

# Multi-topic belief formation through bifurcations over signed social networks

Anastasia Bizyaeva, *Member, IEEE*, Alessio Franci, and Naomi E. Leonard, *Fellow, IEEE*

**Abstract**— We propose and analyze a nonlinear dynamic model of continuous-time multi-dimensional belief formation over signed social networks. Our model accounts for the effects of a structured belief system, self-appraisal, internal biases, and various sources of cognitive dissonance posited by recent theories in social psychology. We prove that strong beliefs emerge on the network as a consequence of a bifurcation. We analyze how the balance of social network effects in the model controls the nature of the bifurcation and, therefore, the belief-forming limit-set solutions. Our analysis provides constructive conditions on how multi-stable network belief equilibria and belief oscillations emerging at a belief-forming bifurcation depend on the communication network graph and belief system network graph. Our model and analysis provide new theoretical insights on the dynamics of social systems and a new principled framework for designing decentralized decision-making on engineered networks in the presence of structured relationships among alternatives.

**Index Terms**— Multi-agent systems, networked control systems, nonlinear systems, belief dynamics, belief system, opinion dynamics, bifurcation, signed networks, collective decision-making, matrix-valued interactions

## I. INTRODUCTION

The study of belief dynamics has a long history of academic inquiry across disciplines. In the social sciences, belief formation models are built to explore belief change in individuals as well as collective phenomena including the spread of social norms, the outcomes of elections, and the emergence of consensus or polarization in societies. In the study of collective animal behavior, similar models are deployed to understand the outcomes of social decisions in animal groups. Belief formation models are also prominent in control systems and robotics, where they inspire the design of new protocols for networked consensus formation, multi-agent decision-making, navigation, task scheduling, control of networks, and human-robot interaction [1]–[7].

Recent work in social psychology points out an important limitation of belief formation models [8], [9]. Existing models either focus on the internal dynamics of beliefs of individuals

without accounting for social network effects or they focus on the dynamics of social networks without incorporating meaningful internal factors, such as the nature of interpersonal relationships and the effects of cognitive dissonance. Emerging evidence suggests instead that belief formation should be viewed through a framework that synthesizes internal and external effects, with individual beliefs that are “embedded in a multidimensional, self-sustaining system of mental representations and shaped and reinforced continuously in the social interactions people have in their communities” [10].

A similar shift in perspective is happening in the collective animal behavior literature. Traditionally, studies have focused on the role of external interaction networks in shaping the behavior of the group [11], [12]. In contrast, recent studies present evidence that internal cognitive processes play a critical role in shaping the outcomes of animal collective decisions, e.g., during spatial movement [13], [14]. These studies suggest that realistic models of collective animal behavior should account for the internal dynamics of cognitive representations in addition to the external social and environmental factors.

In this paper we take a step forward towards a mathematical understanding of the interplay between internal and external factors in shaping collective belief formation. To do this we present and analyze a nonlinear dynamic model of collective belief formation on multiple alternatives that is adapted from the opinion dynamics model in [15],[16]. To account for external interactions our model includes a social network on which agents can have cooperative and antagonistic relationships. To account for internal factors in individual cognition our model incorporates internal biases and networked relationships between internal belief representations. The coupling between social relationships and a structured belief system is the source of various forms of cognitive dissonance in our model.

Formal models of belief formation outside of the statistical physics literature are predominantly linear. Classically, belief formation is modeled as a linear discrete-time DeGroot consensus process in which cooperative agents update scalar beliefs via local averaging [17]. Common variants of consensus dynamics include its continuous-time extension and the inclusion of time delays, time-varying communication networks, and antagonistic relationships between agents [1], [4]. To account for multiple simultaneously evolving beliefs, a number of multidimensional consensus models have recently been explored in discrete-time [18]–[21] and continuous-time [22]–[26]. Unlike their scalar counterparts, these models explicitly consider the simultaneous evolution of multiple beliefs within the agents, and thereby account for some internal

This work was supported in part by ONR grant N00014-19-1-2556, ARO grant W911NF-18-1-0325, and by NSF Graduate Research Fellowship DGE-2039656.

A. Bizyaeva is with the NSF AI Institute in Dynamic Systems and the Department of Mechanical Engineering at the University of Washington, Seattle, WA, 98195 USA; e-mail: anabiz@uw.edu. A. Franci is with the Department of Electrical Engineering and Computer Science at the University of Liège, Belgium; email: afranci@uliege.be. N.E. Leonard is with the Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ, 08544 USA; e-mail: naomi@princeton.edu.

interdependence of belief representations. However, existing multi-belief models are still linear, as they remain grounded in the local averaging assumption of DeGroot. A number of nonlinear models of belief or opinion dynamics have also been considered in the literature, e.g. [27]–[32]. However, these models only account for scalar beliefs and therefore do not capture rich internal systems of belief representations.

There is compelling evidence that realistic belief formation dynamics are nonlinear, which motivates our use of a multidimensional nonlinear model. At the individual level, nonlinear processing of perceptual evidence in the brain plays an important role in the formation of beliefs and in decision-making [33], [34]. At the group level, a recent analysis of political discussions in an online social network suggests that beliefs cluster around several isolated, simultaneously stable equilibria, a dynamical behavior that cannot be reproduced by linear models [35]. The formation of beliefs and decisions in isolated individuals and in social networks is also often linked to phase transitions in computational statistical physics models [13], [14], [36]–[41]. A deterministic analogue of a phase transition is a *bifurcation*, i.e. a parameterized change in the qualitative behavior of a dynamical system. Bifurcations are nonlinear phenomena, and indeed bifurcations will play a central role in our analysis of belief formation dynamics.

The following are the primary contributions of this work. 1) We introduce a new nonlinear multi-dimensional model of belief dynamics that accounts for structured relationships between options. We connect this model and its analysis to recent theories in social psychology. 2) We prove that agents form strong beliefs when the model exhibits a bifurcation. The emergent beliefs on the network post-bifurcation are shaped by properties of the communication graph between agents, their shared belief system, and the balance of various social effects in the model. 3) We prove sufficient conditions for the emergence of multi-stable belief equilibria and of belief oscillations, and describe the agents' beliefs in terms of spectral properties of communication and belief system graphs.

The manuscript is structured as follows. In Section II we introduce mathematical definitions and relevant background on signed graphs. In Section III we define the belief formation model and connect the model to recent work in social psychology. In Section IV we establish conditions for the onset of an indecision-breaking bifurcation in the model. In Section IV-B we analyze the effect of various model parameters on shaping and controlling solutions at the bifurcation. In Section V we derive necessary conditions for the onset of multiple equilibria in a pitchfork bifurcation, characterize the resulting steady-state solutions, and provide sufficient conditions for graph structures that support a pitchfork bifurcation. In Section VI we derive necessary conditions for the onset of belief oscillations in a Hopf bifurcation. We conclude in Section VII.

## II. BACKGROUND

### A. Notation and mathematical preliminaries

For  $x = a + ib = re^{i\phi} \in \mathbb{C}$ ,  $\bar{x} = a - ib = re^{-i\phi}$  is its complex conjugate,  $|x| = \sqrt{x\bar{x}} = r$  its modulus, and  $\arg(x)$  its argument  $\phi$ . For vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ , the inner product is

$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}\bar{\mathbf{y}}^T$ . The norm of  $\mathbf{x} \in \mathbb{C}^n$  is  $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ . For  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , the matrix  $\text{diag}\{\mathbf{x}\} \in \mathbb{R}^{n \times n}$  is a diagonal matrix with  $x_i$  in row  $i$ , column  $i$ . Define  $\mathcal{I}_n \in \mathbb{R}^{n \times n}$  as the identity matrix,  $\mathbf{0}_n \in \mathbb{R}^n$  as the zero vector, and  $\mathbf{1}_n \in \mathbb{R}^n$  as the vector of ones. We utilize two common matrix products: the Kronecker product, denoted by the symbol  $\otimes$ , and the element-wise Hadamard product, denoted by  $\odot$ . Given vectors  $\mathbf{v}, \mathbf{w}$  or matrices  $M, N$ , we say  $\mathbf{v} \succ \mathbf{w}$  if  $v_i > w_i$  for all  $i$  and  $M \succ N$  if  $M_{ij} > N_{ij}$  for all  $i, j$ .

For a square matrix  $A \in \mathbb{R}^{n \times n}$ , its spectrum is the collection of eigenvalues  $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$  where  $\lambda_i A = \lambda_i \mathbf{v}_i$  for some eigenvector  $\mathbf{v}_i \in \mathbb{R}^n$ . The spectral radius of  $A$  is  $\rho(A) = \max\{|\lambda_i|, \lambda_i \in \sigma(A)\}$ . The kernel of  $A$  is  $\mathcal{N}(A) = \{\mathbf{v} \in \mathbb{R}^n \text{ s.t. } A\mathbf{v} = \mathbf{0}\}$ . Two square matrices  $A, B \in \mathbb{R}^n$  are *co-spectral* if  $\sigma(A) = \sigma(B)$ . An eigenvalue  $\lambda \in \sigma(A)$  is a *leading eigenvalue* of  $A$  if  $\text{Re}(\lambda) \geq \text{Re}(\mu)$  for all  $\mu \in \sigma(A)$ . A leading eigenvalue  $\lambda$  is a *dominant eigenvalue* if  $\lambda = \rho(A)$ .  $A$  has the *strong Perron-Frobenius property* if its dominant eigenvalue  $\lambda$  is unique, positive,  $|\lambda_i| < \lambda$  for all  $\lambda_i \neq \lambda$  in  $\sigma(A)$ , and it has a corresponding eigenvector  $\mathbf{v} \succ \mathbf{0}$ .  $A$  is *irreducible* if it cannot be transformed into an upper triangular matrix through similarity transformations.  $A$  is *eventually positive* (*eventually nonnegative*) if there exists a positive integer  $k_0$  such that  $A^k \succ 0_{N \times N}$  ( $A^k \succeq 0_{N \times N}$ ) for all integers  $k > k_0$ . A *permutation matrix* is a square matrix  $P_n \in \{0, 1\}^{n \times n}$  formed by permuting the rows of the identity matrix  $\mathcal{I}_n$ . It has exactly one entry of 1 in each row and each column, with all other entries being zero, and it is an orthonormal matrix that satisfies  $P_n P_n^T = P_n^T P_n = \mathcal{I}_n$ .

*Proposition 2.1 (matrices with Perron-Frobenius property):* i) [42, Theorem 2.2] The following statements are equivalent for a real square  $n \times n$  matrix  $A$ : (1)  $A$  and  $A^T$  have the strong Perron-Frobenius property; (2)  $A$  is eventually positive; (3)  $A^T$  is eventually positive. ii) [43, Theorem 8.4.4] Suppose  $A$  is a real, square, irreducible matrix with nonnegative entries,  $A \succeq 0_{n \times n}$ . Then  $A$  has a unique dominant eigenvalue  $\lambda = \rho(A)$  with a corresponding eigenvector  $\mathbf{v} \succ \mathbf{0}$ .

### B. Signed graphs

The discussion of signed graphs presented here adapts the conventions from [4], [44], [45]. A *signed graph*  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s)$  consists of a set of *vertices*  $\mathcal{V} = (1, \dots, n)$ , a set of *edges*  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ , and a *signature function*  $s : \mathcal{E} \rightarrow \{-1, 1\}$ . An edge  $(i, j) \in \mathcal{E}$  begins at vertex  $i$  and ends at vertex  $j$ . The entries of the *signed adjacency matrix*  $A \in \mathbb{R}^{n \times n}$  corresponding to  $\mathcal{G}$  are  $a_{ij} = s(i, j)$  if  $(i, j) \in \mathcal{E}$  and  $a_{ij} = 0$  otherwise. All graphs in this paper are *simple*, i.e. contain no self-loops  $(i, i) \notin \mathcal{E}$  for all  $i \in \mathcal{V}$  and contain at most one edge  $(i, j)$  that begins at  $i$  and ends at  $j$  for all  $i, j \in \mathcal{V}$ . We say a signed graph is *undirected* if  $(i, j) \in \mathcal{E} \iff (j, i) \in \mathcal{E}$  and  $s(i, j) = s(j, i)$  for all  $(i, j) \in \mathcal{E}$ , and it is *directed* otherwise. For a directed signed graph, a pair of edges sharing the same vertices  $(i, j), (j, i) \in \mathcal{E}$  is a *digon*. When  $s(i, j)s(j, i) = 1$  for any digon in  $\mathcal{E}$  the graph is *digon sign-symmetric*, which means that edges between any two nodes share sign. An undirected signed graph is therefore a digon sign-symmetric graph with the property  $A = A^T$ .

Given a signed graph  $\mathcal{G}$ , its underlying unsigned graph  $|\mathcal{G}|$  with adjacency matrix  $|A|$  is obtained by setting  $s(i, j) = 1$

for all  $(i, j) \in \mathcal{E}$ . A *simple path* from  $i$  to  $j$  is a sequence of edges starting at vertex  $i$  and ending at vertex  $j$  that does not visit the same vertex twice, e.g.  $(i, l), (l, m), \dots, (k, n), (n, j)$ . The *signature of a path* is the product of the signatures of all edges along the path, e.g.  $s(i, l)s(l, m) \dots s(k, n)s(n, j)$ . A *cycle* is a simple path except that it starts and ends at the same vertex. We say  $\mathcal{G}$  is *strongly connected* if there exists a simple directed path in  $\mathcal{E}$  from any vertex to any other vertex in  $\mathcal{V}$ . Equivalently, it is strongly connected if  $|A|$  is irreducible. A signed graph  $\mathcal{G}$  is *structurally balanced* if it admits a bipartition of nodes  $\mathcal{V}_1, \mathcal{V}_2$  such that  $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}$  and  $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$ , with  $s(i, j) = 1$  for all  $(i, j) \in \mathcal{E}$  satisfying  $i, j \in \mathcal{V}_q$ ,  $q \in \{1, 2\}$ , and with  $s(i, j) = -1$  for all  $(i, j) \in \mathcal{E}$  satisfying  $i \in \mathcal{V}_q$ ,  $j \in \mathcal{V}_r$ ,  $q, r \in \{1, 2\}$ ,  $q \neq r$ .

We can define a *switching function*  $\theta : \mathcal{V} \rightarrow \{-1, 1\}$  for a given signed graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s)$ . This function partitions the vertex set  $\mathcal{V}$  into the sets  $\mathcal{V}^+$  and  $\mathcal{V}^-$  such that  $\theta(i) = 1(-1)$  for all  $i \in \mathcal{V}^+(\mathcal{V}^-)$ . We will say that all vertices in  $\mathcal{V}^-$  are in the *switching set*, or are being *switched*. The switching function defines a *switching transformation* that maps  $\mathcal{G}$  to a new graph  $\mathcal{G}' = (\mathcal{V}, \mathcal{E}, s')$  that differs only in its signature from the original graph. The switched signature function  $s' : \mathcal{E} \rightarrow \{-1, 1\}$  is generated through the relationship

$$s'(i, j) = \theta(i)s(i, j)\theta(j). \quad (1)$$

Switching a single vertex flips the signature of all edges in  $\mathcal{E}$  that point into and out of that vertex. When  $\mathcal{V}^-$  contains more than one vertex, following (1) a switching transformation flips the signature of all edges that connect vertices in  $\mathcal{V}^-$  with those in  $\mathcal{V}^+$ , while the signature of edges between two vertices in  $\mathcal{V}^+$  or in  $\mathcal{V}^-$  remains unaltered. Whenever  $\mathcal{G}$  and  $\mathcal{G}'$  are related by a switching transformation, we say they are *switching equivalent*.<sup>1</sup>

A *switching matrix* associated with a switching transformation is a diagonal matrix  $\Theta = \text{diag}(\theta(1), \theta(2), \dots, \theta(n))$ . The adjacency matrices of  $\mathcal{G}$  and  $\mathcal{G}'$  are similar, with  $A' = \Theta A \Theta$ , where we note that  $\Theta = \Theta^{-1}$ . Two signed graphs  $\mathcal{G}$ ,  $\mathcal{G}'$  on  $n$  vertices are *isomorphic* if there exists an isomorphism between  $|\mathcal{G}|$  and  $|\mathcal{G}'|$  that preserves the signature of edges in  $\mathcal{G}$  and  $\mathcal{G}'$ . Formally,  $\mathcal{G}$  and  $\mathcal{G}'$  are isomorphic if there exists a permutation matrix  $P_n$  such that  $A' = P_n A P_n^T$ . If  $\mathcal{G}$  is switching equivalent to a graph that is isomorphic to  $\mathcal{G}'$ , we say the two graphs are *switching isomorphic*. For two switching isomorphic graphs, the adjacency matrices are related by a similarity transformation generated by a switching matrix  $\Theta$  and a permutation matrix  $P_n$ , with

$$A' = P_n \Theta A \Theta P_n^T = \Theta P_n A P_n^T \Theta. \quad (2)$$

A graph  $\mathcal{G}$  is called *bipartite* if it admits a bipartition of nodes  $\mathcal{V}_1, \mathcal{V}_2$  such that  $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}$  and  $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$ , and for all  $(i, j) \in \mathcal{E}$ ,  $i \in \mathcal{V}_q$ ,  $j \in \mathcal{V}_p$ ,  $q, p \in \{1, 2\}$  and  $q \neq p$ . A signed graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, s)$  is said to be *sign-symmetric* if it is switching isomorphic to its negation  $-\mathcal{G} = (\mathcal{V}, \mathcal{E}, -s)$ . A bipartite signed graph is trivially sign-symmetric, and other non-bipartite constructions of sign-symmetric undirected graphs

can be found in [47].

Next, we define two important classes of signed graphs:

*Definition 1:* Consider a signed graph  $\mathcal{G}$  on  $n$  vertices.

*Class I graph.*  $\mathcal{G}$  is switching isomorphic to  $\mathcal{G}'$  with an eventually positive adjacency matrix  $A'$ .

*Class II graph.*  $\mathcal{G}$  is digon-symmetric, strongly connected, and structurally balanced, i.e. switching isomorphic to a  $\mathcal{G}'$  with adjacency matrix  $A' \succeq 0_{n \times n}$ .

The two graph classes identified by Definition 1 are not mutually exclusive. We illustrate the distinction between the two in the following example.

