, and $r = 0.02$. Initial conditions: $x_i(0) \sim U(0, 0.25)$ and $y_i(0) = 0$.

threshold $\sigma_S \mu_S / \sigma_A$ that would trigger an individual to increase their intention x_i . As a result, the intentions of all individuals decrease over time [Fig. 2(d)], and the population remains inactive [Figs. 2(g) and 2(j)]. That is, if $\sigma_A \leq \sigma_S \mu_S / \max\{\alpha_i\}$, then the system is certainly in the “no action” state.

In the “partial action” scenario [Fig. 2(b)], individual 1’s internal attitude surpasses the baseline social-norm threshold ($\alpha_1 > \sigma_S \mu_S / \sigma_A$), so their intention increases from time $t = 0$, eventually leading to sustained action-taking [Figs. 2(e) and 2(k)]. The actions of individual 1 produce nudge effects $y_1(t)$ [Fig. 2(h)]. However, these nudge effects are not strong enough to overcome the low internal attitude of individuals 2–4, whose intentions decrease rapidly. Thus, the population splits into a single active individual (individual 1) and a cluster of inactive individuals (individuals 2–4).

In contrast, in the “full action” scenario [Fig. 2(c)], individuals 2–4 have sufficiently large internal attitudes α_i so

that, although their intentions initially decrease, the successive nudge effects eventually lead all individuals to take actions [Fig. 2(f)]. As each additional individual becomes active, nudge effects increase [Fig. 2(i)] and all active individuals increase their action-taking frequency [Fig. 2(l)]. For the N -individual case, a sufficient condition for full action is given by Corollary 2, which is proved as Corollary A5 in the Appendix.

Corollary 2. If $\sigma_A \alpha_i - \sigma_S \mu_S > 0$ for all i , then the system is classified under the full action state.

An intuitive example of Corollary 2 with $N = 2$ is illustrated in Fig. 3. Here, the behavioral intentions of both individuals 1 and 2 increase in the absence of nudge effects, such that both individuals eventually act. However, the condition $\sigma_A \alpha_i - \sigma_S \mu_S > 0$ for all i is not necessary for the population to achieve full action. For example, in Fig. 2(c), we observe full action even though $\sigma_A \alpha_i - \sigma_S \mu_S < 0$ for individuals $i = 2, 3$, and 4. Thus, the emergence of

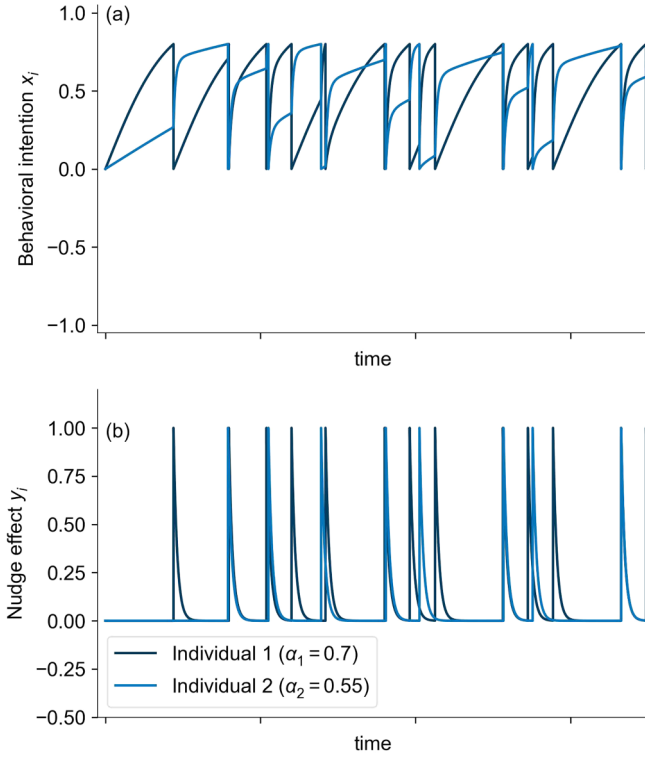


FIG. 3. Increasing behavioral intentions in the absence of nudge effects leads to full action. Behavioral intention (a) and nudge effect (b) for $N = 2$ individuals. Because behavioral intentions of both individuals increase even in the absence of nudge effects, both eventually act, leading to the “full action” outcome (Corollary 2).

action-taking cascades depends both on the internal attitudes of individuals and on recruitment via nudge effects, whereby recruitment of more individuals leads to increased nudge effects and, correspondingly, increased action-taking frequency in the population.

2. Internal attitudes and nudge effects alter the frequency of actions

Our analysis so far has focused on classifying collective-level outcomes in populations with different distributions of

internal attitudes. A related but distinct question is: how does heterogeneity in internal attitudes within a population affect the behavior of individuals themselves? Specifically, how do internal attitudes affect the *frequencies* with which individuals act? To address these questions, we simulate a population of $N = 100$ individuals whose internal attitudes α_i are sampled from a uniform distribution. In general, individuals with larger α_i tend to activate sooner [Fig. 4(a)] and act more frequently [Fig. 4(b)] than those with smaller α_i . Crucially, however, we observe a discontinuous transition in frequency of action-taking [f_i ; Fig. 4(b)]: individuals with small α_i remain inactive over time ($f_i = 0$), and it is only those with α_i above a threshold that take any action ($f_i > 0$). This reflects a competition between internal attitudes and nudge effects: when social reinforcement via nudge effects is initially absent [$y_i(0) = 0$], the behavioral intentions x_i of those with limited intrinsic motivation (α_i small) decay quickly toward -1 , such that x_i may not recover even as nudge effects begin to accumulate. This results in a heterogeneous population where active and inactive individuals coexist. Among those who become active, the frequency of action-taking increases with α_i [Fig. 4(b)]. Note, however, that the threshold observed empirically in Fig. 4(b) is lower than the baseline threshold for a single individual ($\sigma_s \mu_s / \sigma_A$). This is because nudge effects enable even individuals with α_i slightly below the baseline threshold to eventually act.

B. Population-level effects of internal attitudes, perceived social norms, and perceived behavioral control

We have shown that internal attitudes are a key driver of action-taking at the individual level. To characterize the dynamics at the population level, we now consider all three elements of the theory of planned behavior—internal attitudes, perceived social norms, and perceived behavioral control—and study how they impact the collective dynamics of action-taking.

We perform numerical simulations in populations of $N = 100$ individuals with varied strengths of the different components, and we compute the mean frequency of action-taking in the population (Fig. 5). The mean frequency of action-taking tends to increase with the strength σ_A of internal attitudes

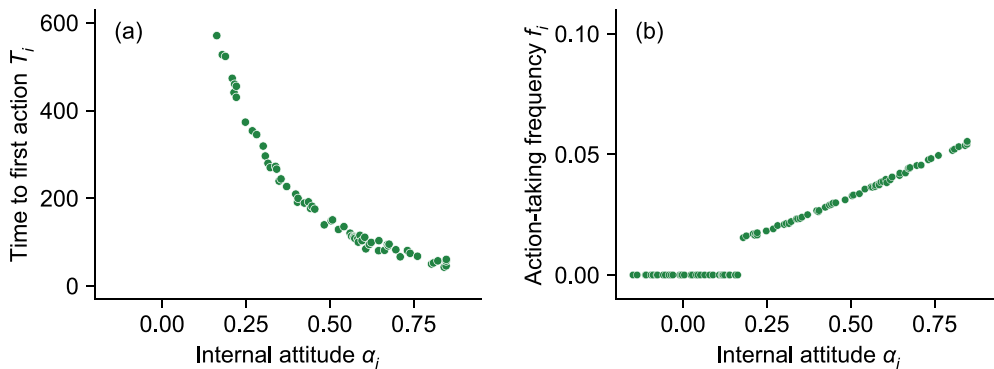


FIG. 4. Stronger internal attitudes promote earlier and more frequent action-taking. We performed numerical simulations of the one-action model with $N = 100$ individuals and with $t = 3000$ time steps. Each dot denotes a single individual i . (a) Time to first action T_i as a function of internal attitude α_i . (b) Time-averaged action-taking frequency f_i as a function of internal attitude α_i . Parameters: $\sigma_A = 1$, $\sigma_s = 1$, $\sigma_C = 0.05$, $\mu_s = 0.5$, $\mu_C = 0.5$, $r = 0.02$, $\tau = 0.5$, and $\alpha_i \sim 0.35 + U(-0.5, 0.5)$. Initial conditions: $x_i(0) \sim U(0, 0.25)$ and $y_i(0) = 0$.

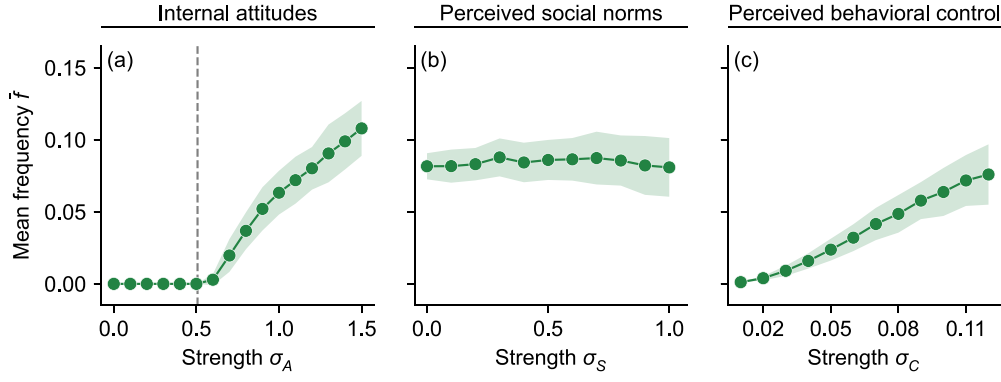


FIG. 5. Frequency of action-taking tends to increase with strengths of internal attitudes and perceived behavioral control, but is insensitive to the strength of perceived social norms. Panels show ensemble mean action-taking frequencies $\bar{f}$ ($\pm$ SD) across 100 replicate simulations, each in a population of $N = 100$ individuals and with $t = 5000$ time steps. (a) We vary the strength σ_A of internal attitudes, with $\sigma_S = 1$ and $\sigma_C = 0.1$. The gray dashed line shows the critical σ_A at which a phase transition occurs (see Corollary 1). (b) We vary the strength σ_S of perceived social norms, with $\sigma_A = 1$ and $\sigma_C = 0.12$. (c) We vary the strength σ_C of perceived behavioral control, with $\sigma_A = 1$ and $\sigma_S = 1$. Other parameters: $\mu_S = 0.5$, $\mu_C = 0.5$, $r = 0.02$, $\tau = 0.5$, and $\alpha_i \sim U(-0.4, 1)$. Initial conditions: $x_i(0) \sim U(0, 0.5)$ and $y_i(0) = 0$.

[Fig. 5(a)]. Importantly, however, the onset of action-taking is marked by a phase transition in σ_A . For subcritical values of σ_A , no individual takes actions (i.e., $\bar{f} = 0$). Once σ_A exceeds a critical value σ_A^* , the mean action-taking frequency increases with the strength of internal attitudes. The critical value σ_A^* has a lower bound $\chi_A = \sigma_S \mu_S / \max_i \{\alpha_i\}$ [dashed gray line in Fig. 5(a)], which we derive from Corollary 1. No actions are taken when $\sigma_A < \chi_A$ because no individual in the population has a sufficiently positive internal attitude to take the first action. Unlike for internal attitudes, we observe no phase transitions associated with the strength σ_S of perceived social norms. Instead, the mean frequency of action-taking $\bar{f}$ remains approximately constant across values of σ_S [Fig. 5(b)]. We can understand this result in terms of the two competing effects of σ_S : increasing σ_S strengthens the social reinforcement of action-taking when nudge effects are strong, but doing so also strengthens the social suppression of action-taking when nudge effects are weak. In the parameter regime considered here, these two effects of σ_S appear to roughly cancel one another. In contrast to the effect of σ_S , the mean action-taking frequency increases monotonically with the strength σ_C of perceived behavioral control [Fig. 5(c)]. This is because σ_C acts as a timescale parameter in the dynamics of the intentions x_i [Eq. (8)]. In sum, the mean frequency of action-taking tends to increase with the strength of internal attitudes or of perceived behavioral control, but is insensitive to the strength of perceived social norms.

C. Classifying collective outcomes for $N = 2$

To more explicitly determine how and when each collective outcome occurs, we focus on the case of two individuals ($N = 2$). In this setting, we are able to obtain a complete classification of the collective outcomes. We consider the regime where one individual acts independently (i.e., in the absence of nudge effects) while the other does not, and we ask: under what conditions does social reinforcement lead the otherwise inactive individual into sustained action? We focus on this regime because the remaining possibilities, i.e., both individuals acting independently or neither acting inde-

pendently, are already characterized by Corollaries 1 and 2. Without loss of generality, we assume $\sigma_A \alpha_1 - \sigma_S \mu_S > 0$ and $\sigma_A \alpha_2 - \sigma_S \mu_S \leq 0$. We also make the following assumptions:

$$\begin{aligned} \sigma_A \alpha_1 - \sigma_S \mu_S &> 0 \quad (h1), \\ \sigma_A \alpha_2 - \sigma_S \mu_S &\leq 0 \quad (h2), \\ x_1(0) = y_1(0) = y_2(0) &= 0 \quad (h3), \\ x_2 \text{ has not yet acted} &\quad (h4). \end{aligned}$$

Assumption (h4) implies that $x_2(0) < \tau$, and, as we will show, if (h4) is relaxed and x_2 acts, then x_2 will act again.

1. Derivation of an explicit solution

Let T denote the time at which individual 1 acts for the first time (see Corollary A1). Prior to the first action of individual 2—that is, under assumption (h4)—individual 1 is unaffected by x_2 and acts at every positive integer multiple of T . We therefore refer to T as the *period* of x_1 . In Lemma 1 we derive an explicit expression for x_2 , valid until either one period of x_1 elapses or x_2 acts. This result is proved as Lemma A3 in the Appendix.

Lemma 1. Assume (h1), (h3), and (h4), and suppose that individual 1 acts at time kT . Then, for $t \in [0, T]$ and initial condition $u_0 = x_2(kT)$, and provided that individual 2 does not act in the interval $[kT, kT + t]$, an explicit solution for x_2 is given by $x_2(kT + t) = V(t; u_0)$, where

$$V(t; u_0) = \tanh(At + Be^{-rt} + Ce^{-2rt} + D(u_0)),$$

with $A = [\sigma_A \alpha_2 - \sigma_S \mu_S] \sigma_C \mu_C$, $B = -\frac{1}{r}[(\sigma_A \alpha_2 - \sigma_S \mu_S) \sigma_C + \sigma_S \sigma_C \mu_C]$, $C = -\frac{\sigma_S \sigma_C}{2r}$, and $D(u_0) = \tanh^{-1}(u_0) - B - C$.

This solution describes the position of x_2 at time $kT + t$, where $k \geq 1$, given starting position u_0 at time kT . This is valid until whichever happens first: $(k + 1)T$ or a time where x_2 crosses the threshold.

2. Useful invariant

The key quantity for determining whether individual 2 is recruited into sustained action is the net change in individual 2's behavioral intention x_2 over a period established by

individual 1. To characterize this net change, we consider the following lemma, which is proved as Lemma A4 in the Appendix.

Lemma 2. Suppose (h1), (h2), and (h4), and suppose that individual 2 does not act on the interval $[kT, (k+1)T]$. Then the following are true:

$$0 < V(T; 0) \Rightarrow u_0 < V(T; u_0),$$

$$0 = V(T; 0) \Rightarrow u_0 = V(T; u_0),$$

$$0 > V(T; 0) \Rightarrow u_0 > V(T; u_0),$$

for any $u_0 \in (-1, 1)$.

Here, $V(T; u_0) - u_0$ gives the net change in x_2 in a single period. Lemma 2 thus states that, while individual 2 has not yet acted, x_2 's behavior on any period of x_1 —net increase, net decrease, or no change over the course of one period (ignoring the threshold)—does not depend on x_2 's starting position. That is, if, prior to its first action, x_2 exhibits a net increase (decrease / no net change) on any given period of x_1 , then x_2 will exhibit the same qualitative behavior over *all* periods of x_1 . To reach this result, we examine the explicit solution from Lemma 1 and we find that, because of the monotonicity of $\tanh^{-1}$, the inequality $u_0 < V(T; u_0)$ is equivalent to an inequality which depends only on system parameters and T . In particular, the sign of this net change is given by comparing $B + C$ with $AT + Be^{-rT} + Ce^{-2rT}$. The details are found in the Appendix in the proof of Lemma A4. For notational convenience, we define $M := AT + Be^{-rT} + Ce^{-2rT} - B - C$. Below, we show that, under assumptions (h1)–(h4), $M \leq 0$ corresponds to the partial action state, and $M > 0$ to the full action state.

3. Classification of partial action

To show that $M \leq 0$ corresponds to the partial action state, we consider the shape of the trajectory of x_2 during a single period. The key observation is that the trajectory has at most one critical point and that this critical point will be a local maximum. This can be seen by recalling that the nudge effect for a single period $y(kT + t) = e^{-rt}$ is strictly decreasing. Now when we examine the dynamics of x on $(-1, 1)$ we see that

$$\begin{aligned} \dot{x}_2(kT + t) &= [\sigma_A \alpha_2 + \sigma_s(e^{-rt} - \mu_s)](\sigma_c(e^{-rt} + \mu_c)) \\ &\quad \times (1 - x_2)(1 + x_2). \end{aligned} \quad (11)$$

Here, the last three factors are always positive (when $x_2 \in (-1, 1)$ (recall that $\mu_c \geq 0$). Thus, the sign of $\dot{x}_2$ is determined entirely by $[\sigma_A \alpha_2 + \sigma_s(e^{-rt} - \mu_s)]$, which is strictly decreasing. This means that x_2 has at most one critical point within a period: x_2 is either increasing then decreasing in one period with a critical point found explicitly in Lemma A5, decreasing on the entire interval, or increasing on the entire interval. In the case where $M \leq 0$ we know that x_2 cannot be increasing on the entire interval. When $M \leq 0$, having an explicit expression for the critical point allows us to show, in Corollary A6, that $x_2(t) \leq x_2(0)$ for all $t > 0$. We argue this by showing that on $[0, T]$, x_2 is decreasing and using the critical point from Lemma A5 we know any local maxima of x_2 must be smaller than $x(0)$. That is to say, x_2 cannot be nudged up in a single period more than it fell in the interval

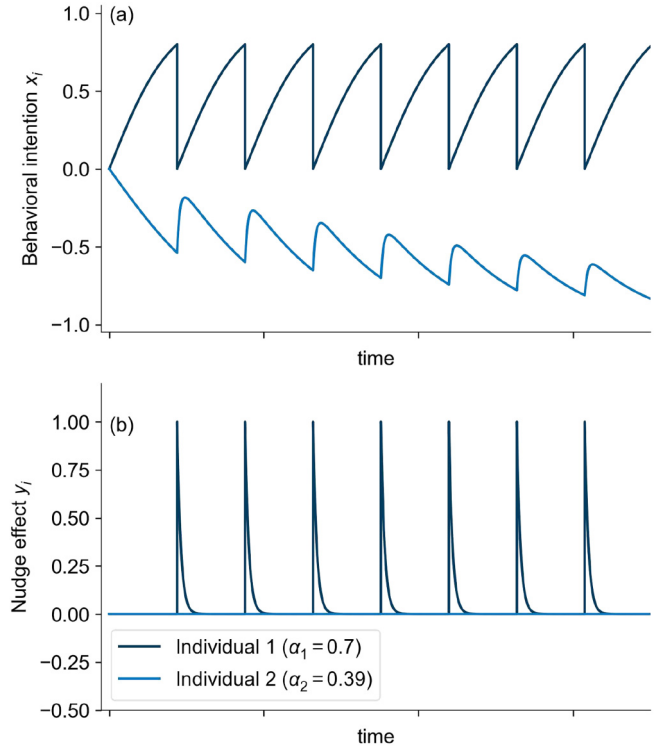


FIG. 6. Partial action in a two-individual system. Behavioral intention (a) and nudge effect (b) for a system with $N = 2$. With each action of individual 1, individual 2's behavioral intention initially increases, but decreases further than it increased within the period determined by individual 1. This results in a net decrease for individual 2, and the solution is classified as partial action.

$[0, T]$. Because of this, we know that as long as x_2 starts below the threshold, τ , it will not act if $M \leq 0$ as in Fig. 6. This result is stated below and proved in the Appendix as Theorem A1.

Theorem 1. Suppose that (h1)–(h3) hold and that $x_2(0) < \tau$. If $M \leq 0$, then the system is classified under partial action.

4. Classification of full action

The case $M > 0$ corresponds to the full action state. Establishing this requires two steps: first, showing that individual 2 acts once, and second, showing that individual 2 continues to act indefinitely. From Lemma 1, we consider

$$V(T, u_0) = \tanh(AT + Be^{-rT} + Ce^{-2rT} + D(u_0))$$

as a function of u_0 , where A, B, C , and D are defined as in Lemma 1. It is helpful to shift perspective from the continuous-time dynamics to the time evolution of x_2 from one period to the next. To this end, we rewrite the function above as

$$f(x) := \tanh(M + \tanh^{-1}(x)).$$

The function f maps x to its value after one period. For example, we have $f(x_2(kT)) = x_2((k+1)T)$, provided that individual 2 does not act during the period $[kT, (k+1)T]$. We can use standard tools from the theory of difference equations to prove the following result. The proof is included in the Appendix under Lemma A7.