*Example 2.1:* Consider graphs  $\mathcal{G}_1, \mathcal{G}_2$  on 4 vertices with adjacency matrices

$$A_1 = \begin{pmatrix} 0 & -1 & 1 & 1 \\ 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

$A_1$  is eventually positive, as  $(A_1)^6$  has all positive entries. However, it is not digon symmetric since  $(A_1)_{12} = -1$  and  $(A_1)_{21} = 1$ . Thus,  $\mathcal{G}_1$  and any graph  $\mathcal{G}'_1$  that is switching isomorphic to  $\mathcal{G}_1$  belongs to Class I of Definition 1 but not to Class II. Meanwhile  $A_2$  is not eventually positive, but it is digon-symmetric, irreducible, and structurally balanced. Thus,  $\mathcal{G}_2$  and any graph  $\mathcal{G}'_2$  that is switching isomorphic to  $\mathcal{G}_2$  belongs to Class II but not to Class I.

In later sections of the manuscript, signed graphs whose adjacency matrices have a simple dominant eigenvalue will be of particular interest in the belief formation problem. In the following Lemma we show that belonging to Class I or Class II of Definition 1 is a sufficient condition for this.

*Lemma 2.2 (Graphs with simple dominant eigenvalue):*

Consider a signed graph  $\mathcal{G}$  on  $n$  vertices that is Class I or Class II or both. Then the following statements are true:

i) The adjacency matrix  $A$  has a unique dominant eigenvalue satisfying  $\lambda = \rho(A) > \text{Re}(\lambda_i)$  for all  $\lambda_i \neq \lambda$  in  $\sigma(A)$ . If  $\mathcal{G}$  is Class I, then additionally  $|\lambda| > |\lambda_i|$  for all  $\lambda_i \neq \lambda$  in  $\sigma(A)$ ;

ii) There exist a positive vector  $\mathbf{v}' \succ \mathbf{0}$ , a switching matrix  $\Theta$ , and a permutation matrix  $P_n$  such that  $\mathbf{v} = \Theta P_n \mathbf{v}'$  is an eigenvector of  $A$  corresponding to its dominant eigenvalue  $\lambda$ .

*Proof:* i) If  $\mathcal{G}$  is Class I, by Definition 1 and Proposition 2.1.i,  $A$  is switching equivalent to a graph  $\mathcal{G}'$  with adjacency matrix  $A'$  that has the strong Perron-Frobenius property. If  $\mathcal{G}$  is Class II,  $A'$  is nonnegative and irreducible. Then by Proposition 2.1.ii,  $A'$  has a simple dominant eigenvalue, and since  $A$  and  $A'$  are similar by the transformation (2), they are co-spectral. ii) The similarity transformation between  $A$  and  $A'$  (2) also relates their eigenvectors. ■

Strongly connected graphs with all-positive edge signatures, strongly connected structurally balanced graphs, and graphs with an eventually positive adjacency matrix are all special cases of graphs satisfying Lemma 2.2, that belong to Class I and/or to Class II. These graph properties are commonly linked to important features of opinion formation, both for linear and nonlinear models [4], [31], [48]–[53].

<sup>1</sup>In [4] this is called *gauge equivalence*, while the term switching equivalence is more common in the signed graph literature e.g. [44], [46].



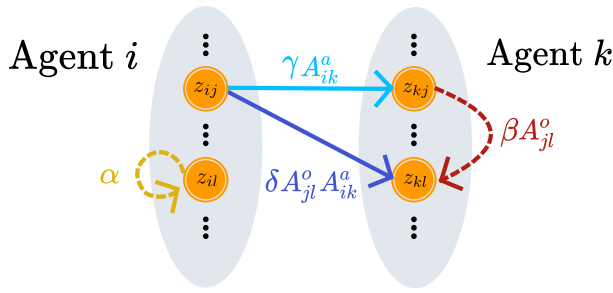


Fig. 1: Four types of communication arrows associated to four different cognitive effects on belief formation in model (3). Arrow direction follows sensing convention.

#### A. Model statement

Agents may form beliefs about options in order to choose among them, for example, when mobile agents choose among alternative heading directions, resource-collecting agents choose among alternative patches, and voters choose among candidates in an election. Agents may also form beliefs about options that present as topics, for example, how strongly they support or reject each of a set of topics such as policy protocols or political issues. In any of these and other scenarios, network relationships may be important: the beliefs among the agents may be interdependent and the beliefs on the different options (or topics) may be logically interdependent.

We study a network of  $N_a$  agents forming beliefs on  $N_o$  options according to a nonlinear update rule introduced in [15], [16], [52]. Let the belief state of agent  $i$  be represented by the vector  $\mathbf{Z}_i = (z_{i1}, z_{i2}, \dots, z_{iN_o}) \in \mathbb{R}^{N_o}$ , where  $z_{ij} \in \mathbb{R}$  is the *belief* agent  $i$  has about option  $j$ . In [16],  $z_{ij}$  is also interpreted as the *value* agent  $i$  assigns to option  $j$ . Let  $\mathbf{Z}_j^\dagger = (z_{1j}, \dots, z_{N_a j}) \in \mathbb{R}^{N_a}$  represent the belief state of the network on option  $j$ . We say agent  $i$  is *neutral* when  $\mathbf{Z}_i = \mathbf{0}$  and *opinionated* otherwise. The strength of belief of agent  $i$  about option  $j$  is proportional to the magnitude of  $z_{ij}$ , with the agent *favoring* option  $j$  when  $z_{ij} > 0$  and *rejecting* option  $j$  when  $z_{ij} < 0$ . The *network state* represents the belief states of all agents,  $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_{N_a}) \in \mathbb{R}^N$ , where  $N = N_a N_o$ .

There are two signed directed graphs that are fundamental to the belief formation process. The first is the *communication graph* among agents,  $\mathcal{G}_a = (\mathcal{V}_a, \mathcal{E}_a, s_a)$  where  $\mathcal{V}_a = \{1, \dots, N_a\}$  is the vertex set corresponding to the  $N_a$  agents,  $\mathcal{E}_a$  is the edge set, and  $s_a : \mathcal{E}_a \rightarrow \{-1, 1\}$  is the signature of the communication graph  $\mathcal{G}_a$ . When  $e_{ik} \in \mathcal{E}_a$ , agent  $k$  is a *neighbor* of agent  $i$  and the belief state of agent  $k$  influences the belief formation of agent  $i$ . When  $s_a(e_{ik}) = 1$ , agent  $i$  is *cooperative* towards agent  $k$ , and whenever  $s_a(e_{ik}) = -1$  it is *competitive* or *antagonistic* towards agent  $k$ . We assume that  $\mathcal{G}_a$  is *simple*, i.e. contains no self-loops  $e_{ii} \notin \mathcal{E}_a$  for all  $i \in \mathcal{V}_a$ , and there is at most one edge  $e_{ik}$  in  $\mathcal{E}_a$  that begins at vertex  $i$  and ends at vertex  $k$  for all  $i, k \in \mathcal{V}_a$ . The signed *adjacency matrix* of the communication graph is the matrix  $A_a \in \mathbb{R}^{N_a \times N_a}$  whose entries are defined as  $(A_a)_{ik} = 0$  if  $e_{ik} \notin \mathcal{E}_a$  and  $s_a(e_{ik})$  otherwise.

The second fundamental graph is the *belief system graph*  $\mathcal{G}_o = (\mathcal{V}_o, \mathcal{E}_o, s_o)$  that encodes the logical, psychological, or social constraints on the relationships between beliefs on

different options [54]. The nodes in the vertex set  $\mathcal{V}_o = \{1, \dots, N_o\}$  correspond to distinct options (topics). An edge  $e_{jl} \in \mathcal{E}_o$  thereby signifies that formation of beliefs about option  $j$  is affected by the beliefs about option  $l$ . The signature function  $\sigma_o : \mathcal{E}_o \mapsto \{-1, 1\}$  describes whether two options are positively or negatively aligned according to the belief system. We also assume that  $\mathcal{G}_o$  is simple, and we define the signed adjacency matrix  $A_o \in \mathbb{R}^{N_o \times N_o}$  whose entries are defined as  $(A_o)_{jl} = 0$  if  $e_{jl} \notin \mathcal{E}_o$  and  $s_o(e_{jl})$  otherwise. In this paper, we assume that all agents share a single global belief system.

We allow each agent  $i$  to form a vector of local observations of its neighbors' belief states  $\mathbf{y}_i = (\mathbf{y}_{ik_1}, \dots, \mathbf{y}_{ik_{m_i}}) \in \mathbb{R}^{m_i N_o}$ , where  $m_i$  is the number of neighbors of agent  $i$  and for each neighbor  $k_l$ ,  $l = 1, \dots, m_i$ ,  $\mathbf{y}_{ik_l} = (y_{ik_l 1}, \dots, y_{ik_l N_o}) = f_i(\mathbf{Z}_{k_l})$ , with  $f_i : \mathbb{R}^{N_o} \rightarrow \mathbb{R}^{N_o}$ . This formulation allows for potentially noisy, partial, or otherwise imperfect information to be shared over the communication network. Agents are assumed to update beliefs in continuous time according to

$$\begin{aligned} \dot{z}_{ij} = & -d_i z_{ij} + u_i \left( S_1 \left( \alpha_i z_{ij} + \gamma_i \sum_{\substack{k=1 \\ k \neq i}}^{N_a} (A_a)_{ik} y_{ikj} \right) \right. \\ & \left. + \sum_{\substack{l \neq j \\ l=1}}^{N_o} S_2 \left( \beta_i (A_o)_{jl} z_{il} + \delta_i (A_o)_{jl} \sum_{\substack{k=1 \\ k \neq i}}^{N_a} (A_a)_{ik} y_{ikl} \right) \right) + b_{ij}. \end{aligned} \quad (3)$$

where  $d_i > 0$  is a damping coefficient that quantifies the agents' resistance towards forming strong beliefs,  $b_{ij} \in \mathbb{R}$  is a *bias* of agent  $i$  towards option  $j$  or an *input* to agent  $i$  about the quality of option  $j$ , and  $\alpha_i, \gamma_i, \beta_i, \delta_i \geq 0$  are the gains that regulate the relative strengths of four cognitive effects within individual  $i$ , each corresponding to a distinct arrow type in Fig. 1. The gain  $\alpha_i$  reflects self-reinforcement or self-appraisal of beliefs within individual  $i$ . The gains  $\gamma_i, \beta_i, \delta_i$  reflect the relative balance of three distinct sources of cognitive dissonance within agent  $i$ . We explain these terms in more detail in the following section.

We let  $\mathbf{b}_i = (b_{i1}, \dots, b_{iN_o})$  be agent  $i$ 's bias vector, and  $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_{N_a})$  the network bias vector. The parameter  $u_i \geq 0$  captures the *attention* agents allocate to social interactions in their evaluation of the options. Alternatively,  $u$  describes the *urgency* agents feel towards forming strong beliefs. The functions  $S_m : \mathbb{R} \rightarrow [-k_{m1}, k_{m2}]$ , with  $k_{m1}, k_{m2} > 0$ ,  $S'_m(y) > 0$ ,  $m = 1, 2$ , saturate social information on each of the options, bounding the net effect of social influence of each topic on the formation of beliefs. We assume that  $S_m(y) = \hat{S}_m(y) + g_m(y)$  with  $S_m(0) = 0$ ,  $\hat{S}_m(-y) = -\hat{S}_m(y)$ ,  $\hat{S}'_m(y) > 0$ ,  $\hat{S}'_m(0) = 1$ ,  $g_m(0) = g'_m(0) = 0$ , and  $g_m(-y) \neq -g_m(y)$ ,  $m = 1, 2$ . When  $g_1(y) = g_2(y) = 0$ , for all  $y$ , the functions  $S_1, S_2$  are odd symmetric and agents are internally equally likely to favor or reject each option. When  $g_m(y) \neq 0$ , for some  $y$ , for  $m = 1$  and/or 2, it perturbs the odd symmetry and captures internal asymmetries in the formation of beliefs. The two saturations  $S_1, S_2$  are distinct in (4) to allow for potentially different bounds on the effects of social imitation, i.e., influence of beliefs along the same option dimension captured by  $S_1$ , and of the ideology, i.e., influence

of beliefs across option dimensions captured by  $S_2$ .<sup>2</sup>

For the analysis in this paper, we will assume perfect observation by letting  $\mathbf{y}_{ik} = f_i(\mathbf{Z}_k) = \mathbf{Z}_k$  for all neighboring pairs  $i, k \in \mathcal{V}_a$ . We also assume agents to be homogeneous by letting  $d_i = d > 0$ ,  $\alpha_i = \alpha \geq 0$ ,  $\gamma_i = \gamma \geq 0$ ,  $\beta_i = \beta \geq 0$ , and  $\delta_i = \delta \geq 0$  for all  $i \in \mathcal{V}_a$ . Note, however, that agents can be affected by heterogeneous biases or inputs  $b_{ij}$ . Under these assumptions, model (3) becomes

$$\begin{aligned} \dot{z}_{ij} = & -d z_{ij} + u \left( S_1 \left( \alpha z_{ij} + \gamma \sum_{\substack{k=1 \\ k \neq i}}^{N_a} (A_a)_{ik} z_{kj} \right) \right. \\ & \left. + \sum_{\substack{l=1 \\ l \neq j}}^{N_o} S_2 \left( \beta (A_o)_{jl} z_{il} + \delta (A_o)_{jl} \sum_{\substack{k=1 \\ k \neq i}}^{N_a} (A_a)_{ik} z_{kl} \right) \right) + b_{ij} \end{aligned} \quad (4)$$

or, equivalently in vector-matrix form,

$$\begin{aligned} \dot{\mathbf{Z}} = & -d\mathbf{Z} + u\mathbf{S}_1((\alpha\mathcal{I}_{N_a} + \gamma A_a) \otimes \mathcal{I}_{N_o})\mathbf{Z} \\ & + \sum_{l=1}^{N_o} \mathbf{S}_2(((\beta\mathcal{I}_{N_a} + \gamma A_a) \otimes M_l)\mathbf{Z}) + \mathbf{b} \end{aligned} \quad (5)$$

where  $\mathbf{S}_m(\mathbf{y}) = (S_m(y_1), \dots, S_m(y_n))$  for  $\mathbf{y} \in \mathbb{R}^n$ , and  $M_l \in \mathbb{R}^{N_a \times N_a}$  is the matrix whose column  $l$  coincides with column  $l$  of  $A_o$ , with zero entries in all other columns.

Model (3) generalizes the opinion dynamics and value formation models of [15] and [16] in the following ways. First, in both of the preceding works, the explicit notion of a belief system is absent and there is all-to-all, same-sign coupling between the options. In contrast, a structured belief system is central to the analysis we present here. Second, motivated by an interpretation of opinions on the simplex, [15] enforces the constraint, not enforced here, that agents' internal beliefs must sum to zero. Lastly, the social networks analyzed in [15] and [16] were constrained to purely cooperative or purely antagonistic interactions between pairs of agents, whereas here we consider networks with potentially mixed social relationships.

## B. Connection with Networks of Beliefs theory

The belief formation model (3) has a compelling interpretation in the context of the Networks of Beliefs (NB) theory recently proposed in social psychology [9]. The NB framework consists of three main premises:

- 1) Beliefs can be represented as nodes in a network, with an *internal* network that describes interactions of beliefs within an individual, and an *external* network that describes interactions between individuals;
- 2) Belief formation seeks to reduce *dissonance* between one's internal beliefs and beliefs on the social network;
- 3) At least three types of dissonance contribute to the formation of beliefs: *personal*, *social*, and *external*. Individuals can allocate different amounts of attention to each.

<sup>2</sup>Equivalently, saturation can instead be applied to all of the dissonance terms together or to individual belief states  $z_{ij}, y_{ikj}$  - see Appendix.

In our model, the belief system graph  $\mathcal{G}_o$  describes the internal network and the graph product  $\mathcal{G}_a \times \mathcal{G}_o$  with adjacency matrix  $A_a \otimes A_o$  describes the external network in the sense of [9]. NB theory also distinguishes between individuals' *personal* beliefs about a topic and individuals' *social* beliefs, i.e. their perceptions about the beliefs of their neighbors. In (3), the belief state  $z_{ij}$  corresponds to the personal belief of agent  $i$  on option  $j$  and the vector of local observations  $\mathbf{y}_{ik}$  is the set of social beliefs of agent  $i$  about agent  $k$ .

The update of an agent's internal belief  $z_{ij}$  in (3) is proportional to  $-(z_{ij} - \frac{u_i}{d_i} F_{ij}(\mathbf{Z}_i, \mathbf{y}_i))$  where  $F_{ij}$  is a shorthand for the nonlinear terms grouped by parentheses in (3). In the terminology of NB theory, the difference in the parentheses represents the magnitude of dissonance individual  $i$  experiences due to inconsistencies between its personal belief on option  $j$  and its other personal and social beliefs. For unbiased agents, a belief  $z_{ij}$  will decrease whenever the dissonance term is positive, and increase whenever it is negative, thus minimizing dissonance until equilibrium is reached. Further observe that for an unbiased agent, dissonance is exactly zero at any equilibrium of the model (3).

Following the third premise of NB theory, the parameters  $\beta_i, \gamma_i, \delta_i$  in (3) are identified with effects of three distinct sources of dissonance within an individual:

- $\beta_i$  is identified with *personal dissonance*, i.e., the inconsistency between an individual's personal belief  $z_{ij}$  and all of its other personal beliefs, evaluated according to some logical, moral, or ideological compass encoded in the belief system graph  $\mathcal{G}_o$ ;
- $\gamma_i$  is identified with *social dissonance*, i.e., the inconsistency between an agent's personal belief  $z_{ij}$  and its social beliefs  $y_{ikj}$  (its perceptions about the beliefs of its neighbors on option  $j$ );
- $\delta_i$  is identified with *external dissonance*, i.e., the inconsistency between an agent's personal belief on option  $z_{ij}$  and its social beliefs  $y_{ikl}$  (its perceptions about the beliefs of its neighbors on all other options), evaluated according to some logical, moral, or ideological compass encoded in the belief system graph  $\mathcal{G}_o$ .

The magnitudes of  $\beta_i, \gamma_i, \delta_i$  modulate the relative amount of attention allocated by the agents towards each of these effects.

For an example of these effects, consider forming beliefs on two political topics: increasing the taxation rate for the wealthy and raising immigration rates. According to the conventional left-right political ideology spectrum, a set of beliefs that either supports or opposes both of these issues simultaneously is considered logically consistent. With this in mind, an individual who supports increased taxes but opposes raising immigration will experience *personal* dissonance. Someone who supports increased taxes but perceives that their friends oppose increased taxation will experience *social* dissonance. Finally, an individual who supports increased taxes but perceives that their friends oppose raising immigration will experience *external* dissonance when updating their belief on taxation.

NB theory is presented in [9] alongside a computational model of social belief formation, which we refer to as the NB model. Our model is both distinct from and complementary to the NB model of [9]. Firstly, the NB model adopts a

statistical physics, i.e., stochastic dynamics, modeling perspective whereas our approach is deterministic. This aspect makes our model amenable to dynamical systems analysis, which we rely on to establish the forthcoming theoretical results. Secondly, our model incorporates factors not included in the NB model, such as the self-appraisal of personal beliefs, parameterized by  $\alpha_i$ , and the impact of internal biases  $b_{ij}$ . Lastly, although both models account for external dissonance, the definition of this effect differs between the two works. In the NB model, external dissonance arises from inconsistencies between individuals' social beliefs  $\mathbf{y}_{ik}$  and the true personal beliefs of their neighbors  $\mathbf{Z}_k$ . This definition requires agents to access  $\mathbf{Z}_k$  in addition to forming the observation  $\mathbf{y}_{ik}$  for each of their neighbors. Our model (3) does not require measuring the true states  $\mathbf{Z}_k$  of neighbors in its full generality. Instead, individuals assess whether their neighbors' perceived beliefs appear internally consistent, according to the belief system  $\mathcal{G}_o$ . Unlike the NB model, in our model agents can experience external dissonance even if their social belief  $\mathbf{y}_{ik}$  accurately reflects the true personal belief  $\mathbf{Z}_k$  of their neighbor, as is the case for the analysis we present.

#### IV. LINEAR BIFURCATION ANALYSIS

##### A. Belief-forming bifurcations

In this section, we begin our analysis of the homogeneous belief formation model (4). The analysis we carry out in this paper draws on standard ideas from the theory of bifurcations in networks and coupled cell systems [55]–[59]. We first consider a network of unbiased agents with  $b_{ij} = 0$  for all  $i \in \mathcal{V}_a, j \in \mathcal{V}_o$ , in which case the neutral state  $\mathbf{Z} = \mathbf{0}$  is an equilibrium of the dynamics for all choices of parameters  $d, u, \alpha, \gamma, \beta, \delta$ . From the vector-matrix form of the dynamic equations (5) we can easily deduce that the Jacobian matrix of the linearization about  $\mathbf{Z} = \mathbf{0}$  has the structure

$$J(\mathbf{0}, u) = (-d + u\alpha)\mathcal{I}_{N_a} \otimes \mathcal{I}_{N_o} + u\gamma A_a \otimes \mathcal{I}_{N_o} \\ + u\beta \mathcal{I}_{N_a} \otimes A_o + u\delta A_a \otimes A_o. \quad (6)$$

The first term on the left hand side of (6) corresponds to the balance of resistance and self-appraisal, the second to social dissonance, the third to personal dissonance, and the fourth to external dissonance. These terms reflect the four distinct coupling arrows identified in Fig. 1. In the following lemma we identify how spectral properties of  $J(\mathbf{0}, u)$  depend on the social network graph  $\mathcal{G}_a$ , the belief system graph  $\mathcal{G}_o$ , and model parameters.

*Lemma 4.1 (Jacobian spectrum [52]):* The following hold for (6). 1) For each  $\eta \in \sigma(J(\mathbf{0}, u))$ , there exists  $\lambda \in \sigma(A_a)$  and  $\mu \in \sigma(A_o)$  so that

$$\eta = -d + u(\alpha + \gamma\lambda + \beta\mu + \delta\lambda\mu) := \eta(u, \lambda, \mu); \quad (7)$$

2) For each pair  $\lambda_i \in \sigma(A_a)$  with an eigenvector  $\mathbf{v}_{a,i}$  and  $\mu_j \in \sigma(A_o)$  with an eigenvector  $\mathbf{v}_{o,j}$ , the vector  $\mathbf{v}_{a,i} \otimes \mathbf{v}_{o,j}$  is an eigenvector of (6) with eigenvalue  $\eta(u, \lambda_i, \mu_j)$ .

*Proof:* See [52, Proposition IV.1] ■

With Lemma 4.1 we have established a bijective correspondence between the Jacobian spectrum  $\sigma(J(\mathbf{0}))$  and the product

set of the adjacency spectra  $\sigma(A_a), \sigma(A_o)$ . For every pair  $\lambda \in \sigma(A_a), \mu \in \sigma(A_o)$  we will say  $\lambda, \mu$  generate the Jacobian eigenvalue  $\eta(u, \lambda, \mu)$  defined in (7) for some constant choice of model parameters  $d, \alpha, \beta, \gamma, \delta$ . Furthermore, eigenspaces of  $J(\mathbf{0})$  coincide with those of  $A_a \otimes A_o$ , independently of the choice of model parameters. We take advantage of these properties to prove our first result, in which we derive necessary and sufficient conditions for the emergence of a low-dimensional locally attracting invariant manifold in the dynamics (4) as the neutral equilibrium  $\mathbf{Z} = \mathbf{0}$  loses stability. This result generalizes [15, Theorem IV.1] to mixed-sign communication and belief networks.

*Theorem 4.2 (Belief-forming bifurcation):* Consider model (4) with  $b_{ij} = 0$  for all  $i \in \mathcal{V}_a, j \in \mathcal{V}_o$  and let

$$\Lambda = \operatorname{argmax}_{(\lambda, \mu) \in \sigma(A_a) \times \sigma(A_o)} \gamma \operatorname{Re}(\lambda) + \beta \operatorname{Re}(\mu) + \delta \operatorname{Re}(\lambda\mu). \quad (8)$$

Let  $(\lambda, \mu) \in \Lambda$  and  $\xi_{max} = \alpha + \gamma \operatorname{Re}(\lambda) + \beta \operatorname{Re}(\mu) + \delta \operatorname{Re}(\lambda\mu)$ , and assume  $\xi_{max} > 0$ .

1) The neutral equilibrium  $\mathbf{Z} = \mathbf{0}$  is locally exponentially stable whenever  $0 \leq u < u^*$  and unstable for  $u > u^*$ , where

$$u^* = \frac{d}{\alpha + \gamma \operatorname{Re}(\lambda) + \beta \operatorname{Re}(\mu) + \delta \operatorname{Re}(\lambda\mu)} = \frac{d}{\xi_{max}}. \quad (9)$$

2) Let  $|\ker(J(\mathbf{0}, u^*))| := k$ . There exists a  $(k + 1)$ -dimensional invariant center manifold  $W^c \subset \mathbb{R}^{N_a N_o + 1}$  passing through  $(\mathbf{Z}, u) = (\mathbf{0}, u^*)$ , tangent to  $\mathcal{N}(J)$  at  $u = u^*$ . All trajectories of (4) starting at  $(\mathbf{Z}, u)$  near  $(\mathbf{0}, u^*)$  converge to  $W^c$  exponentially as  $t \rightarrow \infty$ .

*Proof:* Following Lemma 4.1, for any leading eigenvalue  $\eta(u, \lambda, \mu)$  of (6),  $\operatorname{Re}(\eta(u, \lambda, \mu)) = -d + u\xi$ . When  $u < u^*$ ,  $\operatorname{Re}(\eta) < 0$  and therefore  $\operatorname{Re}(\tilde{\eta}) < 0$  for all  $\tilde{\eta} \in \sigma(J(\mathbf{0}, u))$ ; for  $u > u^*$ ,  $\operatorname{Re}(\eta) > 0$ . Stability conclusions follow by Lyapunov's indirect method [60, Theorem 4.7]. The existence of an attracting center manifold follows by the Center Manifold Theorem [61, Theorem 3.2.1]. ■

There are several key takeaways from the analysis so far:

1) Unbiased agents that are weakly attentive to social interactions remain neutral and do not form strong beliefs;

2) When network attention exceeds the critical value (9), non-neutral beliefs emerge as attracting solution sets;

3) The critical attention value  $u^*$  is determined by the choice of model parameters and by the eigenvalue pairs of  $A_a$  and  $A_o$  contained in the set  $\Lambda$  defined by (8).

A primary aim of this paper is to characterize the properties of solutions of model (4) that arise as a result of the critical transition in which the neutral equilibrium loses stability as  $u$  increases above  $u^*$ . We can already infer the role of the structure of the graphs  $\mathcal{G}_a, \mathcal{G}_o$  in shaping these emerging beliefs in the group. When the neutral equilibrium  $\mathbf{Z}$  loses stability, by Theorem 4.2 we expect the belief trajectories  $\mathbf{Z}(t)$  to settle on the low-dimensional center manifold  $W^c$ . By Lemma 4.1, the center manifold is given, to first order, by the eigenspace of  $A_a \otimes A_o$  associated to eigenvalues in the set (8).<sup>3</sup> A straightforward eigenvalue problem on

<sup>3</sup>More generally,  $W^c$  is approximated by a generalized eigenspace of  $A_a \otimes A_o$ ; for cases considered in this paper the multiplicity of relevant eigenvalues coincides with dimensionality of their eigenspaces so this technical issue does not play any role in the derived results.



$A_a \otimes A_o$  thus provides us information both about *where* (i.e., at what parameter values) to expect a critical belief-forming bifurcation in the group, as well as *what pattern* the agents' beliefs will take on post-bifurcation. We expand on these ideas more formally in upcoming sections by studying the onset of belief-forming equilibria and oscillations in model (4).

### B. Effect of graph structure and internal and social parameters on belief-forming bifurcations

In this section we ask the question: how do the communication and belief system graphs jointly with the model parameters select the eigenstructure of  $A_a \otimes A_o$  (and therefore of  $J(\mathbf{0}, u)$ ) that determines the linear bifurcation behavior of (4), i.e., which eigenvalues belong to the set  $\Lambda$  defined in (8)?

Let an eigenvalue pair  $(\lambda, \mu)$  be a *principal eigenvalue pair* if it belongs to  $\Lambda$ . Let an eigenspace of  $A_a \otimes A_o$  be a *principal eigenspace* if its associated eigenvalue pairs are principal, i.e. it is the leading eigenspace of  $J(\mathbf{0}, u)$ . Also let  $\lambda_{max} = \max_{\lambda_i \in \sigma(A_a)} \{\text{Re}(\lambda_i)\}$ ,  $\mu_{max} = \max_{\mu_j \in \sigma(A_o)} \{\text{Re}(\mu_j)\}$ , and  $(\lambda\mu)_{max} = \max_{\lambda_i \in \sigma(A_a), \mu_j \in \sigma(A_o)} \{\text{Re}(\lambda_i \mu_j)\}$ . We define two sets of eigenvalues that turn out to be key in determining the principal eigenstructure of  $A_a \otimes A_o$ :

- 1) Let  $\Lambda_1$  be a set of ordered pairs  $(\lambda, \mu) \in \sigma(A_a) \times \sigma(A_o)$  for which  $\text{Re}(\lambda) = \lambda_{max}$ ,  $\text{Re}(\mu) = \mu_{max}$ .
- 2) Let  $\Lambda_2$  be a set of ordered pairs  $(\lambda, \mu) \in \sigma(A_a) \times \sigma(A_o)$  for which  $\text{Re}(\lambda\mu) = (\lambda\mu)_{max}$ .

An *indecision-breaking bifurcation* in (4) typically happens either along the product of leading eigenspaces of  $A_a, A_o$ , or along the leading eigenspace of the graph product  $A_a \otimes A_o$ . The set  $\Lambda_1$  is associated with the product of leading eigenspaces of  $A_a, A_o$  and its cardinality reflects the dimension of this eigenspace, while the set  $\Lambda_2$  is similarly associated with the leading eigenspace of  $A_a \otimes A_o$ . Before we formalize this observation, in the following lemma we characterize graphs for which the structure of the sets  $\Lambda_1$  and/or  $\Lambda_2$  is easily identified. Observe that  $\Lambda_1, \Lambda_2$  are not necessarily disjoint and for many common choices of  $\mathcal{G}_a$  and  $\mathcal{G}_o$ ,  $\Lambda_1 \subseteq \Lambda_2$ . Let Class I and Class II graphs be defined as in Definition 1.

**Lemma 4.3 (Graph conditions):** i) Suppose  $\mathcal{G}_a$  and  $\mathcal{G}_o$  are each either Class I or Class II; then  $\Lambda_1 = \{(\lambda_{max}, \mu_{max})\}$ ;

ii) Suppose  $\mathcal{G}_a$  and  $\mathcal{G}_o$  are both Class I; then  $\Lambda_2 = \Lambda_1 = \{(\lambda_{max}, \mu_{max})\}$ ;

iii) Suppose  $\mathcal{G}_a$  and  $\mathcal{G}_o$  are both Class II. If  $\mathcal{G}_a$  and  $\mathcal{G}_o$  are both undirected and sign-symmetric, then

$$\Lambda_2 = \{(\lambda_{max}, \mu_{max}), (-\lambda_{max}, -\mu_{max})\};$$

iv) Suppose  $\mathcal{G}_a$  and  $\mathcal{G}_o$  are both Class II and both undirected. If the characteristic polynomial  $\det(xI - A) = \sum_{i \in \mathcal{V}} c_i x^{n-i}$  for both  $A_a, A_o$  satisfies  $c_i = 0$  for any even  $i$ , the set  $\Lambda_2 = \{(\lambda_{max}, \mu_{max}), (-\lambda_{max}, -\mu_{max})\}$ ; otherwise,  $\Lambda_2 = \Lambda_1 = \{(\lambda_{max}, \mu_{max})\}$ .

*Proof:* See Appendix. ■

In the following results we show under which conditions the principal eigenstructure determining the bifurcation behavior can be characterized by  $\Lambda_1$  and/or  $\Lambda_2$ .

**Proposition 4.4:** Consider (4) with communication graph  $\mathcal{G}_a$  and belief system graph  $\mathcal{G}_o$ .

1) Suppose  $\beta > 0, \gamma > 0$ . There exists small  $\varepsilon$  such that whenever  $\delta < \varepsilon$ ,  $\Lambda = \Lambda_1$ .

2) Suppose  $\delta > 0$ . There exist small  $\varepsilon_1, \varepsilon_2$  such that whenever  $\gamma < \varepsilon_1$  and  $\beta < \varepsilon_2$ ,  $\Lambda = \Lambda_2$ .

*Proof:* Let  $Q(\lambda, \mu) = \gamma \text{Re}(\lambda) + \beta \text{Re}(\mu) + \delta \text{Re}(\lambda\mu)$ . 1) When  $\delta = 0$ ,  $\max_{(\lambda, \mu) \in \sigma(A_a) \times \sigma(A_o)} Q(\lambda, \mu) = \gamma \max_{\lambda \in \sigma(A_a)} \text{Re}(\lambda) + \beta \max_{\mu \in \sigma(A_o)} \text{Re}(\mu)$ . Any pair  $(\lambda, \mu)$  satisfying this condition must be in the set  $\Lambda_1$  by the definition of the set. By continuity of eigenvalues, this will hold for  $\delta < \varepsilon$ . 2) When  $\beta = \gamma = 0$ ,  $\max_{(\lambda, \mu) \in \sigma(A_a) \times \sigma(A_o)} Q(\lambda, \mu) = \delta \max_{(\lambda, \mu) \in \sigma(A_a) \times \sigma(A_o)} \text{Re}(\lambda\mu)$  and the statement follows trivially by the definition of the set  $\Lambda_2$ . By continuity of eigenvalues, this will hold for  $\gamma < \varepsilon_1, \beta < \varepsilon_2$ . ■

In Proposition 4.4 we have shown that when external dissonance is weak ( $\delta$  is zero or small),  $\Lambda_1$  is the principal eigenvalue set, and when social dissonance and personal dissonance are weak ( $\gamma, \beta$  are zero or small),  $\Lambda_2$  is the principal eigenvalue set. Next, we prove a similar result but in the converse limit. Namely, we prove that  $\Lambda_1$  ( $\Lambda_2$ ) is the principal set when social dissonance and/or personal dissonance (external dissonance) are sufficiently strong.

**Proposition 4.5:** Consider (4) with  $u > 0$  a communication graph  $\mathcal{G}_a$ , and a belief system graph  $\mathcal{G}_o$ . Define the second largest real part of eigenvalues of  $A_a$  and  $A_o$  as  $\lambda_2 = \max_{\lambda \in \sigma(A_a) \text{ s.t. } \text{Re}(\lambda) \neq \lambda_{max}} \{\text{Re}(\lambda)\}$ ,  $\mu_2 = \max_{\mu \in \sigma(A_o) \text{ s.t. } \text{Re}(\mu) \neq \mu_{max}} \{\text{Re}(\mu)\}$ .

1) There exists a critical value  $K_c > 0$ , with all other parameters held constant, such that if  $\gamma(\lambda_{max} - \lambda_2) + \beta(\mu_{max} - \mu_2) > K_c$ , then  $(\lambda, \mu) \in \Lambda$  if and only if  $\text{Re}(\lambda) = \lambda_{max}$  and  $\text{Re}(\mu) = \mu_{max}$ , i.e.  $\Lambda = \Lambda_1$ ;

2) There exists a critical value  $\delta_c$ , with all other parameters held constant, such that if  $\delta > \delta_c$ , then  $(\lambda, \mu) \in \Lambda$  if and only if  $\text{Re}(\lambda\mu) = (\lambda\mu)_{max}$ , i.e.  $\Lambda = \Lambda_2$ .

*Proof:* See Appendix. ■

### C. Network of Beliefs theory interpretation

This section interprets the results of Propositions 4.4 and 4.5 in the context of the NB theory. We start with a few key takeaways:

- 1) *Self-appraisal tunes the amount of attention or urgency needed to form strong beliefs.* The parameter  $\alpha$  is inversely proportional to the critical attention magnitude (9). When agents are strongly self-appraising, the belief-forming bifurcation occurs at a lower  $u^*$ . Because  $\alpha$  does not play any role in selecting the principal eigenspace along which belief formation happens, this is the primary effect of self-appraisal.
- 2) *Social dissonance drives network beliefs towards the leading eigenspaces of  $\mathcal{G}_a$  and  $\mathcal{G}_o$ .* By Propositions 4.4, 4.5 when the social dissonance parameter  $\gamma$  dominates, the eigenvalue set  $\Lambda_1$  is principal for any choice of  $\mathcal{G}_a$  and  $\mathcal{G}_o$ . In this case, the principal eigenspace of  $J(\mathbf{0}, u)$  is a product of the leading eigenspaces of  $A_a$  and  $A_o$ . In light of this observation, we call a belief-forming bifurcation for which  $\Lambda_1$  is the principal eigenvalue set *social dissonance-driven*.
- 3) *External dissonance reshapes belief formation but only on specific communication and belief system graphs.*

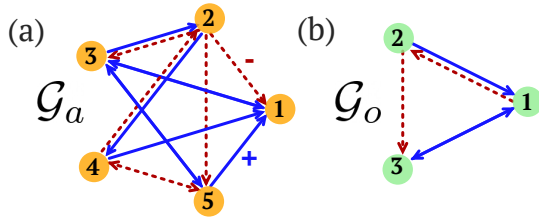


Fig. 2: (a) Communication graph  $\mathcal{G}_a$  for five agents and (b) belief system graph  $\mathcal{G}_o$  for three options with signed adjacency matrices (10). Solid blue edges are positive connections and red dashed edges are negative connections. Arrow direction follows sensing convention.