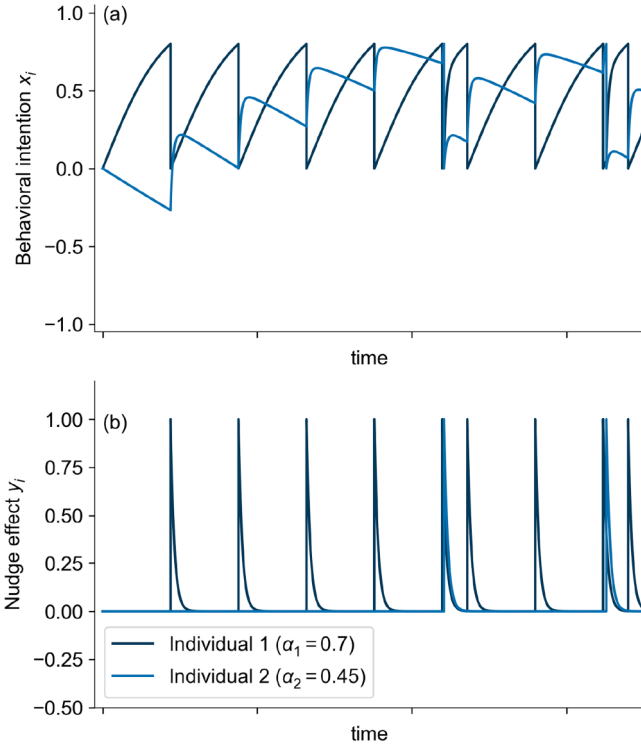


FIG. 7. Full action in a two-individual system. Behavioral intention (a) and nudge effect (b) for a system with $N = 2$. With each action of individual 1, individual 2's behavioral intention initially increases and, although it decreases, the result is a net increase for individual 2's behavioral intention. Thus, individual 2 will act. Knowing that individual 2 acted once is sufficient for us to conclude that individual 2 will continue to act indefinitely.

Lemma 3. Let w_n be a sequence of real numbers in $(-1, 1)$ satisfying $w_n = f(w_{n-1})$. If $M > 0$, then $w_n \rightarrow 1$, if $M < 0$, then $w_n \rightarrow -1$, and if $M = 0$, then $w_n = w_0$.

Lemma 3 immediately implies that, if $M > 0$, $x_2(t)$ will eventually surpass any threshold $\tau < 1$. The proof is straightforward and included in the Appendix under Corollary A7.

Corollary 3. If $M > 0$ and $\tau < 1$, then there is a time at which $x_2(t) = \tau$.

Lemma 3 and Corollary 3 show that, under assumptions (h_1) – (h_4) , x_2 will eventually act. To show that this ultimately leads to the full action state, we relax assumption (h_4) and show that x_2 will act even if it has already acted. The key idea is that an action by individual 2 generates a nudge effect on individual 1, and this nudge effect cannot cause individual 1 to act less frequently. As a result, individual 2 receives nudges from individual 1 *at least as* frequently as before individual 2 first acted. This means that x_2 will continue to experience a net increase between successive actions of individual 1, since $M > 0$ implies a net increase in x_2 over a period even prior to individual 2's first action (Fig. 7). The following Theorem formalizes this argument; the proof follows directly from the previous two results and is included in the Appendix as Theorem A2.

Theorem 2. If (h_1) – (h_3) hold and $M > 0$, then individual 2 will act infinitely many times, thus establishing the full action state.

TABLE II. Complete classification of collective outcomes for $N = 2$. We assume $\alpha_1 \geq \alpha_2$.

$\sigma_A \alpha_1 - \sigma_S \mu_S$	$\sigma_A \alpha_2 - \sigma_S \mu_S$	M	Classification
≤ 0	≤ 0		No Action
> 0	< 0	≤ 0	Partial Action
> 0	< 0	> 0	Full Action
> 0	≥ 0		Full Action

We can now summarize the conditions for the no action, partial action, and full action states in Table II.

These analytical conditions partition the parameter space into regions associated with the three collective outcomes. As illustrated in Fig. 8, the analytical boundaries show strong agreement with numerical simulations. Although the same type of analytical classification is not performed for the general N -individual case due to its complexity, a numerical approach can be used to distinguish between the qualitatively different behaviors observed in the three-individual system, which follow separations similar to the two-individual system (Fig. 9).

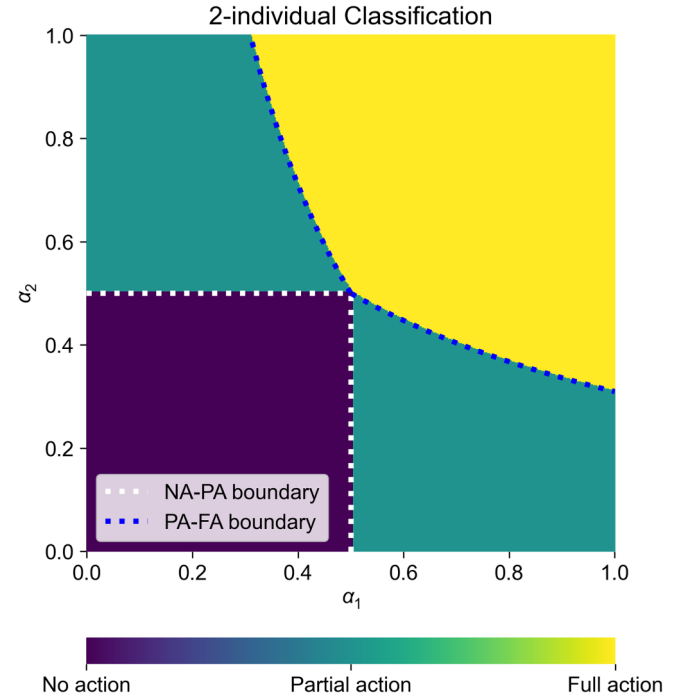


FIG. 8. Analytical classification partitions parameter space into distinct regimes of collective outcomes ($N = 2$). Classification of collective outcomes in the parameter space $(\alpha_1, \alpha_2) \in [0, 1]^2$ with parameters $\sigma_S = 1$, $\sigma_C = 1$, $r = 0.86$, $\mu_C = 0.05$, and $\tau = 0.8$. Solid colors denote outcomes observed in numerical simulations as shown. White dotted segments denote the analytical boundary between no action and partial action (NA-PA). Blue dotted curves denote the analytical boundary between partial action and full action (PA-FA). The analytical boundaries are consistent with the transitions between qualitatively distinct behaviors observed in numerical simulations.

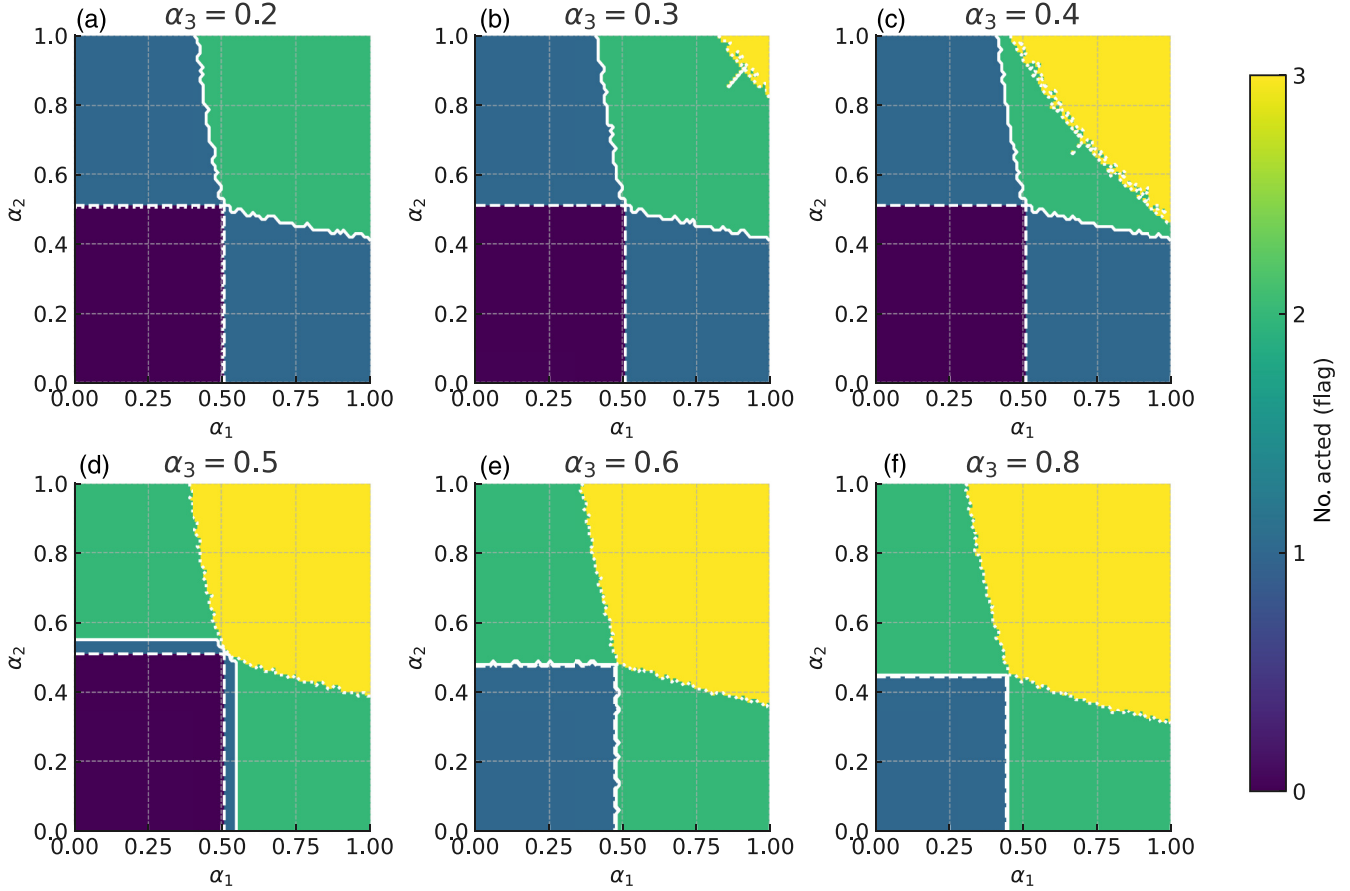


FIG. 9. Numerical classification of collective outcomes for $N = 3$. We classify the simulated behavior of the three-individual system by the number of individuals taking action, denoted by colors. Panels (a)–(f) correspond to different values of α_3 ranging from 0.2 to 0.8; within each panel, α_1 and α_2 are varied ($(\alpha_1, \alpha_2) \in [0, 1]^2$).

D. Delay and inhibition of action cascades via immediacy bias

We have characterized how each component of the theory of planned behavior—internal attitudes, perceived social norms, and perceived behavioral control—affects the frequency of action-taking and the collective outcome, assuming that immediacy bias is weak ($r = 0.02$) and therefore nudge effects decay slowly. However, the strength of immediacy bias may impact whether or not individual actions will induce sustained collective action. To explore this possibility, we analyze the effect of the immediacy parameter r on the emergence of action-taking cascades.

In the absence of immediacy bias (i.e., $r = 0$), the active cluster of a population establishes itself very quickly [Figs. 10(a), 10(d), and 10(g)]. When $r = 0$, the nudge effects y_i never decay; they follow a step function [Fig. 10(d)]. Once an individual takes their first action, they exert the strongest possible nudge effect that an individual can exert on the population, and they will continue to do so for all time. Following the first action in a population, all individuals who can be persuaded to take an action because of a perceived positive social norm will do so, thus increasing their intentions quickly [Fig. 10(a)] and joining the active cluster in rapid succession [Fig. 10(g)].

When $r > 0$, nudge effects decay exponentially with time [Figs. 10(e) and 10(f)], with small (positive) values of r

corresponding to a slow decay of nudge effects [Fig. 10(e)]. Whenever an action occurs, the entire population briefly experiences a strong nudge effect that can lead individuals to increase their intentions [Fig. 10(b)]. However, since the intensity of the nudge effect decays exponentially, the perceived strong social norm is short-lived. If the perceived social norm does not cause other individuals to take action immediately, then it can be necessary for the first action taker to take several actions before a second individual is persuaded to join the active cluster.

When r is sufficiently large, immediacy bias inhibits the emergence of collective action [Figs. 10(c), 10(f), and 10(i)]. The fast decay of nudge effects [Fig. 10(f)] prevents the emergence of positive perceived social norms that would lead individuals to increase their intentions [Fig. 10(c)] and, eventually, take action. However, individuals who have a very large internal attitude will still take actions not motivated by social influence [Fig. 10(i)].

In sum, the number of individuals who take action decreases with the strength of immediacy bias, and full participation in action-taking occurs only when the immediacy parameter is below a certain threshold such that all individuals in the population receive sufficiently strong nudges to act. Above this threshold, the immediacy bias inhibits full action and, at even greater values of r , inhibits all action

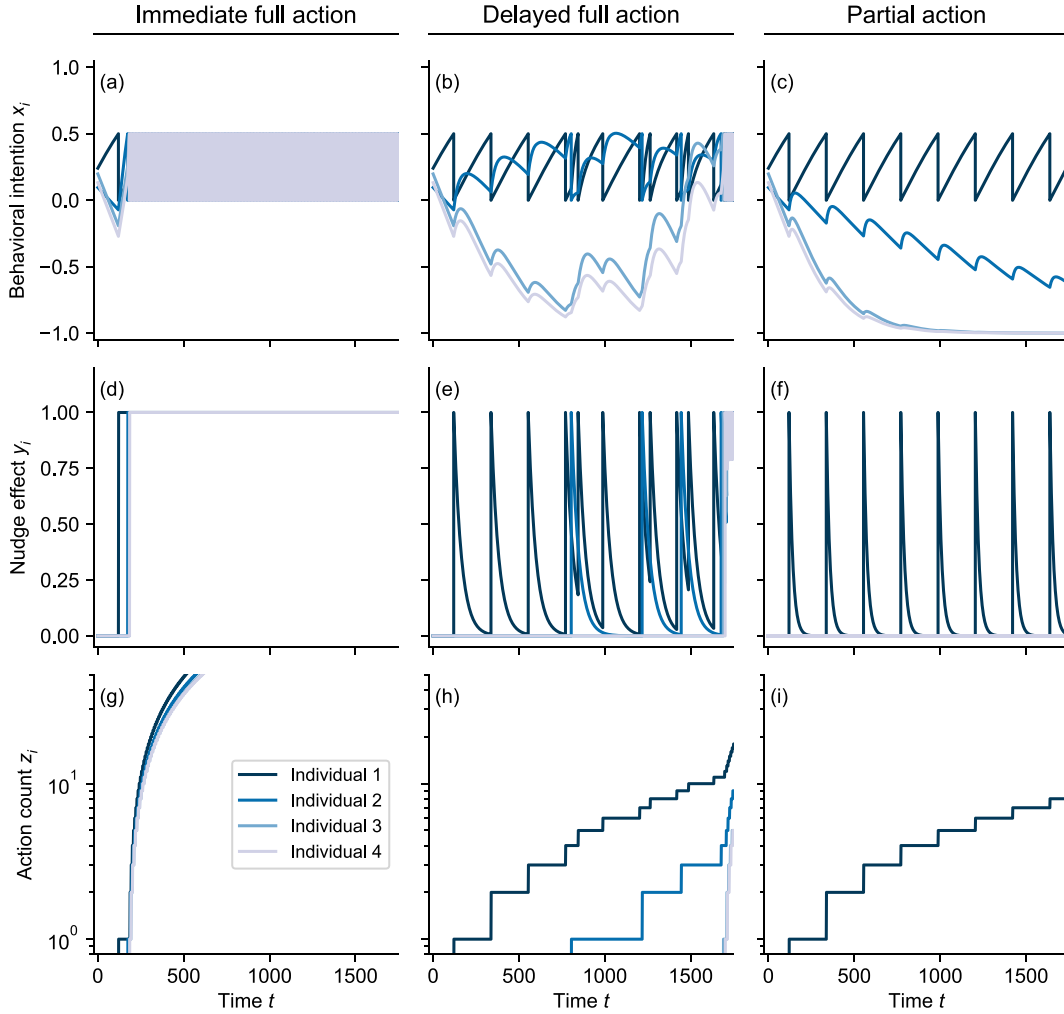


FIG. 10. Immediacy bias alters action-taking dynamics. Behavioral intentions x_i (a)–(c), nudge effects y_i (d)–(f), and action counts z_i (g)–(i) of individuals $i \in \{1, 2, 3, 4\}$, indicated by the colors. The internal attitudes α_i of the four individuals are sampled from $U(-0.4, 1)$. The sample is identical to the values in Fig. 2(c), with $\alpha_1 \approx 0.60$, $\alpha_2 \approx 0.44$, $\alpha_3 \approx 0.37$, and $\alpha_4 \approx 0.36$. Parameters: $\sigma_A = 1$, $\sigma_S = 1$, $\sigma_C = 0.05$, and $\mu_S = \mu_C = 0.5$. The initial conditions $x_i(0) \sim U(0, 0.5)$ and $y_i(0) = 0$ are identical in all three simulations. We obtained the results for immediate full action, delayed full action, and partial action with immediacy parameters $r = 0$, $r = 0.023$, and $r = 0.05$, respectively.

within the collective. We observe these trends with increasing r for almost all combinations of model parameters and initial conditions.

Not only does the immediacy bias parameter's ability to control the emergence of collective action illustrate the crucial role of temporal biases in collective decision-making, but our results in Figs. 2 and 10 highlight the role of intrinsically motivated individuals in kickstarting a behavioral cascade. Individuals who will act even in the absence of nudge effects—because their internal attitudes are sufficiently strong—can nudge other less intrinsically motivated individuals to act, even when the latter would otherwise remain inactive. This leads to an observed cascade in which successively activated individuals contribute to an increasing collective nudge (Fig. 11).

To explain how immediacy bias modulates the dynamics of behavioral cascades, we consider a population in the absence of initial nudge effects, and apply an *immediacy window*, which we define as the time it takes for the nudge effect y_i

of individual i to decay from 1 to 0.01 (1%) in the absence of any additional actions by i or other individuals. If immediacy bias is weak (i.e., r is small) and therefore the immediacy window is long, and there is an underlying diversity of intrinsic attitudes among the population, then the feed-forward cascade results in both more effective recruitment to action and an acceleration in action taking [Figs. 11(a) and 11(c)]. In contrast, if the immediacy window is short (i.e., r large) and the internal attitudes are appropriately spaced, then the most innately motivated individual will act at approximately the frequency set by their internal motivation alone between the actions of others [Figs. 11(b) and 11(d)]. Each action by the most innately motivated individual causes a pulse-like increase in behavioral intentions of other individuals [Fig. 11(d)], which leads them to also act in an approximately periodic manner [Fig. 11(b)]. If, however, the immediacy window is of intermediate length (i.e., r intermediate), then the timing of individual actions may become less tightly correlated without entering a periodic, or approximately periodic, regime. In sum, the interactions

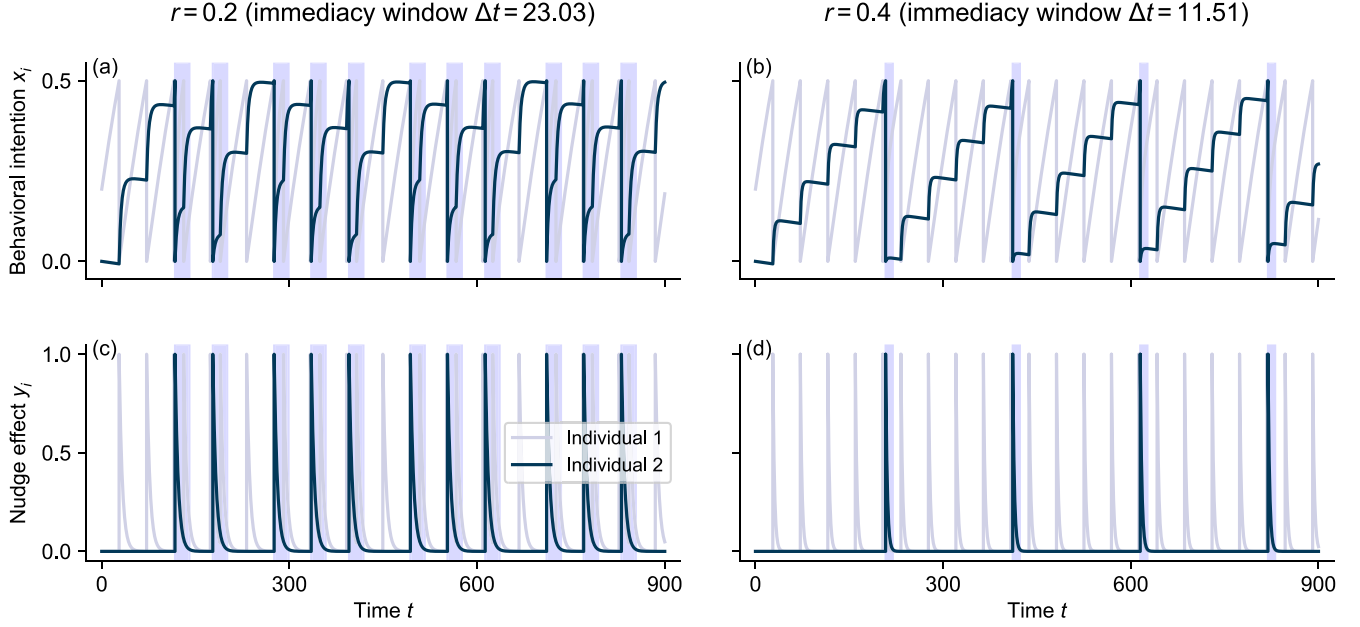


FIG. 11. Successive action-taking modulates with the immediacy window. Behavioral intentions x_i (a), (c) and nudge effects y_i (b), (d) from simulations with $N = 2$ individuals. Each shaded region denotes the *immediacy window* Δt following an action by individual 2, where the immediacy window is the time it takes for an individual i 's nudge effect y_i to decay from 1 to 0.01 (1%) in the absence of any additional actions by i or other individuals (i.e., $\Delta t = -\log_{10}(0.01)/r$). Other parameters: $\mu_S = 0.5$, $\mu_C = 0.5$, $\alpha_1 = 1$, $\alpha_2 = 0.49$, $y_1(0) = y_2(0) = 0$, $x_1(0) = 0.2$, and $x_2(0) = 0$.

among individuals with different internal attitudes, the immediacy window, and the nudge effect give rise to interesting periodicity in the frequency modulation of action-taking at the population level and collective dynamics that would not be observable without considering a more nuanced coupled model with both internal and external drivers of behavior, such as that posed by the theory of planned behavior.

IV. DISCUSSION

We have developed a mathematical model to explore the dynamics of collective action in human populations. This model is grounded in the theory of planned behavior and incorporates well-established temporal biases in human perception.