By Propositions 4.4, 4.5, when the external dissonance parameter  $\delta$  is large enough, the eigenvalue set  $\Lambda_2$  is principal and the principal eigenspace of  $J(\mathbf{0}, u)$  coincides with the leading eigenspace of the graph product  $A_a \otimes A_o$ . Hence, whenever  $\mathcal{G}_a$  and  $\mathcal{G}_o$  are such that  $\Lambda_2 \neq \Lambda_1$ , a sufficiently large external dissonance weight can drastically change the belief-forming behavior for two reasons. First, the manifold along which the belief-forming bifurcation occurs will be different in the external dissonance-driven and the social dissonance-driven case. Second, the *type* of bifurcation that occurs inside the bifurcation manifold will in general be different, e.g. whenever the two sets have different cardinalities. On the same communication and belief system graphs  $\mathcal{G}_a$ ,  $\mathcal{G}_o$ , social dissonance and external dissonance-driven bifurcations in (4) can lead to sharply different belief-forming behaviors.

- 4) *Personal dissonance amplifies the effects of social dissonance.* Another important observation from Propositions 4.4, 4.5 is that the effect of personal dissonance parameter  $\beta$  is coupled with that of the social dissonance parameter  $\gamma$ , rather than with that of the external dissonance parameter  $\delta$ . Even when the social dissonance coupling  $\gamma$  is weak, a strong level of personal dissonance  $\beta$  will lead to a social dissonance-driven bifurcation shaped by the eigenvalue set  $\Lambda_1$ . Agents are more susceptible to social pressure when they are less secure internally about their personal identity with respect to the belief system  $\mathcal{G}_o$ .

*Example 4.1:* Consider a network of 5 agents evaluating 3 options according to (4), with communication graph  $\mathcal{G}_a$  and belief system graph  $\mathcal{G}_o$  pictured in Fig. 2. The signed adjacency matrices of these graphs are

$$A_a = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & -1 & 1 & -1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 \end{pmatrix}, \quad A_o = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix}. \quad (10)$$

Rounded to three decimal points, the eigenvalues of  $A_a$  are  $\lambda_1 \approx 0.823$ ,  $\lambda_{2,3} \approx 0.745 \pm 1.106i$ ,  $\lambda_{4,5} \approx -1.157 \pm 0.327i$  and the eigenvalues of  $A_o$  are  $\mu_1 = 1$ ,  $\mu_{2,3} = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}i$ . The set of leading eigenvalues of  $A_a, A_o$  is  $\Lambda_1 = \{(\lambda_1, \mu_1)\}$  and the set of eigenvalue pairs that maximize the real part of the eigenvalue products of the two graphs is  $\Lambda_2 = \{(\lambda_4, \mu_3), (\lambda_5, \mu_2)\}$ . For belief formation dynamics (4) defined on these graphs, this means that a social dissonance-

driven bifurcation will happen along a one-dimensional manifold, and an external dissonance-driven bifurcation will happen along a different, two-dimensional manifold.

## V. MULTI-STABLE BELIEF-FORMATION: PITCHFORK BIFURCATION

In this section we characterize what types of limit set solutions for dynamics (4) arise at a belief-forming bifurcation. We focus here on the simplest, but also generic, case in which the leading eigenvalue of the Jacobian matrix (6) is real and simple. In light of Theorem 4.2, this amounts to requiring that the set of principal eigenvalues  $\Lambda$  contains exactly one pair of real-valued eigenvalues of  $A_a$  and  $A_o$ . In the following Theorem we prove that under this assumption, the dynamics (4) exhibit a pitchfork bifurcation.

*Theorem 5.1 (Multi-option pitchfork bifurcation):*

Consider (4) with communication graph  $\mathcal{G}_a$  and belief system graph  $\mathcal{G}_o$ , and suppose  $\Lambda = \{(\lambda_a, \mu_o)\}$  with  $\lambda_a, \mu_o \in \mathbb{R}$ . Assume  $K_0 := \alpha + \gamma\lambda_a + \beta\mu_o + \delta\lambda_a\mu_o > 0$ . Let  $\mathbf{v}_a, \mathbf{w}_a \in \mathbb{R}^{N_a}$  and  $\mathbf{v}_o, \mathbf{w}_o \in \mathbb{R}^{N_o}$  be the right and left eigenvectors of  $A_a$  and  $A_o$  corresponding to  $\lambda_a$  and  $\mu_o$ , respectively, normalized to satisfy  $\langle \mathbf{w}_a, \mathbf{v}_a \rangle = 1$ ,  $\langle \mathbf{w}_o, \mathbf{v}_o \rangle = 1$ . Let  $K_1 = (S_1'''(0)(\alpha + \gamma\lambda_a)^3 + S_2'''(0)(\beta\mu_o + \delta\lambda_a\mu_o)^3) \langle \mathbf{w}_a, \mathbf{v}_a^3 \rangle \langle \mathbf{w}_o, \mathbf{v}_o^3 \rangle$  and  $K_2 := (g_1''(0)(\alpha + \gamma\lambda_a)^2 + g_2''(0)(\beta\mu_o + \delta\lambda_a\mu_o)^2) \langle \mathbf{w}_a, \mathbf{v}_a^2 \rangle \langle \mathbf{w}_o, \mathbf{v}_o^2 \rangle \neq 0$  where  $\mathbf{x}^2 = \mathbf{x} \odot \mathbf{x}$  and  $\mathbf{x}^3 = \mathbf{x} \odot \mathbf{x} \odot \mathbf{x}$ .

1) **Unbiased agents with symmetric response.** Suppose  $S_1, S_2$  have odd symmetry ( $g_1(x) = g_2(x) = 0$ ). At  $(\mathbf{Z}, u, \mathbf{b}) = (\mathbf{0}, u^*, \mathbf{0})$ , the system undergoes a symmetric pitchfork bifurcation. All bifurcation branches are tangent at  $(\mathbf{Z}, u) = (\mathbf{0}, u^*)$  to  $\text{span}\{\mathbf{v}_a \otimes \mathbf{v}_o\}$ . If  $K_1 < 0$  ( $> 0$ ), the pitchfork is supercritical (subcritical), that is, the non-trivial equilibria  $\mathbf{Z}^*(u), -\mathbf{Z}^*(u) \neq \mathbf{0}$  bifurcating from  $\mathbf{0}$  exist for  $u > u^*$  ( $u < u^*$ ),  $|u - u^*|$  sufficiently small, and are locally exponentially stable (unstable).

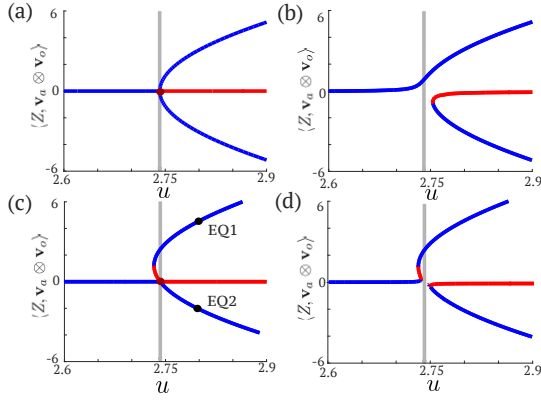
2) **Unbiased agents with asymmetric response.** Suppose  $S_1, S_2$  are asymmetric ( $g_1(x) \neq 0$  and/or  $g_2(x) \neq 0$ ), and  $|g_1''(0)|, |g_2''(0)|$  are sufficiently small. At  $(\mathbf{Z}, u, \mathbf{b}) = (\mathbf{0}, u^*, \mathbf{0})$ , the system undergoes a transcritical bifurcation. All bifurcation branches are tangent at  $(\mathbf{Z}, u) = (\mathbf{0}, u^*)$  to  $\text{span}\{\mathbf{v}_a \otimes \mathbf{v}_o\}$ . The non-trivial equilibria  $\mathbf{Z}_1^*(u), \mathbf{Z}_2^*(u) \neq \mathbf{0}$  exist for  $|u - u^*|$  sufficiently small and satisfy  $\langle \mathbf{v}_a \otimes \mathbf{v}_o, \mathbf{Z}_1^*(u) \rangle > 0$ ,  $\langle \mathbf{v}_a \otimes \mathbf{v}_o, \mathbf{Z}_2^*(u) \rangle < 0$ . If  $K_2 > 0$  ( $< 0$ ),  $\mathbf{Z}_1^*(u)$  ( $\mathbf{Z}_2^*(u)$ ) appears for  $u < u^*$  and is unstable, while  $\mathbf{Z}_2^*(u)$  ( $\mathbf{Z}_1^*(u)$ ) appears for  $u > u^*$  and is locally exponentially stable.

3) **Biased agents.** Suppose  $\mathbf{b} \neq \mathbf{0}$  and  $\|\mathbf{b}\|$  is small. Whenever  $\langle \mathbf{w}_a \otimes \mathbf{w}_o, \mathbf{b} \rangle > 0$  ( $< 0$ ), for sufficiently small  $|u - u^*|$ , the local bifurcation diagram of (4) has a unique, locally exponentially stable equilibrium point  $\mathbf{Z}^*(u)$ . This equilibrium satisfies  $\langle \mathbf{v}_a \otimes \mathbf{v}_o, \mathbf{Z}^*(u) \rangle > 0$  ( $< 0$ ).

*Proof:* See Appendix.  $\blacksquare$

*Example 5.1:* To illustrate Theorem 5.1 we consider model (4) with the communication and belief system graphs of Fig. 2 and Example 4.1. Given model parametrization described in Fig. 3, it can easily be established that  $\Lambda = \Lambda_1 = \{(\lambda_1, \mu_1)\}$ , i.e., the indecision-breaking bifurcation is social dissonance-driven, and the assumptions of Theorem 5.1 are satisfied. In

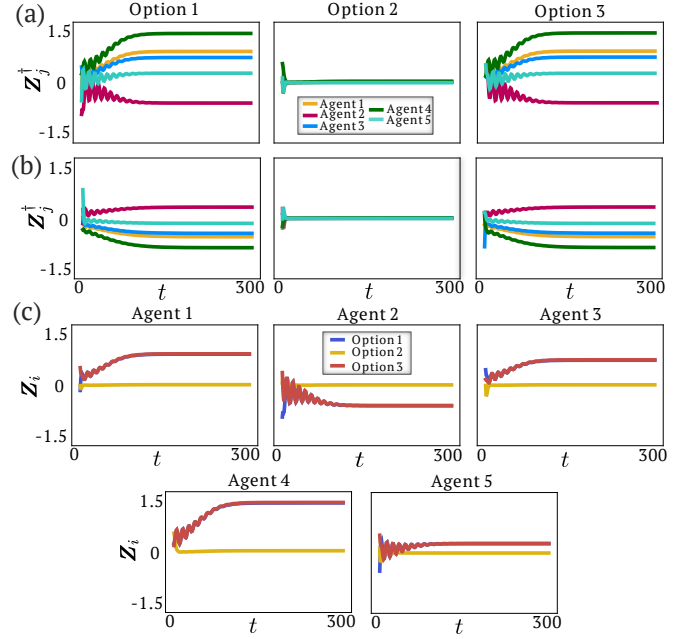




**Fig. 3:** Bifurcation diagrams for Example 5.1, generated with MatCont numerical continuation package [62]. Red (blue) lines represent unstable (stable) equilibria.  $S_1(y) = \tanh(y + \varepsilon_1 \tanh(y^2))$ ,  $S_2(y) = \frac{1}{2} \tanh(2y + 2\varepsilon_2 \tanh(y^2))$ , (a)  $\varepsilon_1 = \varepsilon_2 = 0$ ,  $\langle \mathbf{w}_a \otimes \mathbf{w}_o, \mathbf{b} \rangle = 0$ ; (b)  $\varepsilon_1 = \varepsilon_2 = 0$ ,  $\langle \mathbf{w}_a \otimes \mathbf{w}_o, \mathbf{b} \rangle > 0$ ; (c)  $\varepsilon_1 = \varepsilon_2 = 0.1$ ,  $\langle \mathbf{w}_a \otimes \mathbf{w}_o, \mathbf{b} \rangle = 0$ ; (d)  $\varepsilon_1 = \varepsilon_2 = 0.1$ ,  $\langle \mathbf{w}_a \otimes \mathbf{w}_o, \mathbf{b} \rangle > 0$ . For (b) and (d),  $b_{11} = 0.001$ ,  $b_{22} = 0.003$ ,  $b_{53} = -0.002$  and  $b_{ij} = 0$  for all other  $i \in \mathcal{V}_a, j \in \mathcal{V}_o$ . Vertical gray line indicates the bifurcation point  $u = u^* \approx 2.742$ . Other parameters:  $d = 1$ ,  $\alpha = \gamma = \beta = \delta = 0.1$ .

Fig. 3 we show four bifurcation diagrams that summarize the predictions of Theorem 5.1: a symmetric pitchfork bifurcation for unbiased agents with odd symmetric  $S_1, S_2$  in (a), an unfolding of a symmetric pitchfork for agents with small biases in (b), a transcritical bifurcation for unbiased agents with asymmetric response in (c), and an unfolding of the asymmetric case for agents with small biases in (d). Observe that in the asymmetric diagram (c), the unstable branch of nontrivial equilibria regains stability in a saddle-node bifurcation and for  $u > u^*$  there is a bistability between the two nontrivial branches, similarly to the odd symmetric case. This is expected in general as long as  $S_1''(0)$  and  $S_2''(0)$  are sufficiently small, as a consequence of unfolding theory for a pitchfork bifurcation - see [63, Chapter I §1] for a discussion. In this sense, the bistability of belief equilibria is a feature of dynamics (4) that is robust to asymmetries in the agents' internal and social responses. Fig. 4 shows simulated belief trajectories settling on EQ1 and EQ2 of Fig. 3. At steady state, the vectors  $\mathbf{z}_j^*$ ,  $j = 1, \dots, N_o$ , of agents' beliefs about a given option in (a) and (b) are roughly (modulo higher-order in the pitchfork center manifold expansion) aligned with the communication graph eigenvectors  $(\mathbf{v}_o)_j \mathbf{v}_a$ . Similarly, at steady state, the vectors  $\mathbf{z}_i$ ,  $i = 1, \dots, N_o$ , of an agent's beliefs about the set of option in (c) are roughly aligned with the belief system eigenvectors  $(\mathbf{v}_a)_i \mathbf{v}_o$ . In particular, all beliefs about option 2 remain close to neutral because  $(\mathbf{v}_o)_2 = 0$ .

Next, we establish sufficient conditions on the structure of communication and belief system graphs for the existence of a pitchfork bifurcation in (4). Recall that a social dissonance-driven bifurcation happens whenever the set of bifurcating dominant eigenvalues (8) consists of the leading eigenvalues of signed adjacency matrices  $A_a$  and  $A_o$ . Given a graph  $\mathcal{G}$  with adjacency matrix  $A$  that is either Class I or II (by Definition 1), let  $\theta : \mathcal{V} \rightarrow \{1, -1\}$  be the switching function with switching matrix  $\Theta$ , and let  $P_n$  be a permutation matrix



**Fig. 4:** Trajectories  $z_{ij}(t)$  for belief formation dynamics (4) for 5 agents evaluating 3 options on  $\mathcal{G}_a, \mathcal{G}_o$  of Fig. 2 from random initial conditions. The figure illustrates the distribution of agents' states  $z_{ij}$  along each option dimension as the network settles to (a) EQ1 from Figure 3(c); (b) EQ2 from Figure 3(c); (c) shows same trajectories as (a), grouped by agent instead of option. Belief trajectories on options 1 and 3 overlap for most of the simulation for all agents. Parameters:  $u = u^* + 0.05 \approx 2.792$ , all other parameters as in Fig. 3(c).

for which  $\Theta P_n A P_n^T \Theta$  is eventually positive if  $\mathcal{G}$  is Class I or nonnegative if  $\mathcal{G}$  is Class II.