We find that the three components of the theory of planned behavior—internal attitudes, perceived social norms, and perceived behavioral control—affect collective dynamics in distinct ways. Varying the strength of perceived behavioral control results in a gradual increase in the frequency of action-taking, whereas varying the strength of internal attitudes results in a phase transition, such that individuals remain inactive until their attitudes cross a critical threshold. Interestingly, however, the strength of perceived social norms has little effect on the frequency of action-taking. This likely reflects a balance between the two competing effects of social norms: they reinforce action-taking when nudge effects are strong, but they also reinforce inaction when nudge effects are weak.

The immediacy bias also exhibits a threshold effect: while individuals with a negative internal attitude never take action if the immediacy bias is strong, they can be compelled by

a perceived social norm to take an action if the immediacy bias is weak enough for the nudge effects to persist for a sufficiently long window of time. Intriguingly, the transient dynamics of behavioral intentions and nudge effects can look similar between scenarios where all individuals eventually take action (“delayed full action” in Fig. 10) and scenarios where only a subset of individuals take action (“partial action” in Fig. 10). It can thus be difficult to predict sustained collective action from the initial activity of individuals alone. In this regard, the dynamics of our model mirror those of real-world societal trends and social movements—whose long-term success or failure is often unpredictable in their early stages. This work can therefore be an initial step toward understanding collective behavior and engagement in social movements, such as voting, activism, and community programming.

While our current results hold for the linear functional forms described in Sec. II and have not proven to persist with different functional forms, we suspect that these phenomena would arise in alternative forms of the model based on the analytical results in the two-individual case in Sec. III C. Given that Lemma 1 yields $\tanh(H(t) + C)$ where $H(t) + C$ is an anti-derivative of $h(\gamma(t)) = [f_A(\gamma(t)) + f_S(\gamma(t))]f_C(\gamma(t))$, similar results could be derived as long as h is increasing in γ . The fact that a single pairwise interaction is robust to changes in the functional forms suggests, but does not prove, that the system may be robust to such changes as well.

This analysis was done specifically in the one-action model, but in the simple linear case, the perceived behavioral control being an average over all available actions means that a similar strategy could be used to investigate the multi-action system. In the two-individual case, again, the size of the increase in behavioral intention could be bounded below and

bounded above, even when nudges may come from the other individual taking multiple different actions at different rates. Although the increased dimensionality of the multi-action system gives room for far more complicated patterns, in the linear two-individual example, we can expect that the single-action results give us some information about the multi-action case. The difficulty in doing the same analysis in this case comes from the high dimensionality of the state space, requiring a more complicated partition.

While our model addresses human social behavior, it exhibits dynamics that resemble those of neuronal networks. In our model, individuals act when they are nudged by the actions of their neighbors; similarly, in an integrate-and-fire neuronal model, neurons receive nudges from other neurons and activate (i.e., fire) when they reach a threshold [25]. A network of neurons also exhibits patterns of recruitment and oscillations similar to our behavioral model [26]. Moreover, neuronal networks can display temporal biases during cognitive tasks—mimicking our model’s immediacy effect—driven by plasticity rules that change connectivity strength between neurons [23,27,28]. These parallels between our model and models of neural networks suggest several directions for future work. For example, while we have assumed that an individual’s action always nudges others toward action-taking, nudges could instead increase or decrease the recipients’ motivation to act, mimicking the excitatory and inhibitory properties of neurons. Likewise, while we have assumed that an individual can take actions continuously, a neuron cannot fire for a short period following its activation. Analogous constraints may be relevant to human behavior: for example, an individual’s perceived behavioral control could decline temporarily after taking an action. Finally, neuronal networks display additional temporal biases, such as primacy biases in which initial inputs are weighed more heavily. In the context of neuronal networks, this can result in dynamic changes in connectivity between neurons that may be relevant to human behavior [28,29].

We have shown that immediacy bias alone can give rise to complex transient dynamics (Fig. 10) and frequency modulations (Fig. 11). We anticipate that extending our model to include other types of observational biases, such as primacy bias, would result in relevant changes to the collective dynamics that mimic human behavior. Empirical studies in psychology and sociology have shown that behavioral context and experience can shape the relative propensities toward different types of biases [30–32]: for example, primacy effects tend to drive investment behaviors [33], whereas immediacy effects tend to drive food consumption choices [34]. The relative impacts of these biases can depend on the form and strength of social connections [35]. Moreover, the relative importance of different biases may vary over time, even within a narrow question or domain (e.g., investments or food consumption): in the context of consumer choices, for instance, primacy bias tends to be strong when all available options are undesirable but weakens as the options become more attractive overall [36]. Future work could explore how the emergence and dynamics of behavioral cascades in our model are impacted by such context-dependent, temporally varying cognitive biases, as well as interactions between these biases.

Several assumptions in our model merit discussion. First, we have assumed that every individual can observe the actions of all others in the population. While this assumption may be appropriate for small, close-knit communities, it is unlikely to hold in larger, structured societies, where individuals can only sample the actions of their neighbors from time to time. Future work could explore between-individual or temporal variation in the observability, uncertainty, or interpretations of observed actions, perhaps in heterogeneous networks. Second, while we have adopted a deterministic approach, social phenomena are often inherently stochastic. Incorporating small amounts of noise would be an important direction for future work, as it would both increase the realism of the model and reveal how stochastic fluctuations affect the robustness of the collective outcomes identified here. Third, we have chosen to model elements of the theory of planned behavior that are thought to directly shape behavioral intentions: internal attitudes, perceived social norms, and perceived behavioral control. However, other, secondary elements of the theory—such as normative beliefs and motivation to comply, which are thought to shape the perceived social norm [14]—may also be relevant depending on the behavior of interest and domain of application. Likewise, our work has focused on developing a theoretical model that links individual decision-making to collective outcomes, and this model could be used to generate predictions about human behavior in specific instances. Applying this model to empirical cases and behavioral data is a natural direction for future work.

Understanding human collective behavior requires an account of the interplay between internal motivations and external actions mediated by social influence. Our work provides such an account—a theoretical model for understanding the social and psychological regimes in which actions propagate within a population—laying the groundwork for future empirical and theoretical studies of collective behavior grounded in psychological theory.

ACKNOWLEDGMENTS

We thank The Banff International Research Station for Mathematical Innovation and Discovery (BIRS) for supporting the Workshop on Building Networks: Women in Complex and Nonlinear Systems, where our collaboration formed and the ideas we have developed were initially proposed. We thank Irina Popovici, Namrata Shukla, Rebecca Hardenbrook, Xie He, Peter Mucha, and Anna Vasenina for fruitful discussions. A.C.S. was supported by funding to Peter J. Mucha’s research group at Dartmouth College, including the Army Research Office MURI Award No. W911NF-18-1-0244, and startup funds provided by Dartmouth College. M.K. was supported by a Postdoctoral Fellowship Award from the James S. McDonnell Foundation (doi:10.37717/2021-3209). T.L.E. is supported by NIH BRAIN Initiative R00NS127855. R.D., J.S.M., and N.H.F. were supported by NSF Grants No. DBI 2412115 and No. DBI 2622265 as part of the U.S. NSF Center for Analysis and Prediction of Pandemic Expansion (APPEX).

DATA AVAILABILITY

Code for this paper is openly available [37].

APPENDIX: PROOFS RELATED TO THE CLASSIFICATION OF THE TWO-INDIVIDUAL GAME

Proposition A1. Suppose $\gamma_i(t) = 0$ and $x_i(0) = u_0$. Then x_i has the explicit solution $x_i(t) = \tanh(A_i t + \tanh^{-1}(u_0))$, where $A_i = [\sigma_A \alpha_i - \sigma_S \mu_S] \sigma_C \mu_C$.

Proof. While no individual has acted, the system is uncoupled. Thus, every individual acts autonomously. As such, and for convenience, we drop the index i on x_i , α_i , and A_i . With the given conditions

$$\frac{dx}{dt} = A(1-x)(1+x) = A(1-x^2),$$

which (for the purpose of solving the separable ODE) gives us

$$\frac{dx}{1-x^2} = A dt.$$

Through elementary ODE techniques, we get that

$$\frac{1}{2} \ln \left(\left| \frac{1+x}{1-x} \right| \right) = At + C,$$

and we are able to drop the absolute value bars due to the constraints on x . Using the initial condition $x(0) = u_0$,

$$\frac{1}{2} \ln \left(\frac{1+u_0}{1-u_0} \right) = \tanh^{-1}(u_0) = C,$$

$$t = \frac{1}{A} (\tanh^{-1}(x) - \tanh^{-1}(u_0)).$$

Solving for x , we get $x = \tanh(At + \tanh^{-1}(u_0))$, the desired result. ■

Corollary A1. Suppose that for individual i , $x_i(0) = u_0 < \tau$, $\gamma_i(t) = 0$ for all $t \leq \frac{1}{A} (\tanh^{-1}(\tau) - \tanh^{-1}(u_0))$ and $\sigma_A \alpha_i - \sigma_S \mu_S > 0$. Then, the time until individual i 's first action is $T = \frac{1}{A} (\tanh^{-1}(\tau) - \tanh^{-1}(u_0))$.

Proof. Until x_i reaches the threshold τ , it is C^∞ , so we are able to evaluate the expression derived in Proposition A1 with $x_i = \tau$ to get

$$T = \frac{1}{A} (\tanh^{-1}(\tau) - \tanh^{-1}(u_0))$$

as desired. ■

Lemma A1. Suppose $x_i(t_0) = x_j(t_0)$ and $\dot{x}_i(t_0) > \dot{x}_j(t_0)$. Then there is a $\delta > 0$ and an interval where $t \in (t_0, t_0 + \delta)$ satisfies $x_i(t) > x_j(t)$.

Proof. First, if $x_i(t_0) \geq \tau$, then $x_i(t_0 + dt) = 0$; otherwise, $x_i(t_0 + dt) < \tau$. Note that by definition of the derivative,

$$x(t_0 + dt) = \dot{x}(t_0)dt + x(t_0) + o(dt) \text{ as } dt \rightarrow 0.$$

Express $o(dt)/dt = \varepsilon(dt)$, so

$$o(dt) = \varepsilon(dt)dt \text{ as } dt \rightarrow 0.$$

Hence,

$$x_i(t_0 + dt) - x_j(t_0 + dt) = (\dot{x}_i(t_0) - \dot{x}_j(t_0))dt + \varepsilon(dt)dt$$

as $dt \rightarrow 0$. Let $c = \dot{x}_i(0) - \dot{x}_j(0) > 0$. Since $\varepsilon(dt) \rightarrow 0$, there exists $\delta > 0$, where

$$|dt| < \delta \Rightarrow |\varepsilon(dt)| < c/2.$$

Then, for $dt \in (0, \delta)$,

$$\begin{aligned} x_i(t_0 + dt) - x_j(t_0 + dt) &= cdt + \varepsilon(dt)dt \\ &\geq cdt - |\varepsilon(dt)|dt \\ &> \frac{c}{2}dt > 0. \end{aligned}$$

That is, for $t \in (t_0, t_0 + \delta)$, it must be that $x_i(t) > x_j(t)$. ■

Corollary A2. Suppose that $x_i(0) = 0$ and $\sigma_A \alpha_i - \sigma_S \mu_S > 0$. Then, the time until individual i 's first action is less than or equal to $T = \frac{1}{A} \tanh^{-1}(\tau)$.

Proof. Since γ_i is always nonnegative,

$$\begin{aligned} \dot{x}_i &= [\sigma_A \alpha_i + \sigma_S (\gamma_i - \mu_S)] \sigma_C (\gamma_i + \mu_C) (1 - x_i) (1 + x_i) \\ &\geq A_i (1 - x_i) (1 + x_i) \end{aligned}$$

for all t . Define the system with only one individual given by $\dot{x}^* := A_i (1 - x_i) (1 + x_i)$ and $x^*(0) = 0$. Suppose, for contradiction, that individual i acts after time T . Then x_i, x^* are both continuous until T . Applying Lemma A1, we see that for $t \in (0, \delta)$, it must be that $x_i(t) > x^*(t)$. By supposition x^* acts before x_i , so the intermediate value theorem tells us there is some (first) time $\delta < t' < T$ where $x^*(t') = x_i(t')$, and applying Lemma A1 again gives us the desired contradiction since x^* could not possibly act before x_i if $x^* \leq x_i$. ■

Corollary A3. Suppose that for individual i , $\sigma_A \alpha_i - \sigma_S \mu_S > 0$. Then, individual i will act countably infinitely many times after any fixed time $t_{\text{start}} \geq 0$.

Proof. t is unbounded above. We know from Corollary A2 that individual i will act at time $t^* \leq T$, and $x_i(t^* + dt) = 0$. We can then apply Corollary A2 at $t^* + dt$ again to see that x_i will act again. We can apply this argument every time that individual i acts, so it must act infinitely many times. In the rest of this proof, we show that an individual cannot act an uncountable number of times. To do so, we show that the interval between actions is bounded from below. Consider

$$\dot{x}_i = [\sigma_A \alpha_i + \sigma_S (\gamma_i - \mu_S)] \sigma_C (\gamma_i + \mu_C) (1 - x_i) (1 + x_i).$$

The presence of nonzero γ_i only results in a larger derivative, and $\max \gamma_i = 1$. Additionally $(1 - x_i)(1 + x_i)$ is maximized at 0, where it is 1. Then,

$$\dot{x}_i \leq [\sigma_A \alpha_i + \sigma_S - \sigma_S \mu_S] \sigma_C (1 + \mu_C).$$

Regardless of $x_i(0)$, the first time this individual acts at time t , we will have $x_i(t + dt) = 0$. Then the variable interval between actions T_{var} is bounded as follows:

$$T_{\text{var}} \geq \frac{\tau}{[\sigma_A \alpha_i + \sigma_S - \sigma_S \mu_S] \sigma_C (1 + \mu_C)},$$

and the right-hand side is constant. ■

Lemma A2. Suppose we have $x_j(0) = y_j(0) = 0$ for all individuals, j . Suppose the individual i with the smallest α_i acts at time t . Then all other individuals must also act. Furthermore, all other individuals could not have acted later than individual i .

Proof. We start by considering $N = 2$. If $\alpha_1 = \alpha_2$, then the dynamics for each individual are identical and the claim is trivial. Then, without loss of generality, let $\alpha_1 < \alpha_2$, and for all $t \geq 0$, $x_2(t) < \tau$. Let $x_1(0) = x_2(0) = y_1(0) = y_2(0) = 0$. Suppose for contradiction that there exist time(s) $t^* \geq 0$

where $x_1(t^*) = \tau$, and for simplicity, say that t_0 is the smallest such time. Note first that with the initial conditions above,

$$\dot{x}_i(0) = [\sigma_A \alpha_i - \sigma_S \mu_S] \sigma_C \mu_C.$$

This implies that $\dot{x}_1(0) < \dot{x}_2(0)$. Applying Lemma A1, we see that all $t \in (0, \delta)$ satisfy $x_2(t) > x_1(t)$. When neither individual has acted, the system is continuous. Since $x_1(t) < \tau$ for all $0 \leq t \leq t_0$, and $x_2(t) < \tau$ for all $t \geq 0$, both by assumption, there exists at least one time t where $x_1(t) = x_2(t)$ by the intermediate value theorem. Call $t' \leq t_0$ the smallest such t . Note that $y_i(t) = 0$ for $t < t_0$, so $\dot{x}_2(t') > \dot{x}_1(t')$. Again, we use Lemma A1 to see that $dt \in (0, \delta')$ satisfies

$$x_2(dt + t') > x_1(dt + t'),$$

thus $x_1(t) > x_2(t)$ is impossible for any $t \leq t_0$, which is a contradiction. So we have shown that $x_1(t) = \tau$ for some t implies $x_2(p) = \tau$ for some $p \leq t$ given the conditions in the supposition. It is easy to show that this generalizes to the case with N individuals. We can apply the same argument above to each of x_i , $i \geq 3$. We can do this because for every individual who has not yet acted, γ_i is the same, and they all had $x_i(0) = 0$. ■

Proposition A2. Suppose $\gamma_i(t) = 0$ for all $t \in [0, T)$, $x_i(0) < \tau$, and $\sigma_A \alpha_i - \sigma_S \mu_S \leq 0$. Then individual i will not act in $[0, T)$.

Proof. $\dot{x}_i \leq 0$ for all t , so $x_i(t) = \tau > 0$ is impossible. ■

Corollary A4. A system falls under the “no action” classification [Fig. 2(a)] if and only if the individual i with the largest α_i satisfies $\sigma_A \alpha_i - \sigma_S \mu_S \leq 0$ and $x_j(0) < \tau$ for all individuals i .

Proof. If the system is classified in the no action state, then the contrapositive of Corollary A1 tells us that $\sigma_A \alpha_i - \sigma_S \mu_S \leq 0$ is satisfied for each individual. For the other direction, α_i satisfies the condition above. Suppose by way of contradiction that an individual i acts first at time T . Because this is the first action, $\gamma_i(t) = 0$ for $t \leq T$. The contrapositive of Proposition A2 means that $\sigma_A \alpha_i - \sigma_S \mu_S > 0$. However, if $\sigma_A \alpha_i - \sigma_S \mu_S \leq 0$ for the individual with the largest α_i , the same inequality must hold for all individuals. This contradiction completes the proof that no individual will act. ■

Corollary A5. If $\sigma_A \alpha_i - \sigma_S \mu_S > 0$ for all i , then the system is classified under the full action state [Fig. 2(c)].

Proof. This follows directly from Corollary A3. ■

Lemma A3. Assume (h1), (h3), and (h4). If $x_2(T) = u_0$ and x_1 acts at time T , then $x_2(T + t) = V(t; u_0)$ depending on the time $t \in [0, T]$ and the initial condition u_0 as long as x_2 itself does not act in this interval, where

$$V(t; u_0) = \tanh(At + Be^{-rt} + Ce^{-2rt} + D(u_0)),$$

with $A = [\sigma_A \alpha_2 - \sigma_S \mu_S] \sigma_C \mu_C$, $B = -\frac{1}{r}[(\sigma_A \alpha_2 - \sigma_S \mu_S) \sigma_C + \sigma_S \sigma_C \mu_C]$, $C = -\frac{\sigma_S \sigma_C}{2r}$, and $D(u_0) = \tanh^{-1}(u_0) - B - C$.

Proof. In the two individual system, immediately after x_1 acts, the system is

$$\begin{aligned} \dot{x}_1 &= [\sigma_A \alpha_1 + \sigma_S (y_2 - \mu_S)] [\sigma_C (y_2 + \mu_C)] (1 - x_1^2), \\ \dot{x}_2 &= [\sigma_A \alpha_2 + \sigma_S (y_1 - \mu_S)] [\sigma_C (y_1 + \mu_C)] (1 - x_2^2), \\ \dot{y}_1 &= -ry_1, \\ \dot{y}_2 &= -ry_2. \end{aligned} \quad (\text{A1})$$

The initial conditions are $x_1(0) = 0$, $x_2(0) = u_0$, $y_1(0) = 1$, $y_2(0) = \tilde{y}$, where u_0 is the position of x_2 when x_1 acts, and $\tilde{y}$ similarly. Notice that for our constraints, the system is decoupled, so we need only consider x_2 and y_1 . Furthermore, $y_1(t) = e^{-rt}$ so we have a single separable ODE. The computation follows through grouping and keeping track of the constants:

$$\dot{x}_2 = [C_6 + C_7 e^{-rt} + C_8 e^{-2rt}] (1 - x_2^2),$$

where $C_6 = A_2 = (\sigma_A \alpha_2 - \sigma_S \mu_S) (\sigma_C \mu_C)$, $C_7 = (\sigma_A \alpha_2 - \sigma_S \mu_S) \sigma_C + \sigma_S \sigma_C \mu_C$, $C_8 = \sigma_S \sigma_C$. For brevity, we write x_2 just as x .

$$\begin{aligned} \frac{1}{1 - x^2} \frac{dx}{dt} &= C_6 + C_7 e^{-rt} + C_8 e^{-2rt}, \\ \int \frac{1}{1 - x^2} \frac{dx}{dt} dt &= \int (C_6 + C_7 e^{-rt} + C_8 e^{-2rt}) dt, \\ \int \frac{1}{1 - x^2} dx &= \left[C_6 t - \frac{C_7}{r} e^{-rt} - \frac{C_8}{2r} e^{-2rt} \right] + C_0, \end{aligned}$$

where C_0 is the constant of integration. Through partial fraction decomposition on the left-hand side, we obtain the following:

$$\frac{1}{2} \ln \left(\frac{1 + x}{1 - x} \right) = \left[C_6 t - \frac{C_7}{r} e^{-rt} - \frac{C_8}{2r} e^{-2rt} \right] + C_0.$$

As x is bounded in $(-1, 1)$, we need not use the absolute values. Using this, we can solve for $C_0(u_0)$ through $x(0) = u_0$, which gives us

$$C_0(u_0) = \frac{1}{2} \ln \left(\frac{1 + u_0}{1 - u_0} \right) + \frac{C_7}{r} + \frac{C_8}{2r}.$$

We rearrange the equation again to get that

$$\frac{1 + x}{1 - x} = \exp \left(2 \left(\underbrace{C_6 t - \frac{C_7}{r} e^{-rt} - \frac{C_8}{2r} e^{-2rt} + C_0(u_0)}_{h(t, u_0)} \right) \right).$$

The left-hand side of the function is injective from $(-1, 1) \rightarrow (0, \infty)$ so we can solve for the unique solution

$$x = \frac{e^{2h(t, u_0)} - 1}{e^{2h(t, u_0)} + 1} = \tanh(h(t, u_0)).$$

This means that if x_2 is at position u_0 when x_1 acts with period T , then the position of x_2 at time $t \in [0, T]$ is given by $u(T + t) = V(t; u_0)$ defined as

$$V(t; u_0) := \tanh(At + Be^{-rt} + Ce^{-2rt} + D(u_0)),$$

where $A = A_2 = (\sigma_A \alpha_2 - \sigma_S \mu_S) (\sigma_C \mu_C)$, $B = -\frac{1}{r}[(\sigma_A \alpha_2 - \sigma_S \mu_S) \sigma_C + \sigma_S \sigma_C \mu_C]$, $C = -\frac{\sigma_S \sigma_C}{2r}$, $D(u_0) = \frac{1}{2} \ln \left(\frac{1 + u_0}{1 - u_0} \right) - B - C = \tanh^{-1}(u_0) - B - C$. ■