*Corollary 5.1:* Suppose  $A_a, A_o$  each has a simple leading eigenvalue  $\lambda_{max}, \mu_{max}$ , respectively. Let  $\mathbf{v}_a, \mathbf{v}_o$  be the corresponding eigenvectors. Define the sets  $\mathcal{V}_a^0 = \{i \in \mathcal{V}_a \text{ s.t. } (\mathbf{v}_a)_i = 0\} \subset \mathcal{V}_a$  and  $\mathcal{V}_o^0 = \{j \in \mathcal{V}_o \text{ s.t. } (\mathbf{v}_o)_j = 0\} \subset \mathcal{V}_o$ . The following statements hold:

i) Suppose  $\mathcal{G}_a$  and  $\mathcal{G}_o$  are each either Class I or Class II; then any social dissonance-driven indecision-breaking bifurcation in (4) must be a pitchfork bifurcation. Let  $\mathbf{Z}^* = (z_{11}^*, z_{12}^*, \dots, z_{N_a N_o}^*)$  be an equilibrium on one of the new solution branches of the pitchfork bifurcation with sufficiently small  $|u - u^*|$ . Then  $z_{ij}^* \neq 0$  for all  $i \in \mathcal{V}_a, j \in \mathcal{V}_o$ .

ii) Suppose  $\mathcal{G}_a$  and  $\mathcal{G}_o$  are both Class I; then for any choice of model parameters, the indecision-breaking bifurcation in (4) is a social dissonance-driven pitchfork bifurcation.

iii) Suppose  $\mathcal{G}_a$  is either Class I or Class II,  $b_{ij} = 0$  for all  $i \in \mathcal{V}_a, j \in \mathcal{V}_o$ , and the indecision-breaking bifurcation in (4) is a social dissonance-driven pitchfork bifurcation. Let  $\mathbf{Z}^* = (z_{11}^*, \dots, z_{N_a N_o}^*)$  be an equilibrium on one of the new solution branches with sufficiently small  $|u - u^*|$ . The switching function  $\theta_a$  induces a bipartition  $\mathcal{V}_{a1}, \mathcal{V}_{a2}, \mathcal{V}_{a1} \cup \mathcal{V}_{a2} = \mathcal{V}_a, \mathcal{V}_{a1} \cap \mathcal{V}_{a2} = \emptyset$ , of the agent set  $\mathcal{V}_a$  such that  $i \in \mathcal{V}_{a1}$  whenever  $\theta_a(i) = 1$  and  $i \in \mathcal{V}_{a2}$  whenever  $\theta_a(i) = -1$ . Whenever  $i, k \in \mathcal{V}_p, p \in \{a1, a2\}$ , and  $j \in \mathcal{V}_o \setminus \mathcal{V}_o^0$ ,  $\text{sign}(z_{ij}^*) = \text{sign}(z_{kj}^*)$  and  $\text{sign}(z_{ij}^*) \neq \text{sign}(z_{kj}^*)$  otherwise. If  $|(\mathbf{v}_a)_i| > |(\mathbf{v}_a)_k|$ , then  $|z_{ij}^*| > |z_{kj}^*|$  for all  $j \in \mathcal{V}_o$ .

iv) Suppose  $\mathcal{G}_o$  is either Class I or Class II,  $b_{ij} = 0$  for all  $i \in \mathcal{V}_a, j \in \mathcal{V}_o$ , and the indecision-breaking bifurcation

in (4) is a social dissonance-driven pitchfork bifurcation. Let  $\mathbf{Z}^* = (z_{11}^*, \dots, z_{N_a N_o}^*)$  be an equilibrium on one of the new solution branches with sufficiently small  $|u - u^*|$ . The switching function  $\theta_o$  induces a bipartition  $\mathcal{V}_{o1}, \mathcal{V}_{o2}, \mathcal{V}_{o1} \cup \mathcal{V}_{o2} = \mathcal{V}_o$ ,  $\mathcal{V}_{o1} \cap \mathcal{V}_{o2} = \emptyset$ , of options set  $\mathcal{V}_o$  such that  $j \in \mathcal{V}_{o1}$  whenever  $\theta_o(j) = 1$  and  $j \in \mathcal{V}_{o2}$  whenever  $\theta_o(j) = -1$ . Whenever  $j, l \in \mathcal{V}_p$ ,  $p \in \{o1, o2\}$ , and  $i \in \mathcal{V}_a \setminus \mathcal{V}_a^0$ ,  $\text{sign}(z_{ij}^*) = \text{sign}(z_{il}^*)$  and  $\text{sign}(z_{ij}^*) \neq \text{sign}(z_{il}^*)$  otherwise. If  $|(\mathbf{v}_o)_j| > |(\mathbf{v}_o)_l|$ , then  $|z_{ij}^*| > |z_{il}^*|$  for all  $i \in \mathcal{V}_a$ .

*Proof:* i) By Lemma 4.3.i,  $\Lambda_1 = \{(\lambda_{max}, \mu_{max})\}$ . Since  $\lambda_{max}, \mu_{max} > 0$ ,  $\alpha + \gamma\lambda_{max} + \beta\mu_{max} + \delta\lambda_{max}\mu_{max} > 0$  and the conditions of Theorem 5.1 are satisfied whenever  $\Lambda = \Lambda_1$ . ii) By Lemma 4.3.ii,  $\Lambda = \Lambda_1$  for any choice of model parameters. By the argument in part i), the conditions of Theorem 5.1 are satisfied and the statement follows. iii) By Theorem 4.2, the equilibria lie on a manifold that is approximated arbitrarily closely by the span of  $\mathbf{v}_a \otimes \mathbf{v}_o$  when  $|u - u^*|$  is small. Following Lemma 2.2.ii, there exists a vector  $\mathbf{v}' \succ \mathbf{0}$  such that  $\mathbf{v}_a = \Theta P_n \mathbf{v}'$ . The vector  $P_n \mathbf{v}' \succ \mathbf{0}$ , and it follows that  $(\mathbf{v}_a)_i > 0 (< 0)$  if and only if  $\theta(i) = 1 (-1)$ . The belief  $z_{ij}^* \approx (\mathbf{v}_a)_i (\mathbf{v}_o)_j$ ; as long as  $(\mathbf{v}_o)_j \neq 0$ ,  $\text{sign}(z_{kj}^*) = \text{sign}((\mathbf{v}_a)_k (\mathbf{v}_o)_j) = \text{sign}((\mathbf{v}_a)_i (\mathbf{v}_o)_j) = \text{sign}(z_{ij}^*)$  if and only if  $\theta(i) = \theta(k)$ . iv) The statement follows by analogous argument to part iii). ■

For a social dissonance-driven bifurcation with  $\mathcal{G}_a$  and  $\mathcal{G}_o$  in Class I or II, Corollary 5.1 predicts relative signs and ordering of strength of equilibrium opinions, the latter according to the ordering of components in the graph eigenvector  $\mathbf{v}_a$  or  $\mathbf{v}_o$ . According to the Corollary, each graph belonging to Class I and/or Class II of Definition 1 is a sufficient condition for a pitchfork bifurcation. Importantly, these are not *necessary* conditions as we already observed with Example 5.1 (Figs. 3-4). The belief system graph  $\mathcal{G}_o$  of Fig. 2(b) used in Example 5.1 is not digon symmetric.  $\mathcal{G}_o$  is also not eventually structurally balanced:  $A_o^2$  and  $A_o^3 = I$  have zero entries, and  $A_o^4 = A_o$ . Yet the bifurcation observed in Example 5.1 is a social dissonance-driven pitchfork bifurcation. In general, when the structure of social network graph  $\mathcal{G}_a$  and belief system graph  $\mathcal{G}_o$  are known, the conditions of Theorem 5.1 are easily verified by computing eigenvalues of adjacency matrices  $A_a, A_o$ .

Nonetheless, the sufficient conditions of Corollary 5.1 are particularly meaningful in the context of belief formation. Structural balance in a graph corresponds to a bipartition of the nodes, with strictly positive edges within each subgroup and strictly negative edges between subgroups. For a communication graph this bipartition corresponds to the split of a group into two mutually antagonistic subgroups, as may occur in a polarized social network. For a belief system graph, structural balance may correspond, for example, to a partition of political issues into two subgroups based on their alignment with left-leaning or right-leaning ideological views. Corollary 5.1.iv is additionally related to the notion of *coherence* of beliefs [35], [54]. An individual holds a maximally coherent set of beliefs when their beliefs don't violate the logical constraints imposed by the belief system  $\mathcal{G}_o$ . According to Corollary 5.1.iv, when the graph  $\mathcal{G}_o$  is bipartite, all individuals on the network form maximally coherent beliefs, as their belief vector at steady-state directly reflects the bipartition of beliefs encoded in the

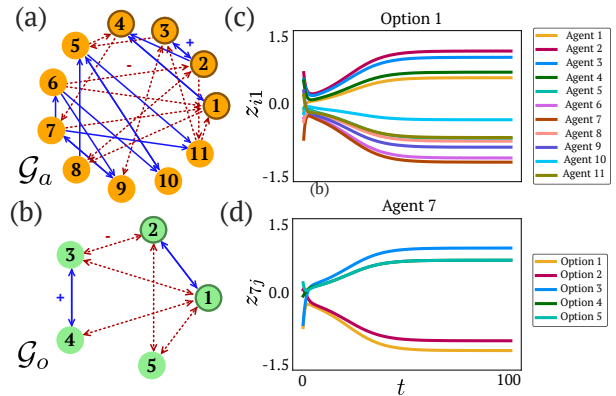


Fig. 5: (a) Strongly connected, structurally balanced communication graph with node partitions  $\mathcal{V}_{a1} = \{1, 2, 3, 4\}$ ,  $\mathcal{V}_{a2} = \{5, 6, 7, 8, 9, 10, 11\}$ ; (b) Strongly connected, structurally balanced belief system graph with node partitions  $\mathcal{V}_{o1} = \{1, 2\}$ ,  $\mathcal{V}_{o2} = \{3, 4, 5\}$ ; (c) belief trajectories of all 11 agents on option 1; (d) belief trajectories of agent 7 on all 5 options; belief trajectories on options 4 and 5 overlap for most of the simulation. Parameters:  $u = u^* + 0.05 \approx 0.7627$ ,  $\alpha = \gamma = \beta = \delta = 0.1$ ,  $d = 1$ ,  $b_{ij} = 0$  for all  $i \in \mathcal{V}_a, j \in \mathcal{V}_o$ ,  $S_1(\cdot) = \tanh(\cdot)$ ,  $S_2(\cdot) = \frac{1}{2} \tanh(2\cdot)$ .

structure of  $\mathcal{G}_o$ .

*Example 5.2:* Consider the belief dynamics model (4) for 11 agents forming beliefs on 5 options with structurally balanced  $\mathcal{G}_a, \mathcal{G}_o$  pictured in Fig. 5(a),(b). With the indicated choice of parameters, the system satisfies the conditions of Theorem 5.1 and exhibits a social dissonance-driven pitchfork bifurcation. According to Corollary 5.1.iii, due to the structural balance of the communication graph we expect the beliefs of agents 1-4 on each option to share sign at steady-state, while agents 5-11 will form same-sign beliefs that are opposite-sign from agents 1-4 following the bipartition in  $\mathcal{G}_a$ . Belief trajectories simulated from a random initial condition in Fig. 5(c) confirm this prediction. Analogously, according to Corollary 5.1.iv we expect each agent at steady-state to simultaneously favor options 1 and 2 while rejecting the rest, or simultaneously favor options 3-5 while rejecting 1 and 2 following the bipartition in  $\mathcal{G}_o$ . This prediction is confirmed by the simulated belief trajectories in Fig. 5(d).

## VI. OSCILLATORY BELIEF-FORMATION: HOPF BIFURCATION

So far we have explored the emergence of multistable belief equilibria through pitchfork bifurcations, a generic outcome of indecision-breaking in the belief formation model (4). We now study under which conditions a belief-forming bifurcation can lead to an oscillatory belief state. Oscillations in belief systems have received scant attention to date, both in formal models and in empirical studies of social systems. Within an individual, belief oscillations are thought by some to be an important dynamic feature of cognition that plays a role in decision-making and attitude change, a view that is supported by indirect experimental evidence [64], [65]. At the group level, oscillations in beliefs and preferences are commonly observed in human societies, e.g., in recurring fashion trends [66] and in recurring swings between conservative and liberal

political attitudes [67]. We derive sufficient and necessary conditions for belief oscillations that arise from a Hopf bifurcation in the belief formation model (4). The results in this section are partially adapted from the recent conference paper [52].

A Hopf bifurcation is a local bifurcation in which a limit cycle appears near a singular equilibrium of a nonlinear system. Consider the belief formation model (4). A necessary condition for the indecision-breaking bifurcation point  $(\mathbf{Z}, u) = (\mathbf{0}, u^*)$  of Theorem 4.2 to be a *Hopf bifurcation point* is for the Jacobian  $J(\mathbf{0}, u)$  (6) to have a complex conjugate pair of leading eigenvalues  $\eta_+(u) = \eta_r(u) + i\eta_i(u)$ ,  $\eta_-(u) = \bar{\eta}_+(u) = \eta_r(u) - i\eta_i(u)$ , with  $\eta_+(u^*) = i\eta_i(u^*)$  and  $\eta_-(u^*) = -i\eta_i(u^*)$  being the only eigenvalues of  $J(\mathbf{0}, u^*)$  on the imaginary axis [61, Theorem 3.4.2]. To enforce this necessary condition we introduce the following assumption.

*Assumption 1 (Complex-conjugate leading eigenvalues):*

The set  $\Lambda$  of leading eigenvalues of  $J(\mathbf{0}, u)$ , defined in (8), satisfies one of the following three conditions:

- 1) *Complex eigenvalue in belief system graph:*  $\Lambda = \{(\lambda, \mu), (\lambda, \bar{\mu})\}$  where  $\lambda \in \mathbb{R}$ ,  $\mu \in \mathbb{C}$ ,  $\lambda, \text{Re}(\mu), \text{Im}(\mu) \neq 0$ ;
- 2) *Complex eigenvalue in social network graph:*  $\Lambda = \{(\lambda, \mu), (\bar{\lambda}, \mu)\}$  where  $\lambda \in \mathbb{C}$ ,  $\mu \in \mathbb{R}$ ,  $\mu, \text{Re}(\lambda), \text{Im}(\lambda) \neq 0$ ;
- 3) *Interaction of complex eigenvalues:*  $\Lambda = \{(\lambda, \mu), (\bar{\lambda}, \bar{\mu})\}$  where  $\lambda, \mu \in \mathbb{C}$ ,  $\text{Re}(\lambda), \text{Im}(\lambda), \text{Re}(\mu), \text{Im}(\mu) \neq 0$ .

*Theorem 6.1 (Hopf bifurcation):* [52, Theorem IV.3] Consider (4) with communication graph  $\mathcal{G}_a$  and belief system graph  $\mathcal{G}_o$ . Let Assumption 1 hold. Suppose  $\lambda^\dagger \in \sigma(A_a)$ ,  $\mu^\dagger \in \sigma(A_o)$  generate  $\eta_+(u)$ , i.e.  $\eta_+(u) = \eta(u, \lambda^\dagger, \mu^\dagger)$ , and assume  $\alpha + \gamma \text{Re}(\lambda^\dagger) + \beta \text{Re}(\mu^\dagger) + \delta \text{Re}(\lambda^\dagger \mu^\dagger) > 0$ . Let  $\mathbf{w}_a, \mathbf{v}_a \in \mathbb{C}^{N_a}$  be the left and right eigenvectors of  $A_a$  corresponding to  $\lambda^\dagger$  and  $\bar{\lambda}^\dagger$ , respectively; let  $\mathbf{w}_o, \mathbf{v}_o \in \mathbb{C}^{N_o}$  be the left and right eigenvectors of  $A_o$  corresponding to  $\mu^\dagger$  and  $\bar{\mu}^\dagger$ , respectively, satisfying the biorthogonal normalization condition

$$\langle \mathbf{w}_a \otimes \mathbf{w}_o, \mathbf{v}_a \otimes \mathbf{v}_o \rangle = 2, \quad \langle \overline{\mathbf{w}_a \otimes \mathbf{w}_o}, \mathbf{v}_a \otimes \mathbf{v}_o \rangle = 0.$$

For sufficiently small  $|g_1''(0)|$ ,  $|g_2''(0)|$ , the following hold:

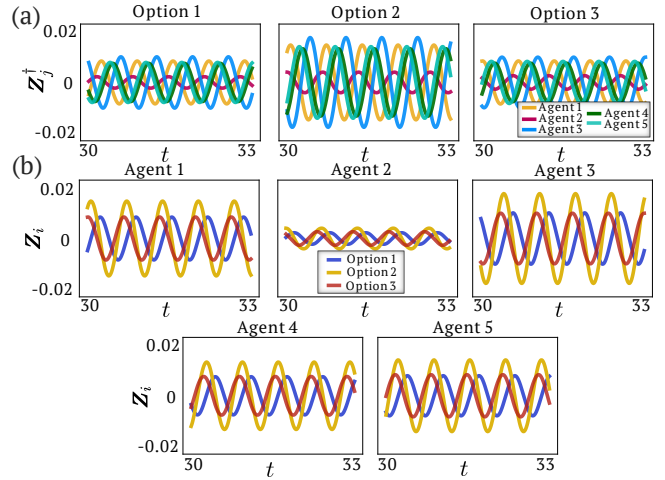
1) There is a unique 3-dimensional center manifold  $W^c \subset \mathbb{R}^{N_a N_o} \times \mathbb{R}$  passing through  $(\mathbf{Z}, u) = (\mathbf{0}, u^*)$ , tangent to  $\text{span}\{\text{Re}(\mathbf{v}_a \otimes \mathbf{v}_o), \text{Im}(\mathbf{v}_a \otimes \mathbf{v}_o)\}$  at  $u = u^* = d/(\alpha + \gamma \text{Re}(\lambda^\dagger) + \beta \text{Re}(\mu^\dagger) + \delta \text{Re}(\lambda^\dagger \mu^\dagger))$ . There is a family of periodic orbits of (4) that bifurcates from the neutral equilibrium  $\mathbf{Z} = \mathbf{0}$  along  $W_c$  at  $u = u^*$ ;

2) Let  $K = \text{Re}((S_1'''(0)(\alpha + \gamma \lambda^\dagger)|\alpha + \gamma \lambda^\dagger|^2 + S_2'''(0)(\beta + \delta \lambda^\dagger)\mu^\dagger|\beta + \delta \lambda^\dagger|^2|\mu^\dagger|^2)\langle \mathbf{w}_a \otimes \mathbf{w}_o, \mathbf{v}_a \otimes \mathbf{v}_o \rangle + \text{Im}(\mathbf{v}_a \otimes \mathbf{v}_o))$  where  $|\mathbf{x}|^2 = \bar{\mathbf{x}} \odot \mathbf{x}$ . Whenever  $K < 0$  the periodic solutions appear supercritically (for  $u > u^*$ ) and are locally asymptotically stable; whenever  $K > 0$ , the solutions appear subcritically (for  $u < u^*$ ) and are unstable;

3) When  $|u - u^*|$  is small, the period of the solutions  $\mathbf{Z}^*(t)$  is near  $T = 2\pi/(\omega^*|\gamma \text{Im}(\lambda^\dagger) + \beta \text{Im}(\mu^\dagger) + \delta \text{Im}(\lambda^\dagger \mu^\dagger)|)$ , the difference in phase between  $z_{ij}^*(t)$  and  $z_{kl}^*(t)$  is near  $\varphi_{ik}^{jl} = \arg((\mathbf{v}_a)_i(\mathbf{v}_o)_j) - \arg((\mathbf{v}_a)_k(\mathbf{v}_o)_l)$ , and the amplitude of  $z_{ij}^*(t)$  is greater than the amplitude of  $z_{kl}^*(t)$  if and only if  $|(\mathbf{v}_a)_i| |(\mathbf{v}_o)_j| > |(\mathbf{v}_a)_k| |(\mathbf{v}_o)_l|$ .