Lemma A4. Suppose (h1), (h2), and (h4), and suppose that individual 2 does not act on the interval $[kT, (k + 1)T]$. Then the following are true:

$$\begin{aligned} 0 < V(T; 0) &\Rightarrow u_0 < V(T; u_0), \\ 0 = V(T; 0) &\Rightarrow u_0 = V(T; u_0), \\ 0 > V(T; 0) &\Rightarrow u_0 > V(T; u_0). \end{aligned}$$

Proof. In the first case suppose that $u_0 < \tanh(AT + Be^{-rT} + Ce^{-2rT} + D(u_0))$. Since $\tanh^{-1}$ is monotonic, we have that

$$\tanh^{-1}(u_0) < AT + Be^{-rT} + Ce^{-2rT} + D(u_0).$$

Recall $D(u_0) = \tanh^{-1}(u_0) - B - C$, so we have that

$$B + C < AT + Be^{-rT} + Ce^{-2rT},$$

which is an expression entirely independent of u_0 . Each step is reversible and so this implies that

$$\begin{aligned} V(0; v_0) &< \tanh(AT + Be^{-rT} + Ce^{-2rT} + D(v_0)) \\ &< V(T; v_0) \end{aligned}$$

for any $v_0 \in (-1, 1)$. What we have shown is that, if the inequality is satisfied for any initial condition u_0 , it is satisfied for any other condition u_0 . The same thing can be done in the ($=$) and ($>$) case. ■

Lemma A5. Suppose we have that $\sigma_A \alpha_1 - \sigma_S \mu_S > 0$ while $\sigma_A \alpha_2 - \sigma_S \mu_S < 0$. Then, in the case where $\dot{x}_2(T) > 0$, we can derive a critical time t_{crit} where for each k before x_2 acts (if it does at all), $x_2(kT + t_{\text{crit}})$ achieves a local maximum on $[kT, (k+1)T]$ given by

$$t_{\text{crit}} = -\frac{1}{r} \ln \left(\left| \mu_S - \frac{\sigma_A \alpha_2}{\sigma_S} \right| \right).$$

Proof. Notice that in the ODE for the position of individual i ,

$$\dot{x}_i = [\sigma_A \alpha_i + \sigma_S (\gamma_i - \mu_S)] \sigma_C (\gamma_i + \mu_C) (1 - x_i) (1 + x_i),$$

the only term affecting the sign of $\dot{x}_i$ is

$$\sigma_A \alpha_i + \sigma_S (\gamma_i - \mu_S).$$

Since we are in the case with only two individuals, $\gamma_2(t) = e^{-rt}$ when $t < T$. We can look for the critical point of $x_2(t)$ in this interval:

$$0 = \sigma_A \alpha_2 + \sigma_S e^{-rt} - \sigma_S \mu_S,$$

and we find the solution

$$-\frac{1}{r} \ln \left(\left| \mu_S - \frac{\sigma_A \alpha_2}{\sigma_S} \right| \right) = t_{\text{crit}}.$$

Note that if $t_{\text{crit}} < T$ and x_2 has not yet acted, then, there is a critical time $t_{\text{crit}} + kT \in [kT, (k+1)T]$. ■

Corollary A6. Suppose (h_1) – (h_4) hold and that $M = 0$. Then, the system is in the partial action state. That is, individual 1 acts but individual 2 never acts.

We seek to show that x_2 has a local maximum below the threshold. We suppose for contradiction that this local maximum is above the threshold. Then, we examine x_2 as if this individual ignores the threshold, with the intent of applying Lemma A4 to reach our contradiction. Below is a sketch of how we achieve the contradiction. x_2 was only decreasing on $[0, T]$, and on every future period, x_2 is increasing during the start of $[kT, (k+1)T]$. x_2 must return to $x_2(T)$ at the beginning of each period. Putting these facts together, we realize that if x_2 were able to cross $x_2(0)$, $|\dot{x}_2|$ is too small to return to $x_2(T)$ by the beginning of the next period. This tells us that we can pick any starting value lower than the threshold and land in a unique oscillatory equilibrium. That is,

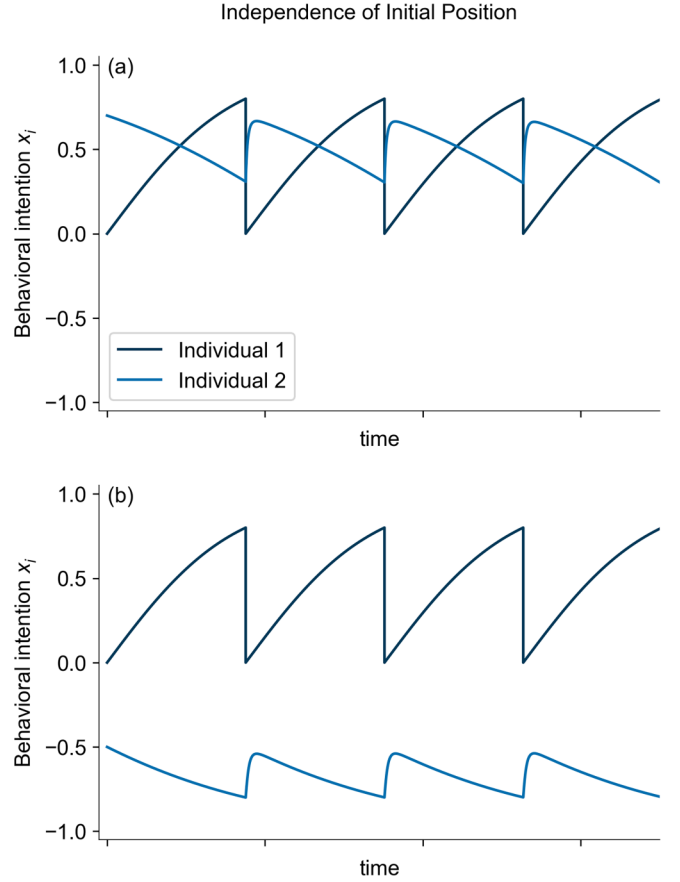


FIG. 12. Both are examples where the net change on each period is 0. In panel (a), we can see an example where initial position is high, yet individual 2 cannot act. This behavior is qualitatively no different than the case in panel (b) where individual 2's initial position is quite low.

x_2 will oscillate between $x_2(T)$ and $x_2(T + t_{\text{crit}})$ with exactly the same behavior over each period (Fig. 12).

Proof. If x_2 does not act on $[T, 2T]$, then it will never act, as the same behavior will be exactly replicated on $[kT, (k+1)T]$ for all $k \geq 1$. That is, there will be no net change over each interval. Suppose for contradiction that there is some point t' on $[T, 2T]$ where $x_2(t') \geq \tau$, then $x_2(t') > x_2(0)$ by supposition. For the moment, say that x_2 ignores the threshold and examine what x'_2 's behavior would be—it must be that $x_2(2T) = x_2(T)$ by Lemma A4. We show this is not possible. To see why, note that for $t \in (0, T)$, $\dot{x}_2(t) < 0$ was always true. We can make the following comparison as the action of x_1 at T can only increase the derivative of x_2 :

$$\dot{x}_2(T + t) > \dot{x}_2(t), \quad t \in (0, T).$$

Since $x_2(T) = x_2(2T)$, but for any small $\varepsilon > 0$, $x'_2(T + \varepsilon) > 0$ for $x_2(t') \geq \tau$, then x_2 must, at some point, be decreasing on the period $(T, 2T)$, so by the intermediate value theorem, there is a point where the derivative is 0. We have found an exact expression for this in Lemma A5. Call that point $T + t_{\text{crit}}$, and $x_2(T + t_{\text{crit}}) \geq x_2(t') \geq \tau > x_2(t_{\text{crit}})$. Then,

$$x_2(T + t_{\text{crit}}) - x_2(2T) > x_2(t_{\text{crit}}) - x_2(T)$$

since $x_2(T) = x_2(2T)$. However, $\dot{x}_2(T + t) > \dot{x}_2(t)$ for $t \in (0, T)$, so it cannot be true that $x_2(2T) = x_2(T)$. Restated: the

derivative is *larger* in sign on $(T, 2T)$ than on $(0, T)$, and the position of x_2 at $T + t_{\text{crit}}$ is *larger* than the position of x_2 at t_{crit} , but the remaining time $2T - t_{\text{crit}} - T = T - t_{\text{crit}}$ to reach the position $x_2(T) = x_2(2T)$ is equal. ■

Lemma A6. Suppose that $\alpha_i = \alpha_j$, and neither x_i, x_j acts on some interval $[t_0, t_1]$. Suppose also that $x_i(t_0) \geq x_j(t_0)$, then $x_i \geq x_j$ on $[t_0, t_1]$.

Proof. In the equality case, this follows from the previous result, so suppose that the inequality is strict. Suppose for contradiction that there is some point where $x_i(t) < x_j(t)$, then by the intermediate value theorem, there is some point t' such that $x_i(t') = x_j(t')$. Since $\alpha_i = \alpha_j$, $\dot{x}_i(t) = \dot{x}_j(t)$ on $[t', t_1]$, so $x_j(t) = x_i(t)$. ■

Theorem A1. Suppose (h_1) – (h_3) hold, and suppose $x_2(0) < \tau$. If $M \leq 0$, then the system is classified under partial action (Fig. 6).

Proof. We covered the $M = 0$ case in Corollary A6. This proof will be very similar. Suppose we have $M < 0$. As in the last proof, if x_2 does not act on $[T, 2T]$, it will never act. This is because x_2 was only decreasing on $[0, T]$, and there is a net decrease over each following period, so we can apply Lemma A6. If $\dot{x}_2(T + \varepsilon) \leq 0$ for every small $\varepsilon > 0$, then we are done as x_2 will never be increasing, so suppose for contradiction that x_2 acts on $[T, 2T]$. As a consequence, x_2 must be increasing on (T, T') for some $T < T' < 2T$. We show that this contradicts $M < 0$. To see this, examine the behavior of x_2 as if it ignores the threshold. $x_2(t)$ must be decreasing over some interval $[T', 2T]$, or else we would not have the net decrease prescribed by $M < 0$. Then, the intermediate value theorem tells us that there is a point where $\dot{x}_2 = 0$, and we have derived in Lemma A5 that this point is $T + t_{\text{crit}}$. Then, $x_2(T + t_{\text{crit}}) \geq x_2(T') \geq \tau > x_2(t_{\text{crit}})$. Then,

$$x_2(T + t_{\text{crit}}) - x_2(2T) > x_2(t_{\text{crit}}) - x_2(T),$$

since $x_2(T) > x_2(2T)$ by supposition. However, $\dot{x}_2(T + t) > \dot{x}_2(t)$ for $t \in (0, T)$, so it cannot be true that $x_2(2T) < x_2(T)$, as reasoned similarly in Corollary A6. ■

Lemma A7. Let w_n be a sequence of real numbers in $(-1, 1)$ satisfying $w_n = f(w_{n-1})$. If $M > 0$ then $w_n \rightarrow 1$, if $M < 0$ then $w_n \rightarrow -1$, and if $M = 0$ then $w_n = w_0$.