*Proof:* See Appendix. ■

*Example 6.1:* To illustrate the predictions of Theorem 6.1 we use again the communication and belief system graphs



**Fig. 6:** Trajectories  $z_{ij}(t)$  of stable oscillation described in Example 6.1 for (4) on communication and belief system graphs of Fig. 2; (a) grouped by option; (b) grouped by agent. Parameters:  $d = 10$ ,  $u = 1$ ,  $\alpha = \gamma = \beta = 0.1$ ,  $\delta = 12$ ,  $b_{ij} = 0$  for all  $i \in \mathcal{V}_a$ ,  $j \in \mathcal{V}_o$ ,  $S_1(\cdot) = \tanh(\cdot)$ ,  $S_2(\cdot) = \frac{1}{2} \tanh(2 \cdot)$ .

$\mathcal{G}_a, \mathcal{G}_o$  of Fig. 2. Recall from Example 4.1 that for these graphs the set of graph eigenvalue pairs  $(\lambda, \mu) \in \sigma(A_a) \times \sigma(A_o)$  that maximize  $\text{Re}(\lambda\mu)$  has the structure  $\Lambda_2 = \{(\lambda^\dagger, \mu^\dagger), (\bar{\lambda}^\dagger, \bar{\mu}^\dagger)\}$  and satisfies the third case in Assumption 1. It follows from Theorem 6.1 and Proposition 4.5 that any external dissonance-dominant bifurcation on these graphs will be a Hopf bifurcation, and this bifurcation is realized for sufficiently large values of the external dissonance parameter  $\delta$ . Consider (4) on these graphs with the parameters as in Fig. 6. The periodic orbits bifurcate supercritically and are stable. The ordering of phases and amplitudes of these oscillations along each option, as in Fig. 6(a), and internally, as in Fig. 6(b) are predicted by the complex eigenvectors  $\mathbf{v}_a$  and  $\mathbf{v}_o$ .

The three types of eigenvalue combinations in Assumption 1 lead to qualitatively different types of oscillation. Case 1 (Case 2) lead to phase-synchronized intra-agent (inter-agent) beliefs, i.e., the components of  $\mathbf{Z}_i(t)$  ( $\mathbf{Z}_j^\dagger(t)$ ) oscillate with the same phase for each  $i$  (each  $j$ ). Case 3 generically leads to more complex oscillatory phase patterns, unless some kind of symmetry is present in one of the graphs. The following corollary formalizes these observations. Let Class I and Class II be defined as in Definition 1. Given graph  $\mathcal{G}$  with adjacency matrix  $A$  that is either Class I or II, let  $\theta : \mathcal{V} \rightarrow \{1, -1\}$  be the switching function with switching matrix  $\Theta$ , and let  $P_n$  be a permutation matrix for which  $\Theta P_n A P_n^T \Theta$  is eventually positive if  $\mathcal{G}$  is Class I or nonnegative if  $\mathcal{G}$  is Class II.

*Corollary 6.1:* i.1) Suppose that  $\mathcal{G}_a$  is either Class I or Class II with associated dominant eigenvalue  $\lambda_{max}$  and corresponding eigenvector  $\mathbf{v}_a$ , and that  $A_o$  has a complex conjugate pair of leading eigenvalues  $\mu, \bar{\mu}$ . Then, a social dissonance-driven indecision-breaking bifurcation in (4) is a Hopf bifurcation. For sufficiently small  $|u - u^*|$  and  $\mathbf{b} = \mathbf{0}$ , let  $\mathbf{Z}^*(t) = (z_{11}^*(t), \dots, z_{N_a N_o}^*(t))$  be the periodic solution of (4) emerging at the Hopf bifurcation.

i.2) Let  $\mathcal{V}_a^0 = \{i \in \mathcal{V}_a \text{ s.t. } (\mathbf{v}_a)_i = 0\} \subset \mathcal{V}_a$ . Also let  $\mathcal{V}_{a1}, \mathcal{V}_{a2} \subset \mathcal{V}_a$ , with  $\mathcal{V}_{a1} \cup \mathcal{V}_{a2} = \mathcal{V}_a$ ,  $\mathcal{V}_{a1} \cap \mathcal{V}_{a2} = \emptyset$



be the bipartition of  $\mathcal{V}_a$  induced by the switching function  $\theta_a$  such that  $i \in \mathcal{V}_{a1}$  whenever  $\theta_a(i) = 1$  and  $i \in \mathcal{V}_{a2}$  whenever  $\theta_a(i) = -1$ . Suppose  $\|(\mathbf{Z}_j^*)^*(t)\| > 0$  for some  $t \in \mathbb{R}$  and  $j \in \mathcal{V}_o \setminus \mathcal{V}_o^0$ . If  $i, k \in \mathcal{V}_p$ ,  $p \in \{a1, a2\}$ , then  $\text{sign}(z_{ij}^*(t)) = \text{sign}(z_{kj}^*(t))$ . Conversely, if  $i \in \mathcal{V}_p$  and  $k \in \mathcal{V}_s$ ,  $p, s \in \{a1, a2\}$ ,  $p \neq s$ , then  $\text{sign}(z_{ij}^*(t)) \neq \text{sign}(z_{kj}^*(t))$ .

ii.1) Suppose that  $\mathcal{G}_o$  is either Class I or Class II with associated dominant eigenvalue  $\mu_{max}$  and corresponding eigenvector  $\mathbf{v}_o$ , and that  $\mathcal{G}_a$  has a complex conjugate pair of leading adjacency eigenvalues  $\lambda, \bar{\lambda}$ . Then a social dissonance-driven indecision-breaking bifurcation in (4) is a Hopf bifurcation. For sufficiently small  $|u - u^*|$  and  $\mathbf{b} = \mathbf{0}$ , let  $\mathbf{Z}^*(t) = (z_{11}^*(t), \dots, z_{N_a N_o}^*(t))$  be a periodic solution of (4) emerging at the Hopf bifurcation.

ii.2) Let  $\mathcal{V}_o^0 = \{j \in \mathcal{V}_o \text{ s.t. } (\mathbf{v}_o)_j = 0\} \subset \mathcal{V}_a$ . Also let  $\mathcal{V}_{o1}, \mathcal{V}_{o2}, \mathcal{V}_{o1} \cup \mathcal{V}_{o2} = \mathcal{V}_o, \mathcal{V}_{o1} \cap \mathcal{V}_{o2} = \emptyset$  be the bipartition of options induced by the switching function  $\theta_o$  such that  $j \in \mathcal{V}_{o1}$  whenever  $\theta_o(j) = 1$  and  $j \in \mathcal{V}_{o2}$  whenever  $\theta_o(j) = -1$ . Suppose  $\|\mathbf{Z}_i^*(t)\| > \varepsilon$  for some  $t \in \mathbb{R}$ ,  $i \in \mathcal{V}_a \setminus \mathcal{V}_a^0$ , and some small  $\varepsilon$ . If  $j, l \in \mathcal{V}_p$ ,  $p \in \{o1, o2\}$ , then  $\text{sign}(z_{ij}^*(t)) = \text{sign}(z_{il}^*(t))$ . Conversely, if  $j \in \mathcal{V}_p$ ,  $l \in \mathcal{V}_s$ ,  $p, s \in \{a1, a2\}$ ,  $p \neq s$ , then  $\text{sign}(z_{ij}^*(t)) \neq \text{sign}(z_{il}^*(t))$ .

*Proof:* i) The set  $\Lambda_1 = \{(\lambda_{max}, \mu), (\lambda_{max}, \bar{\mu})\}$  corresponding to  $\mathcal{G}_a, \mathcal{G}_o$  satisfies the first case for the necessary condition in Assumption 1 and any social dissonance-driven bifurcation is a Hopf bifurcation. By part 3) of Theorem 6.1, phase difference between  $z_{ij}^*(t)$  and  $z_{kj}^*(t)$  is near  $\varphi_{ik}^{jj} = \arg((\mathbf{v}_a)_i(\mathbf{v}_o)_j) - \arg((\mathbf{v}_a)_k(\mathbf{v}_o)_j) = \arg((\mathbf{v}_a)_i) + \arg((\mathbf{v}_o)_j) - \arg((\mathbf{v}_a)_k) - \arg((\mathbf{v}_o)_j) = \arg\left(\frac{(\mathbf{v}_a)_i}{(\mathbf{v}_a)_k}\right)$  which means, since all entries of  $\mathbf{v}_a$  are real-valued,  $\varphi_{ik}^{jj} = 0$  if  $\text{sign}((\mathbf{v}_a)_i) = \text{sign}((\mathbf{v}_a)_k)$  and  $\varphi_{ik}^{jj} = \pi$  if  $\text{sign}((\mathbf{v}_a)_i) \neq \text{sign}((\mathbf{v}_a)_k)$ . Following Lemma 2.2.ii, there exists a vector  $\mathbf{v}' \succ \mathbf{0}$  such that  $\mathbf{v}_a = \Theta P_n \mathbf{v}'$ .  $P_n \mathbf{v}' \succ \mathbf{0}$ , and it follows that  $(\mathbf{v}_a)_i > 0 (< 0)$  if and only if  $\theta(i) = 1 (-1)$ . So  $\varphi_{ik}^{jj} = 0$  whenever  $i, k$  both belong to either  $\mathcal{V}_{a1}$  or  $\mathcal{V}_{a2}$ ; otherwise  $\varphi_{ik}^{jj} = \pi$ . Part ii) follows by an analogous argument. ■

As a consequence of Corollary 6.1, social dissonance-driven oscillations on structurally balanced communication graphs reflect the partition of a graph into two polarized subgroups because the beliefs of agents in the two subgroups oscillate exactly out-of-phase with one another, i.e., the agents in the two subgroups will always disagree with each other despite the oscillatory trajectories of their beliefs. Analogously, social dissonance-driven oscillations on structurally balanced belief system graphs will reflect the induced bipartition of options in the internal ordering of each agent's beliefs. In other words, despite their oscillatory trajectories, each agent's beliefs remain coherent at all times and do not violate the logical relationships encoded in the structure of belief system  $\mathcal{G}_o$ .

*Example 6.2:* Consider the belief dynamics (4) with belief system graph  $\mathcal{G}_o$  and three different communication graphs, pictured in Fig. 7(a) with parameter values indicated in the figure caption. The leading eigenvalues of  $A_o$  are a complex-conjugate pair. The graphs  $\mathcal{G}'_a$  and  $\mathcal{G}''_a$  are both Class II and switching equivalent, with  $A'_a$  having the Perron-Frobenius property while  $A''_a$  is structurally balanced with a bipartition of nodes  $\mathcal{V}_{a1} = \{1, 2\}$ ,  $\mathcal{V}_{a2} = \{3, 4\}$ . Following Corollary

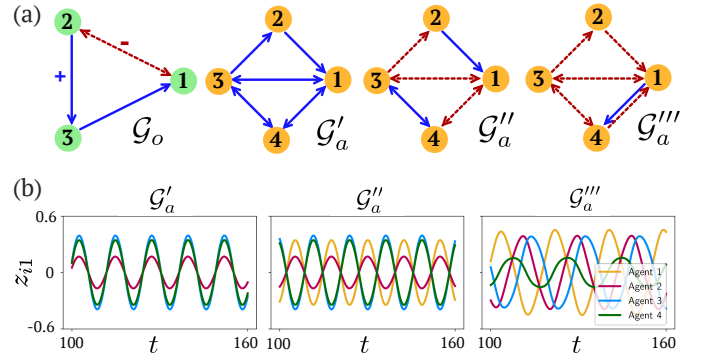


Fig. 7: Figure adapted from [52]. (a) Belief system graph  $\mathcal{G}_o$  and three communication graphs  $\mathcal{G}'_a, \mathcal{G}''_a, \mathcal{G}'''_a$ . (b) Trajectories of belief dynamics for all four agents on option 1 with belief system graph  $\mathcal{G}_o$  and each of the three communication graphs shown in (a). Parameters:  $d = 1$ ,  $\alpha = \gamma = 0.1$ ,  $\beta = \delta = 0.25$ ,  $b_{ij} = 0$  for all  $i \in \mathcal{V}_a$ ,  $j \in \mathcal{V}_o$ ;  $u = 1.25$  for  $\mathcal{G}'_a$  and  $\mathcal{G}''_a$ ;  $u = 1.7$  for  $\mathcal{G}'''_a$ ;  $S_1(\cdot) = \tanh(\cdot)$ ,  $S_2(\cdot) = \frac{1}{2} \tanh(2 \cdot)$ .

6.1.i, in a social dissonance-driven Hopf bifurcation, beliefs of agents on  $\mathcal{G}'_a$  synchronize in phase along every option while beliefs of agents 1 and 2 on  $\mathcal{G}''_a$  oscillate out of phase from agents 3 and 4, reflecting the bipartition. Sample simulated trajectories are shown in Fig. 7(b) for option 1. Meanwhile,  $\mathcal{G}'''_a$  is neither class I or II, and has an adjacency matrix with a complex conjugate pair of leading eigenvalues. A social dissonance-driven bifurcation on this graph is still a Hopf bifurcation; however, due to the fact that both  $\mathbf{v}_a$  and  $\mathbf{v}_o$  of Theorem 6.1 are complex, beliefs do not synchronize in phase internally or between agents along any option. A sample trajectory for option 1 is shown in Fig. 7(b).

We next derive sufficient conditions for graphs  $\mathcal{G}_a, \mathcal{G}_o$  that either rule out oscillations altogether or guarantee oscillations when social dissonance is dominant in the belief dynamics (4).

*Corollary 6.2:* 1) [52, Proposition V.1] Suppose  $\mathcal{G}_a$  and  $\mathcal{G}_o$  are both undirected. Then the indecision-breaking bifurcation of Theorem 4.2 *cannot* be a Hopf bifurcation.

2) Suppose  $\mathcal{G}_a$  ( $\mathcal{G}_o$ ) belongs to Class I and/or Class II of Definition 1. Suppose  $\mathcal{G}_o$  ( $\mathcal{G}_a$ ) has an odd number of nodes and is switching isomorphic to a cycle graph  $\mathcal{G}$  with all-negative edge signatures, i.e.,  $\mathcal{G}$  has circulant adjacency matrix  $A$  with the first row  $(0, 0, \dots, 0, -1)$ , and every subsequent row generated by shifting the preceding row to the right by one element. Any social dissonance-driven indecision-breaking bifurcation in (4) on these graphs is a Hopf bifurcation.

*Proof:* 1) Due to the symmetry of the adjacency matrices, eigenvalues  $\lambda \in \sigma(A_a)$  and  $\mu \in \sigma(A_o)$  must be real-valued, which violates the necessary conditions in Assumption 1.

2) Without loss of generality, let  $\mathcal{G}_o$  be the circulant graph. By Lemma 2.2,  $A_a$  has a simple dominant eigenvalue  $\lambda$ . Since  $A$  is circulant, it has  $N = N_o$  eigenvalues  $\mu_m = -\exp(2\pi mi/N)$ ,  $m = 0, \dots, N-1$  [43, 2.2.P10]. A leading eigenvalue of  $A$ , and therefore of  $A_o$  since the two are co-spectral, has real part  $\lambda_{max} = \max_{m=0, \dots, N-1} \text{Re}(\mu_m) = \max_{m=0, \dots, N-1} -\cos(2\pi m/N)$ , which happens when  $m = (N-1)/2$  and  $m = (N+1)/2$  since  $N$  is odd. Observe that  $\mu_{(N-1)/2} = \bar{\mu}_{(N+1)/2}$ . The set of leading eigenvalue pairs is

then  $\Lambda_1 = \{(\lambda, \mu_{(N+1)/2}), (\lambda, \bar{\mu}_{(N+1)/2})\}$  which satisfies the necessary condition in Assumption 1. ■

## VII. DISCUSSION AND FINAL REMARKS

We have proposed a nonlinear, multi-dimensional belief formation model and shown how agents form strong beliefs as a consequence of an indecision-breaking bifurcation determined by the communication network and belief system network. We have described the multi-stable belief equilibria and belief oscillations that form through a pitchfork bifurcation and a Hopf bifurcation, respectively.

While these two outcomes of belief formation are most common across all graphs, other more complex bifurcations can occur in the model. For example, the set  $\Lambda_2$  described in parts iii and iv of Lemma 4.3 corresponds to two equal real-valued eigenvalues in the Jacobian  $J(0, \mathbf{u})$ . External dissonance-driven bifurcations on these graphs will lead to a different bifurcation than those described in Sections V and VI, and belief trajectories will settle on a 2-dimensional attractor. The dissertation [68] contains an example of graphs  $\mathcal{G}_a, \mathcal{G}_o$  on which the social dissonance-driven bifurcation results in belief trajectories settling on an ergodic torus in phase space. The analysis presented in [16] considers bifurcations with a belief system graph  $\mathcal{G}_o$  that is all-to-all connected and same-sign. In this case,  $\Lambda$  can consist of  $N_o$  eigenvalue pairs, and many new symmetric equilibria can emerge at the corresponding equivariant bifurcation. When symmetry is present in one of the graphs  $\mathcal{G}_a, \mathcal{G}_o$ , tools from equivariant bifurcation theory [69] can be used to classify the indecision-breaking bifurcation.

The proposed model can thus be understood to parametrize a set of belief-forming behaviors that are extremely rich but nevertheless interpretable with a small number of parameters that capture realistic effects. Our analyses provide novel insights about the dynamics of social systems. They also serve as building blocks, as in [5], [6], [51], for new classes of decentralized algorithms for decision-making, task allocation, and other types of control for engineered teams and networks.

## APPENDIX

### Alternative model formulations

The model (3) can be formulated alternatively as

$$\begin{aligned} \dot{z}_{ij} = & -d_i z_{ij} + u_i S \left( \alpha_i z_{ij} + \gamma_i \sum_{\substack{k=1 \\ k \neq i}}^{N_a} (A_a)_{ik} y_{ikj} \right) \\ & + \beta_i \sum_{\substack{l \neq j \\ l=1}}^{N_o} (A_o)_{jl} z_{il} + \delta_i \sum_{l=1}^{N_o} \sum_{\substack{k=1 \\ k \neq i}}^{N_a} (A_o)_{jl} (A_a)_{ik} y_{ikl} \Big) + b_{ij} \end{aligned} \quad (11)$$

where  $S : \mathbb{R} \rightarrow [-k_1, k_2]$  with  $k_1, k_2 > 0$  is a saturating function  $S(y) = \hat{S}(y) + g(y)$  satisfying  $S(0) = 0$ ,  $\hat{S}(-y) = -\hat{S}(y)$ ,  $\hat{S}'(0) = 1$ ,  $g(0) = 0$ ,  $g'(0) = 0$ ,  $g(-y) \neq -g_m(y)$ . In the alternative model (11) agents process all dissonance terms together, whereas in the model (3) saturation is applied to social information along each option separately.

Another reasonable assumption may be that belief states are saturated individually, limiting the influence of any single

strong belief. This is captured in a third model formulation:

$$\begin{aligned} \dot{z}_{ij} = & -d_i z_{ij} + u_i \left( \alpha_i S(z_{ij}) + \gamma_i \sum_{\substack{k=1 \\ k \neq i}}^{N_a} (A_a)_{ik} S(y_{ikj}) \right) \\ & + \beta_i \sum_{\substack{l \neq j \\ l=1}}^{N_o} (A_o)_{jl} S(z_{il}) + \delta_i \sum_{l=1}^{N_o} \sum_{\substack{k=1 \\ k \neq i}}^{N_a} (A_o)_{jl} (A_a)_{ik} S(y_{ikl}) \Big) + b_{ij}. \end{aligned} \quad (12)$$

Due to the local and topological nature of the bifurcation analysis in this paper, all of the results proved for (3) apply directly to (11) and (12). The three models can be used interchangeably, and selected based on the interpretation of nonlinear processing that best fits an application context. The magnitudes of opinions individual agents arrived at, given the same set of model parameters, will differ between (3), (11), and (12) by some nonlinear scaling; however, the local topological features of the flow near the indecision-breaking bifurcation point will be shared by the three models.

### Specialization to one topic

All of the results in this paper apply to the special case  $N_o = 1$ , in which agents form scalar beliefs  $\mathbf{Z}_i = z_{i1} \in \mathbb{R}$  and the belief system graph  $\mathcal{G}_o$  is trivial. We denote  $z_{i1}$  by  $z_i$ ,  $y_{ik1}$  by  $y_{ik}$ , and  $b_{i1}$  by  $b_i$ ; the dynamics (3) are

$$\dot{z}_i = -d z_i + u S_1 \left( \alpha z_i + \gamma \sum_{\substack{k=1 \\ k \neq i}}^{N_a} (A_a)_{ik} y_{ik} \right) + b_i. \quad (13)$$

The results in Theorems 4.2, 5.1, 6.1, and their various corollaries specialize to (13) with perfect observation  $y_{ik} = z_k$  with the substitutions  $A_o = \mathbf{v}_o = \mathbf{w}_o = \mu = 1$ ,  $\beta = \delta = 0$ .

### Proof of Lemma 4.3

i) By Lemma 2.2,  $\lambda_{max} = \rho(A_a)$  ( $\mu_{max} = \rho(A_o)$ ) is a simple dominant eigenvalue of  $A_a$  ( $A_o$ ), where  $\text{Re}(\lambda) < \lambda_{max}$  for any  $\lambda \in \sigma(A_a) \setminus \{\lambda_{max}\}$  and  $\text{Re}(\mu) < \mu_{max}$  for any  $\mu \in \sigma(A_o) \setminus \{\mu_{max}\}$ . The rest follows by the definition of  $\Lambda_1$ .

ii) By Lemma 2.2,  $|\lambda_i| < \lambda_{max}$  for any  $\lambda_i \in \sigma(A_a) \setminus \{\lambda_{max}\}$  and  $|\mu_j| < \mu_{max}$  for any  $\mu_j \in \sigma(A_o) \setminus \{\mu_{max}\}$ . Then  $|\lambda_i \mu_j| = |\lambda_i| |\mu_j| < \lambda_{max} \mu_{max}$  unless  $\lambda_i = \lambda_{max}$  and  $\mu_j = \mu_{max}$ . The proposition follows by the definition of  $\Lambda_2$ .

iii) For undirected  $\mathcal{G}_a, \mathcal{G}_o$ , the matrices  $A_a, A_o$  are symmetric with real-valued spectra. By Lemma 2.2,  $\lambda_{max} = \rho(A_a)$  ( $\mu_{max} = \rho(A_o)$ ) is a simple dominant eigenvalue of  $A_a$  ( $A_o$ ). Sign-symmetric graphs have symmetric spectra by [45, Theorem 3.2], which implies  $\lambda_{min} = -\lambda_{max}$  is a simple eigenvalue of  $A_a$  and  $\mu_{min} = -\mu_{max}$  is a simple eigenvalue of  $A_o$ . Suppose  $\lambda_i \in \sigma(A_a)$ ,  $\mu_j \in \sigma(A_o)$ . Then  $|\lambda_i \mu_j| = |\lambda_i| |\mu_j| < \rho(A_a) \rho(A_o)$  unless  $\lambda_i = \pm \lambda_{max}$  and  $\mu_j = \pm \mu_{max}$ . The proposition follows by the definition of  $\Lambda_2$ .

iv) For undirected  $\mathcal{G}_a, \mathcal{G}_o$ , the matrices  $A_a$  and  $A_o$  are symmetric and their spectra are real. By Lemma 2.2,  $\lambda_{max} = \rho(A_a)$  ( $\mu_{max} = \rho(A_o)$ ) is a simple dominant eigenvalue of  $A_a$  ( $A_o$ ). The spectra of  $\sigma(A_a)$ ,  $\sigma(A_o)$  are symmetric if and only if both matrices have characteristic polynomial with no even-degree terms [70, Theorem 3.1]. When this is satisfied, by the same argument as presented for statement iii),

$\Lambda_2 = \{(\lambda_{max}, \mu_{max}), (-\lambda_{max}, -\mu_{max})\}$ . Suppose  $\mathcal{G}_a$  does not satisfy the stated algebraic condition. Then  $|\lambda_i| < \lambda_{max}$  for any  $\lambda_i \in \sigma(A_a) \setminus \{\lambda_{max}\}$  and  $|\lambda_i \mu_j| = |\lambda_i| |\mu_j| < \rho(A_a) \rho(A_o)$  unless  $\lambda_i = \lambda_{max}$  and  $\mu_j = \mu_{max}$ . An analogous argument holds if  $\mathcal{G}_o$  does not satisfy the algebraic condition.

### Proof of Proposition 4.5

1) Suppose  $\lambda, \tilde{\lambda} \in \sigma(A_a)$ ,  $\mu, \tilde{\mu} \in \sigma(A_o)$ ,  $\text{Re}(\lambda) = \lambda_{max}$ ,  $\text{Re}(\mu) = \mu_{max}$  and at least one of the following holds:  $\text{Re}(\tilde{\lambda}) < \lambda_{max}$ ,  $\text{Re}(\tilde{\mu}) < \mu_{max}$ , i.e.  $(\lambda, \mu) \in \Lambda_1$  and  $(\tilde{\lambda}, \tilde{\mu}) \notin \Lambda_1$ . Then  $\text{Re}(\eta(u, \lambda, \mu)) - \text{Re}(\eta(u, \tilde{\lambda}, \tilde{\mu})) = u\gamma(\lambda_{max} - \text{Re}(\tilde{\lambda})) + u\beta(\mu_{max} - \text{Re}(\tilde{\mu})) + u\delta(\text{Re}(\lambda\mu) - \text{Re}(\tilde{\lambda}\tilde{\mu})) > 0$  whenever  $\gamma(\lambda_{max} - \text{Re}(\tilde{\lambda})) + \beta(\mu_{max} - \text{Re}(\tilde{\mu})) > -\delta(\text{Re}(\lambda\mu) - \text{Re}(\tilde{\lambda}\tilde{\mu}))$ . Let  $Q(\gamma, \beta, \lambda, \tilde{\lambda}, \mu, \tilde{\mu})$  be the left-hand side of this inequality and define  $K := \min_{\lambda, \tilde{\lambda} \in \sigma(A_a), \mu, \tilde{\mu} \in \sigma(A_o)} (\text{Re}(\lambda\mu) - \text{Re}(\tilde{\lambda}\tilde{\mu})) < 0$ . Observe that  $\frac{\partial Q}{\partial \gamma} = \lambda_{max} - \text{Re}(\tilde{\lambda}) \geq 0$  and  $\frac{\partial Q}{\partial \beta} = \mu_{max} - \text{Re}(\tilde{\mu}) \geq 0$ , with at least one of the two inequalities being strict. Therefore increasing the magnitude of  $\gamma$  and/or of  $\beta$  will eventually make the inequality  $Q(\gamma, \beta, \lambda, \tilde{\lambda}, \mu, \tilde{\mu}) > -\delta K := K_c > 0$  valid for any choice of  $(\lambda, \tilde{\lambda}) \in \sigma(A_a) \times \sigma(A_o) \setminus \Lambda_1$ . Furthermore, suppose  $(\hat{\lambda}, \hat{\mu}) \in \sigma(A_a) \times \sigma(A_o)$ ,  $\text{Re}(\hat{\lambda}) = \lambda_2$ , and  $\text{Re}(\hat{\mu}) = \mu_2$ . For any  $(\tilde{\lambda}, \tilde{\mu}) \in \sigma(A_a) \times \sigma(A_o) \setminus \Lambda_1$ ,  $Q(\gamma, \beta, \lambda, \tilde{\lambda}, \mu, \tilde{\mu}) \geq Q(\gamma, \beta, \hat{\lambda}, \hat{\mu})$ . Then the inequality  $Q(\gamma, \beta, \lambda, \tilde{\lambda}, \mu, \tilde{\mu}) > K_c$  implies that  $\text{Re}(\eta(u, \lambda, \mu)) > \text{Re}(\eta(u, \tilde{\lambda}, \tilde{\mu}))$  for any  $(\tilde{\lambda}, \tilde{\mu}) \in \sigma(A_a) \times \sigma(A_o) \setminus \Lambda_1$ . This shows  $(\lambda, \mu) \in \Lambda_1 \implies (\lambda, \mu) \in \Lambda$ .

Now suppose  $(\lambda, \mu) \in \Lambda$  and  $Q(\gamma, \beta, \hat{\lambda}, \hat{\mu}) > K_c > 0$  for all  $(\tilde{\lambda}, \tilde{\mu}) \in \sigma(A_a) \times \sigma(A_o) \setminus \Lambda_1$ . Then by the definition of the set  $\Lambda$ ,  $\text{Re}(\eta(\lambda, \mu, u)) - \text{Re}(\eta(\tilde{\lambda}, \tilde{\mu}, u)) \geq 0$  for any  $(\tilde{\lambda}, \tilde{\mu}) \in \sigma(A_a) \times \sigma(A_o)$ , with equality holding if and only if  $(\tilde{\lambda}, \tilde{\mu}) \in \Lambda$ . Suppose  $\text{Re}(\tilde{\lambda}) = \lambda_{max}$ ,  $\text{Re}(\tilde{\mu}) = \mu_{max}$ . Then  $\text{Re}(\eta(u, \lambda, \mu)) - \text{Re}(\eta(u, \tilde{\lambda}, \tilde{\mu})) = u\gamma(\text{Re}(\lambda) - \lambda_{max}) + u\beta(\text{Re}(\mu) - \mu_{max}) + u\delta(\text{Re}(\lambda\mu) - \text{Re}(\tilde{\lambda}\tilde{\mu})) \geq 0$  which means  $-Q(\gamma, \beta, \lambda, \mu) \geq -\delta(\text{Re}(\lambda\mu) - \text{Re}(\tilde{\lambda}\tilde{\mu}))$ . If  $(\lambda, \mu) \notin \Lambda_1$ , then  $-K_c > -Q(\gamma, \beta, \lambda, \mu) > -\delta(\text{Re}(\lambda\mu) - \text{Re}(\tilde{\lambda}\tilde{\mu})) \implies K_c < -\delta(\text{Re}(\tilde{\lambda}\tilde{\mu}) - \text{Re}(\lambda\mu))$ . This is a contradiction of the definition  $K_c = \max_{\lambda, \tilde{\lambda} \in \sigma(A_a), \mu, \tilde{\mu} \in \sigma(A_o)} -\delta(\text{Re}(\lambda\mu) - \text{Re}(\tilde{\lambda}\tilde{\mu}))$ . Therefore  $(\lambda, \mu) \in \Lambda_1$  and the proposition follows.

2) Suppose  $\lambda, \tilde{\lambda} \in \sigma(A_a)$ ,  $\mu, \tilde{\mu} \in \sigma(A_o)$ ,  $\text{Re}(\lambda\mu) = (\lambda\mu)_{max}$ ,  $\text{Re}(\tilde{\lambda}\tilde{\mu}) < (\lambda\mu)_{max}$ , i.e.  $(\lambda, \mu) \in \Lambda_2$  and  $(\tilde{\lambda}, \tilde{\mu}) \notin \Lambda_2$ . Then  $\text{Re}(\eta(u, \lambda, \mu)) - \text{Re}(\eta(u, \tilde{\lambda}, \tilde{\mu})) = u\gamma(\text{Re}(\lambda) - \text{Re}(\tilde{\lambda})) + u\beta(\text{Re}(\mu) - \text{Re}(\tilde{\mu})) + u\delta((\lambda\mu)_{max} - \text{Re}(\tilde{\lambda}\tilde{\mu})) > 0$  whenever

$$\delta > \frac{-\beta(\text{Re}(\mu) - \text{Re}(\tilde{\mu})) - \gamma(\text{Re}(\lambda) - \text{Re}(\tilde{\lambda}))}{(\lambda\mu)_{max} - \text{Re}(\tilde{\lambda}\tilde{\mu})}. \quad (14)$$

Define the right-hand side of (14) to be  $h(\lambda, \mu, \tilde{\lambda}, \tilde{\mu})$  and let  $\delta_c := \max_{\lambda, \tilde{\lambda} \in \sigma(A_a), \mu, \tilde{\mu} \in \sigma(A_o), (\tilde{\lambda}, \tilde{\mu}) \notin \Lambda_2} h(\lambda, \mu, \tilde{\lambda}, \tilde{\mu})$ . Then assuming  $\delta > \delta_c$ ,  $\text{Re}(\eta(u, \lambda, \mu)) > \text{Re}(\eta(u, \tilde{\lambda}, \tilde{\mu}))$  for any  $(\tilde{\lambda}, \tilde{\mu}) \notin \Lambda_2$ , which means  $(\lambda, \mu) \in \Lambda_1$  implies  $(\lambda, \mu) \in \Lambda$ .

Next, suppose  $\delta > \delta_c$ ,  $(\lambda, \mu) \in \Lambda$ , and  $\text{Re}(\lambda\mu) < (\lambda\mu)_{max}$ . Then since  $\delta > \delta_c$ , for any eigenvalue pair  $\hat{\lambda} \in \sigma(A_a)$ ,  $\hat{\mu} \in \sigma(A_o)$  with  $\text{Re}(\hat{\lambda}\hat{\mu}) = (\lambda\mu)_{max}$  it must hold that  $\text{Re}(\eta(\hat{\lambda}, \hat{\mu}, u)) > \text{Re}(\eta(\lambda, \mu, u))$ . However, by definition of set  $\Lambda$  (8), it must hold that  $\text{Re}(\eta(\lambda, \mu, u)) \geq \text{Re}(\eta(\hat{\lambda}, \hat{\mu}, u))$ , which

is a contradiction. Therefore, whenever  $\gamma > \gamma_c$  and  $(\lambda, \mu) \in \Lambda$ , it must hold that  $\text{Re}(\lambda\mu) = (\lambda\mu)_{max}$ , i.e.  $(\lambda, \mu) \in \Lambda_2$ .

### Proof of Theorem 5.1

To identify the pitchfork bifurcation in the model (4) we rely on the singularity theory of bifurcations [63]. For a concise overview of the main ideas of this approach we refer the reader to Chapter 2.3 of the doctoral dissertation [68].

1) From Theorem 4.2 we know  $\mathbf{Z} = \mathbf{0}$  is stable for  $u < u^*$  and unstable for  $u > u^*$ ; when  $u = u^*$ , a bifurcation happens along a manifold tangent to  $\text{span}\{\mathbf{v}_a \otimes \mathbf{v}_o\}$ . To identify this bifurcation we compute a polynomial expansion of a Lyapunov-Schmidt (LS) reduction  $f(y, u, \mathbf{b})$  of (4) following the procedure outlined in [63, Chapter I]. We derive the coefficients through third order. With the odd symmetry assumption on  $S_m$ ,  $S'_m(0) = 0$  and a LS reduction approximation reads  $f(y, u, \mathbf{0}) = K_0(u - u^*)y + dK_1/K_0 y^3$ . To establish a pitchfork bifurcation, we verify that this LS reduction satisfies the necessary and sufficient conditions outlined in [63, Chapter II, Proposition 9.2]:  $f(0, u^*, \mathbf{0}) = \frac{\partial f}{\partial y}(0, u^*, \mathbf{0}) = \frac{\partial^2 f}{\partial y^2}(0, u^*, \mathbf{0}) = \frac{\partial f}{\partial u}(0, u^*, \mathbf{0}) = 0$ ,  $\frac{\partial^2 f}{\partial y \partial u}(0, u^*, \mathbf{0}) \neq 0$ ,  $\frac{\partial^2 f}{\partial y^3}(0, u^*, \mathbf{0}) \neq 0$ . Linearization of the system about each of the two bifurcating fixed points shares the  $N - 1$  stable eigenvalues with the origin, and the bifurcating eigenvalue is negative (or positive) under the conditions in the theorem statement by [63, Chapter I, Theorem 4.1].