Proof. The result is accomplished by elementary theory of difference equations. Let $g(x) = f(x) - x = \tanh(M + \tanh^{-1}(x)) - x$. Notice that $\lim_{x \rightarrow -1+} g(x) = \lim_{x \rightarrow 1-} g(x) = 0$ and that $g(x)$ is continuous on the interval $(-1, 1)$. Let $w \in (-1, 1)$ and observe

$$\tanh(M + \tanh^{-1}(x)) = x + g(x),$$

$$M + \tanh^{-1}(x) = \tanh^{-1}(x + g(x)).$$

Because $\tanh^{-1}$ is monotonic increasing, we know that if $M > 0$, then $g(x) > 0$ and if $M < 0$, then $g(x) < 0$ for all $x \in (-1, 1)$. This means that a sequence satisfying $w_n = f(w_{n-1})$ is monotonic increasing if $M > 0$ and monotonic decreasing if $M < 0$. The sequence is stationary if $M = 0$. Moreover, $g(x)$ has no internal zeros. In the case that $M > 0$, the sequence is monotonic increasing and bounded, so it converges to a limit. If that limit $w_n \rightarrow \hat{w} \in (-1, 1)$, then $g(\hat{w}) = 0$, which is not true, so that limit must converge to 1. Likewise, if $M < 0$ the limit must converge to -1 . ■

Corollary A7. If $M > 0$ and $\tau < 1$, then there is a time at which $x_2(t) = \tau$.

Proof. Observe that, until x_2 reaches the threshold τ , it satisfies the relation $x_2(kT) = f(x_2((k-1)T))$. If $M > 0$, then there is a K for which $w_k > \tau$ if it satisfies $w_n = f(w_{n-1})$ by Lemma 3. This means at some time $t \leq KT$, $x_2(t) > \tau$. ■

Theorem A2. Given (h_1) – (h_3) and $M > 0$, then x_2 will act infinitely many times, thus establishing the full action state (Fig. 7).

Proof. Suppose that x_2 has just acted at time t^* . We show it must act again. Since $M > 0$, we know that if x_2 receives a nudge from x_1 at some time t' , then $x_2(t') < x_2(t)$ for any t in $(t', t' + T)$. In other words, $M > 0 \Rightarrow V(t, u_0) > u_0$ for all $t \in [0, T]$. Since x_2 acted, the times between x_1 's actions are all less than or equal to T by Corollary A2. Call this varying time T_{var} . From Corollary A3, we have a lower bound on T_{var} , which we will name T' . This means $T_{\text{var}} \in [T', T]$. Let $\{T_i\}$ be the sequence of times at which x_1 acts after x_2 has acted for the first time. We can say $T_i - T_{i-1} \in [T', T]$. Observe that, because $V(t, u_0)$ is continuous and concave down for every u_0 on the interval $[0, T]$, we know that after any time $T_i - T_{i-1} \in [T', T]$,

$$x_2(T_i) \geq \min\{V(T', x_2(T_{i-1})), V(T, x_2(T_{i-1}))\}.$$

This recurrence inequality tells us the same thing that the original recurrence relation of Lemma A7 tells us. If $M > 0$, then $\min\{V(T', w), V(T, w)\} - w > 0$. This means that any sequence satisfying $w_k = \min\{V(T', w_{k-1}), V(T, w_{k-1})\}$ is a monotonic increasing sequence which is bounded in $(-1, 1)$. It must have a limit and, because $\min\{V(T', w) - w, V(T, w) - w\} > 0$ for all $w \in (-1, 1)$, the only possible limit for such a sequence is $w_k \rightarrow 1$. We have shown that $x_2(T_i)$ is an increasing sequence whose limit is 1 and thus there must be a time t such that $x_2(t) > \tau$ again for any $\tau < 1$. We can repeat this process indefinitely to show that x_2 will act infinitely many times and thus we attain the full action state. ■

- [1] R. M. Bond, J. E. Settle, C. J. Fariss, J. J. Jones, and J. H. Fowler, Social endorsement cues and political participation, *Political Commun.* **34**, 261 (2017).
- [2] W. Jennings, G. Stoker, H. Bunting, V. O. Valgarðsson, J. Gaskell, D. Devine, L. McKay, and M. C. Mills, Lack of trust, conspiracy beliefs, and social media use

predict COVID-19 vaccine hesitancy, *Vaccines* **9**, 593 (2021).

- [3] R. Cont and J.-P. Bouchaud, Herd behavior and aggregate fluctuations in financial markets, *Macroecon. Dyn.* **4**, 170 (2000).
- [4] P. Rollin and S. Bamberg, It's all up to my fellow citizens. descriptive norms as a decisive mediator in the relationship

- between infrastructure and mobility behavior, *Front. Psychol.* **11**, 610343 (2021).
- [5] B. Beckage, L. J. Gross, K. Lacasse, E. Carr, S. S. Metcalf, J. M. Winter, P. D. Howe, N. Fefferman, T. Franck, A. Zia, *et al.*, Linking models of human behaviour and climate alters projected climate change, *Nat. Clim. Change* **8**, 79 (2018).
 - [6] M. J. Silk, S. Carrignon, R. A. Bentley, and N. H. Fefferman, Observations and conversations: How communities learn about infection risk can impact the success of non-pharmaceutical interventions against epidemics, *BMC Public Health* **22**, 13 (2022).
 - [7] Y. Katz, K. Tunström, C. C. Ioannou, C. Huepe, and I. D. Couzin, Inferring the structure and dynamics of interactions in schooling fish, *Proc. Natl. Acad. Sci. USA* **108**, 18720 (2011).
 - [8] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, Statistical mechanics for natural flocks of birds, *Proc. Natl. Acad. Sci. USA* **109**, 4786 (2012).
 - [9] K. Tunström, Y. Katz, C. C. Ioannou, C. Huepe, M. J. Lutz, and I. D. Couzin, Collective states, multistability and transitional behavior in schooling fish, *PLoS Comput. Biol.* **9**, e1002915 (2013).
 - [10] X. Wang, L. Zhang, Y. Lin, Y. Zhao, and X. Hu, Computational models and optimal control strategies for emotion contagion in the human population in emergencies, *Knowledge-Based Syst.* **109**, 35 (2016).
 - [11] I. Iacopini, G. Petri, A. Barrat, and V. Latora, Simplicial models of social contagion, *Nat. Commun.* **10**, 2485 (2019).
 - [12] G. Schwenk and T. Reimer, Simple heuristics in complex networks: Models of social influence, *J. Artif. Soc. Social Simul.* **11**, 4 (2008).
 - [13] P. Wang, L.-J. Zhang, X.-J. Xu, and G. Xiao, Heuristic strategies for persuader selection in contagions on complex networks, *PLoS One* **12**, e0169771 (2017).
 - [14] I. Ajzen, The theory of planned behavior, *Org. Behav. Hum. Decis. Process.* **50**, 179 (1991).
 - [15] I. Ajzen, From intentions to actions: A theory of planned behavior, in *Action Control: From Cognition to Behavior* (Springer, Berlin, 1985), pp. 11–39.
 - [16] D. Feinstein and E. Patel, Max-plus opinion dynamics with temporal confidence, *arXiv:2101.10805*.
 - [17] J. Lorenz, Continuous opinion dynamics under bounded confidence: A survey, *Int. J. Mod. Phys. C* **18**, 1819 (2007).
 - [18] M. A. Porter and J. P. Gleeson, Dynamical systems on networks: A tutorial, *Front. Appl. Dyn. Syst.: Rev. Tutorials* **4** (2016).
 - [19] W. Chu, Q. Li, and M. A. Porter, Inference of interaction kernels in mean-field models of opinion dynamics, *SIAM J. Appl. Math.* **84**, 1096 (2024).
 - [20] C. Koertje and H. Sayama, Collective group drift in a partial-differential-equation-based opinion dynamics model with biased perception kernels, *Phys. Rev. E* **109**, 034304 (2024).
 - [21] H. Sayama and J. Yamanoi, Beyond social fragmentation: Coexistence of cultural diversity and structural connectivity is possible with social constituent diversity, in *Proceedings of the 6th International Winter School and Conference on Network Science (NetSci-X'20)* (Springer, Berlin, 2020), pp. 171–181.
 - [22] S. Bullock and H. Sayama, Agent heterogeneity mediates extremism in an adaptive social network model, in *Proceedings of the 2023 Artificial Life Conference: Ghost in the Machine (ALIFE'23)* (MIT Press, Cambridge, MA, 2023).
 - [23] B. B. Murdock, Jr., The serial position effect of free recall, *J. Exp. Psychol.* **64**, 482 (1962).
 - [24] P. E. Tetlock, Accountability and the perseverance of first impressions, *Soc. Psych. Quarterly* **46**, 285 (1983).
 - [25] L. F. Abbott, Lapicque's introduction of the integrate-and-fire model neuron (1907), *Brain Res. Bull.* **50**, 303 (1999).
 - [26] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, *J. Comput. Neurosci.* **8**, 183 (2000).
 - [27] B. Ballintyn, B. Shlaer, and P. Miller, Spatiotemporal discrimination in attractor networks with short-term synaptic plasticity, *J. Comput. Neurosci.* **46**, 279 (2019).
 - [28] A. J. Greene, C. Prepscius, and W. B. Levy, Primacy versus recency in a quantitative model: Activity is the critical distinction, *Learn. Memory* **7**, 48 (2000).
 - [29] F. Fiebig and A. Lansner, A spiking working memory model based on Hebbian short-term potentiation, *J. Neurosci.* **37**, 83 (2017).
 - [30] W. D. Crano, Primacy versus recency in retention of information and opinion change, *J. Soc. Psych.* **101**, 87 (1977).
 - [31] N. Miller and D. T. Campbell, Recency and primacy in persuasion as a function of the timing of speeches and measurements., *J. Abnormal Soc. Psych.* **59**, 1 (1959).
 - [32] J. H. Han, G. Davies, and A. Grimes, Recency effects in the buffering of negative news by corporate social responsibility advertising, *CCIJ* **26**, 382 (2021).
 - [33] O. Arian, A. E. Gozluclu, G. H. Kim, and H. Sakaguchi, Primacy in stock market participation: The effect of initial returns on market re-entry decisions, *Eur. J. Finance* **25**, 883 (2019).
 - [34] E. N. Garbinsky, C. K. Morewedge, and B. Shiv, Interference of the end: Why recency bias in memory determines when a food is consumed again, *Psychol. Sci.* **25**, 1466 (2014).
 - [35] R. P. Bhargava and N. V. Montgomery, My recency, our primacy: How social connection influences evaluations of sequences, *J. Behav. Decision Making* **28**, 382 (2015).
 - [36] Y. Li and N. Epley, When the best appears to be saved for last: Serial position effects on choice, *J. Behav. Decision Making* **22**, 378 (2009).
 - [37] R. Dadlani, J. S. McAlister, A. Schwarze, M. Kawakatsu, S. Iams, T. Eissa, and N. H. Fefferman, Temporally-biased planned behavior [Computer software], Zenodo, 2026, <https://doi.org/10.5281/zenodo.22713604>.