2) and 3) With nonzero  $\mathbf{b}$ ,  $g_1, g_2$ , the coefficients of the polynomial expansion of  $f(y, u, \mathbf{b})$  through third order take the form  $f(y, u, \mathbf{b}) = K_0(u - u^*)y + K_b + K_3y^3 + dK_2/K_0y^2$  where  $K_b = \langle \mathbf{w}_a \otimes \mathbf{w}_o, \mathbf{b} \rangle$  and  $K_3 = dK_1/K_0 - 3\langle \mathbf{w}_a \otimes \mathbf{w}_o, (u^*)^2(g''_1(0)(\alpha + \gamma\lambda)^2 + g''_2(0)\mu^2(\beta + \delta\lambda)^2)(g'_1(0)(\alpha + \gamma\lambda)\mathbf{v}_a \otimes \mathbf{v}_o \odot ((\alpha\mathcal{I}_{N_a} + \gamma A_a) \otimes \mathcal{I}_{N_o})(J^{-1}P(\mathbf{v}_a \otimes \mathbf{v}_o)^2) + g'_2(0)\mu(\beta + \delta\lambda)\mathbf{v}_a \otimes \mathbf{v}_o \odot ((\beta\mathcal{I}_{N_a} + \delta A_a) \otimes A_o)(J^{-1}P(\mathbf{v}_a \otimes \mathbf{v}_o)^2)) \rangle$  where  $J^{-1}$  is the inverse of the restriction of  $J(\mathbf{0}, u^*)$  to its range, and  $P = \mathcal{I}_{N_a N_o} - \frac{1}{\|\mathbf{v}_a \otimes \mathbf{v}_o\|^2}(\mathbf{v}_a \otimes \mathbf{v}_o)(\mathbf{v}_a \otimes \mathbf{v}_o)^T$  is a projection onto the range of  $J(\mathbf{0}, u^*)$ . Observe that whenever  $g'_1(0) = 0$  and  $g'_2(0) = 0$ , the cubic coefficient  $K_3 = dK_1/K_0$ . Therefore by continuity there exist small positive  $\varepsilon_{g1}, \varepsilon_{g2}$  such that  $\text{sign}(K_3) = \text{sign}(K_1)$  whenever  $|g'_1(0)| < \varepsilon_{g1}$  and  $|g'_2(0)| < \varepsilon_{g2}$ . Observe that  $K_2, K_b$  are small. The bifurcation diagram  $f(y, u, \mathbf{b}) = 0$  is an unfolding (i.e. a perturbation) of a symmetric pitchfork bifurcation diagram, and the characterization of equilibria as a function of  $K_b, K_2$  follow from standard analysis of a pitchfork bifurcation universal unfolding [63, Chapter I §1].

### Proof of Theorem 6.1

This proof is mostly verbatim from [52, Theorem IV.3], with the argument generalized slightly to accommodate asymmetric  $S_1, S_2$ . Let  $C = \alpha + \gamma \text{Re}(\lambda^\dagger) + \beta \text{Re}(\mu^\dagger) + \delta \text{Re}(\lambda^\dagger \mu^\dagger)$ .

1) To establish existence of periodic orbits we check that the system (4) under the stated assumptions satisfies the conditions of the Hopf bifurcation theorem [61, Theorem 3.4.2]. When  $u = u^* = d/C$ , the leading eigenvalues of (6) are a simple purely imaginary pair  $\eta_\pm(u^*) = \pm iu^*|\gamma \text{Im}(\lambda^\dagger) + \beta \text{Im}(\mu^\dagger) + \delta \text{Im}(\lambda^\dagger \mu^\dagger)| \neq 0$ , which satisfies the eigenvalue condition



(H1) of the Hopf theorem. Next, we check that the leading eigenvalues cross the imaginary axis with nonzero speed as  $u$  is varied, i.e.  $\frac{d}{du} \operatorname{Re}(\eta_{\pm}(u)) = C > 0$ , which satisfies the nonzero crossing speed condition (H2) of the Hopf theorem. Existence of periodic orbits directly follows by the Hopf theorem. By this theorem and by the definition of a center manifold [61, Theorem 3.2.1], the solutions appear along a unique  $W^s$  which is tangent at  $u = u^*$  to  $\operatorname{Ker}(J(\mathbf{0}, u^*)) = \operatorname{span}\{\operatorname{Re}(\mathbf{v}_a \otimes \mathbf{v}_o), \operatorname{Im}(\mathbf{v}_a \otimes \mathbf{v}_o)\}$ .

To show 2) and 3) we first compute a third-order approximation of (4) following the Lyapunov-Schmidt reduction for a Hopf bifurcation [63, Chapter VIII, Proposition 3.3]:  $f(y, u) = Cy(u - u^*) + \frac{1}{16}u^*K_3y^3$ , where  $K_3 = K + K_g(g_1''(0), g_2''(0))$  and  $K_g(0, 0) = 0$ . As long as  $K_3 \neq 0$ , by [63, Chapter VIII, Theorems 2.1 and 3.2] the reduced bifurcation equation  $f(y, u)$  possesses a pitchfork bifurcation which is supercritical for  $K_3 < 0$  and subcritical for  $K_3 > 0$ . By continuity there exist small positive  $\varepsilon_{g1}, \varepsilon_{g2}$  such that  $\operatorname{sign}(K_3) = \operatorname{sign}(K)$  whenever  $|g_1''(0)| < \varepsilon_{g1}$  and  $|g_2''(0)| < \varepsilon_{g2}$  from which the theorem statement follows.

When  $|u - u^*|$  is small, solutions to  $f(y, u) = 0$  are in one-to-one correspondence with orbits of small amplitude periodic solutions to the system (4) with period near  $2\pi/(u^*|\gamma \operatorname{Im}(\lambda^\dagger) + \beta \operatorname{Im}(\mu^\dagger) + \delta \operatorname{Im}(\lambda^\dagger \mu^\dagger)|) := 1/\omega$ . For  $u$  near  $u^*$ , the small amplitude oscillations can be approximated to first order as scalar multiples of  $e^{i\omega t} \mathbf{v}_a \otimes \mathbf{v}_o$  from which the conclusions on phase and amplitude difference between agents follow. When  $K_3 < 0$  ( $> 0$ ), and therefore by the same continuity argument as above when  $K < 0$  ( $> 0$ ), the bifurcating periodic solutions are stable (unstable) as established in [63, Chapter VIII, Theorem 4.1].

## REFERENCES

- [1] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, 2004.
- [2] M. A. Montes de Oca, E. Ferrante, N. Mathews, M. Birattari, and M. Dorigo, "Opinion dynamics for decentralized decision-making in a robot swarm," in *Swarm Intelligence. ANTS 2010.*, vol. 6234. Springer, 2010, pp. 251–262.
- [3] M. A. Montes de Oca, E. Ferrante, A. Scheidler, C. Pinciroli, M. Birattari, and M. Dorigo, "Majority-rule opinion dynamics with differential latency: a mechanism for self-organized collective decision-making," *Swarm Intelligence*, vol. 5, pp. 305–327, 2011.
- [4] C. Altafini, "Consensus problems on networks with antagonistic interactions," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 935–946, 2012.
- [5] H. Hu, K. Nakamura, K.-C. Hsu, N. E. Leonard, and J. F. Fisac, "Emergent coordination through game-induced nonlinear opinion dynamics," *To appear Proc. IEEE CDC; arXiv:2304.02687 [eess.SY]*, 2023.
- [6] C. Cathcart, M. Santos, S. Park, and N. E. Leonard, "Proactive opinion-driven robot navigation around human movers," *To appear Proc. IROS; arXiv:2210.01642 [cs.RO]*, 2023.
- [7] H. C. Leung, Z. Li, B. She, and P. E. Paré, "Leveraging opinions and vaccination to eradicate networked epidemics," in *2023 European Control Conference*, 2023, pp. 1–6.
- [8] M. Galesic, W. Bruine de Bruin, J. Dalege, S. L. Feld, F. Kreuter, H. Olsson, D. Prelec, D. L. Stein, and T. van Der Does, "Human social sensing is an untapped resource for computational social science," *Nature*, vol. 595, no. 7866, pp. 214–222, 2021.
- [9] J. Dalege, M. Galesic, and H. Olsson, "Networks of beliefs: An integrative theory of individual-and social-level belief dynamics," *OSF Preprints*, 2023.
- [10] M. Vlasceanu, A. M. Dyckovsky, and A. Coman, "A network approach to investigate the dynamics of individual and collective beliefs: Advances and applications of the bending model," *Perspect. Psychol. Sci.*, 2023.
- [11] A. Strandburg-Peshkin, C. R. Twomey, N. W. Bode, A. B. Kao, Y. Katz, C. C. Ioannou, S. B. Rosenthal, C. J. Torney, H. S. Wu, S. A. Levin *et al.*, "Visual sensory networks and effective information transfer in animal groups," *Current Biology*, vol. 23, no. 17, pp. R709–R711, 2013.
- [12] S. B. Rosenthal, C. R. Twomey, A. T. Hartnett, H. S. Wu, and I. D. Couzin, "Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 112, no. 15, pp. 4690–4695, 2015.
- [13] V. H. Sridhar, L. Li, D. Gorboson, M. Nagy, B. R. Schell, T. Sorochnik, N. S. Gov, and I. D. Couzin, "The geometry of decision-making in individuals and collectives," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 118, no. 50, 2021.
- [14] L. Oscar, L. Li, D. Gorboson, I. Couzin, and N. Gov, "A simple cognitive model explains movement decisions during schooling in zebrafish," *bioRxiv*, pp. 2023–02, 2023.
- [15] A. Bizyaeva, A. Franci, and N. E. Leonard, "Nonlinear opinion dynamics with tunable sensitivity," *IEEE Trans. Autom. Control*, vol. 68, no. 3, pp. 1415–1430, 2022.
- [16] A. Franci, M. Golubitsky, I. Stewart, A. Bizyaeva, and N. E. Leonard, "Breaking indecision in multiagent, multioption dynamics," *SIAM J. Appl. Dynamical Syst.*, vol. 22, no. 3, pp. 1780–1817, 2023.
- [17] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.
- [18] N. E. Friedkin, A. V. Proskurnikov, R. Tempo, and S. E. Parsegov, "Network science on belief system dynamics under logic constraints," *Science*, vol. 354, no. 6310, pp. 321–326, 2016.
- [19] S. E. Parsegov, A. V. Proskurnikov, R. Tempo, and N. E. Friedkin, "Novel multidimensional models of opinion dynamics in social networks," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2270–2285, 2016.
- [20] M. Ye, J. Liu, L. Wang, B. D. Anderson, and M. Cao, "Consensus and disagreement of heterogeneous belief systems in influence networks," *IEEE Trans. Autom. Control*, vol. 65, no. 11, pp. 4679–4694, 2019.
- [21] G. He, Z. Ci, X. Wu, and M. Hu, "Opinion dynamics with antagonistic relationship and multiple interdependent topics," *IEEE Access*, vol. 10, pp. 31 595–31 606, 2022.
- [22] L. Pan, H. Shao, M. Mesbahi, Y. Xi, and D. Li, "Bipartite consensus on matrix-valued weighted networks," *IEEE Trans. Circuits Syst. II*, vol. 66, no. 8, pp. 1441–1445, 2018.
- [23] A. Nedić, A. Olshevsky, and C. A. Uribe, "Graph-theoretic analysis of belief system dynamics under logic constraints," *Sci. Rep.*, vol. 9, no. 1, p. 8843, 2019.
- [24] M. Ye, M. H. Trinh, Y.-H. Lim, B. D. Anderson, and H.-S. Ahn, "Continuous-time opinion dynamics on multiple interdependent topics," *Automatica*, vol. 115, p. 108884, 2020.
- [25] H.-S. Ahn, Q. Van Tran, M. H. Trinh, M. Ye, J. Liu, and K. L. Moore, "Opinion dynamics with cross-coupling topics: Modeling and analysis," *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 3, pp. 632–647, 2020.
- [26] C. Wang, L. Pan, H. Shao, D. Li, and Y. Xi, "Characterizing bipartite consensus on signed matrix-weighted networks via balancing set," *Automatica*, vol. 141, p. 110237, 2022.
- [27] J. Lorenz, "Continuous opinion dynamics under bounded confidence: A survey," *Int. J. Mod. Phys. C*, vol. 18, no. 12, pp. 1819–1838, 2007.
- [28] P. Dandekar, A. Goel, and D. T. Lee, "Biased assimilation, homophily, and the dynamics of polarization," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 15, pp. 5791–5796, 2013.
- [29] R. Gray, A. Franci, V. Srivastava, and N. E. Leonard, "Multiagent decision-making dynamics inspired by honeybees," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 2, pp. 793–806, 2018.
- [30] W. Xia, M. Ye, J. Liu, M. Cao, and X.-M. Sun, "Analysis of a nonlinear opinion dynamics model with biased assimilation," *Automatica*, vol. 120, p. 109113, 2020.
- [31] L. Wang, Y. Hong, G. Shi, and C. Altafini, "Signed social networks with biased assimilation," *IEEE Trans. Autom. Control*, vol. 67, no. 10, pp. 5134–5149, 2022.
- [32] W. Mei, F. Bullo, G. Chen, J. M. Hendrickx, and F. Dörfler, "Micro-foundation of opinion dynamics: Rich consequences of the weighted-median mechanism," *Phys. Rev. Res.*, vol. 4, no. 2, 2022.
- [33] M. Usher and J. L. McClelland, "The time course of perceptual choice: The leaky, competing accumulator model," *Psychol. Rev.*, vol. 108, no. 3, pp. 550–592, 2001.
- [34] R. Bogacz, M. Usher, J. Zhang, and J. L. McClelland, "Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice," *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, vol. 362, no. 1485, pp. 1655–1670, 2007.
- [35] J. Introne, "Measuring belief dynamics on twitter," in *Proc. of the Int. AAAI Conf. on Web and Social Media*, vol. 17, 2023, pp. 387–398.

- [36] N. Rodriguez, J. Bollen, and Y.-Y. Ahn, "Collective dynamics of belief evolution under cognitive coherence and social conformity," *PLoS one*, vol. 11, no. 11, 2016.
- [37] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, "Modeling echo chambers and polarization dynamics in social networks," *Phys. Rev. Lett.*, vol. 124, no. 4, 2020.
- [38] M. J. Brandt and W. W. Sleegers, "Evaluating belief system networks as a theory of political belief system dynamics," *Personality and Social Psychology Review*, vol. 25, no. 2, pp. 159–185, 2021.
- [39] Ł. G. Gajewski, J. Sienkiewicz, and J. A. Holyst, "Transitions between polarization and radicalization in a temporal bilayer echo-chamber model," *Phys. Rev. E*, vol. 105, no. 2, p. 024125, 2022.
- [40] L. Li, A. Zeng, Y. Fan, and Z. Di, "Modeling multi-opinion propagation in complex systems with heterogeneous relationships via Potts model on signed networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 32, no. 8, 2022.
- [41] J. Ojer, M. Starnini, and R. Pastor-Satorras, "Modeling explosive opinion depolarization in interdependent topics," *Phys. Rev. Lett.*, vol. 130, no. 20, p. 207401, 2023.
- [42] D. Noutsos, "On Perron–Frobenius property of matrices having some negative entries," *Linear Algebra Appl.*, vol. 412, no. 2-3, pp. 132–153, 2006.
- [43] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 2012.
- [44] T. Zaslavsky, "Matrices in the theory of signed simple graphs," *Proc. ICDM 2008, RMS-Lecture Notes Series*, no. 13, pp. 207–229, 2010.
- [45] F. Belardo, S. M. Cioabă, J. H. Koolen, and J. Wang, "Open problems in the spectral theory of signed graphs," *Art Discrete Appl. Math.*, vol. 1, no. P2.10, pp. 1–23, 2018.
- [46] T. Zaslavsky, "Signed graphs," *Discrete Appl. Math.*, vol. 4, no. 1, pp. 47–74, 1982.
- [47] E. Ghorbani, W. H. Haemers, H. R. Maimani, and L. P. Majd, "On sign-symmetric signed graphs," *arXiv preprint arXiv:2003.09981*, 2020.
- [48] C. Altafini and G. Lini, "Predictable dynamics of opinion forming for networks with antagonistic interactions," *IEEE Trans. Autom. Control*, vol. 60, no. 2, pp. 342–357, 2014.
- [49] Y. Jiang, H. Zhang, and J. Chen, "Sign-consensus of linear multi-agent systems over signed directed graphs," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 5075–5083, 2016.
- [50] A. Fontan and C. Altafini, "The role of frustration in collective decision-making dynamical processes on multiagent signed networks," *IEEE Trans. Autom. Control*, vol. 67, no. 10, pp. 5191–5206, 2021.
- [51] A. Bizyaeva, G. Amorim, M. Santos, A. Franci, and N. E. Leonard, "Switching transformations for decentralized control of opinion patterns in signed networks: application to dynamic task allocation," *IEEE Control Syst. Lett.*, vol. 6, pp. 3463–3468, 2022.
- [52] A. Bizyaeva, A. Franci, and N. E. Leonard, "Sustained oscillations in multi-topic belief dynamics over signed networks," in *Proc. ACC*. IEEE, 2023, pp. 4296–4301.
- [53] L. Wang, Y. Hong, G. Shi, and C. Altafini, "A biased assimilation model on signed graphs," in *Proc. IEEE CDC*. IEEE, 2020, pp. 494–499.
- [54] P. E. Converse, "The nature of belief systems in mass publics (1964)," *Crit. Rev.*, vol. 18, no. 1-3, pp. 1–74, 2006.
- [55] M. Golubitsky and R. Lauterbach, "Bifurcations from synchrony in homogeneous networks: linear theory," *SIAM J. Appl. Dynamical Syst.*, vol. 8, no. 1, pp. 40–75, 2009.
- [56] I. Stewart and M. Golubitsky, "Synchrony-breaking bifurcation at a simple real eigenvalue for regular networks 1: 1-dimensional cells," *SIAM J. Appl. Dynamical Syst.*, vol. 10, no. 4, pp. 1404–1442, 2011.
- [57] I. Stewart, "Synchrony-breaking bifurcation at a simple real eigenvalue for regular networks 2: Higher-dimensional cells," *SIAM J. Appl. Dynamical Syst.*, vol. 13, no. 1, pp. 129–156, 2014.
- [58] E. Nijholt, B. Rink, and J. Sanders, "Center manifolds of coupled cell networks," *SIAM Rev.*, vol. 61, no. 1, pp. 121–155, 2019.
- [59] M. Golubitsky and I. Stewart, *Dynamics and Bifurcation in Networks*. SIAM, 2023.
- [60] H. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2000.
- [61] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York, NY: Springer-Verlag, 2013, vol. 42.
- [62] A. Dhooge, W. Govaerts, Y. Kuznetsov, H. Meijer, and B. Sautois, "New features of the software MatCont for bifurcation analysis of dynamical systems," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 14, no. 2, pp. 147–175, 2008.
- [63] M. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, ser. Applied Mathematical Sciences. New York, NY: Springer-Verlag, 1985, vol. 51.
- [64] E. L. Fink, S. Kaplowitz, G. Barnett, and W. Richards, "Oscillation in beliefs and cognitive networks," *Progress in Communication Sciences*, vol. 12, pp. 247–272, 1993.
- [65] E. L. Fink, S. A. Kaplowitz, and S. M. Hubbard, "Oscillation in beliefs and decisions," *The Persuasion Handbook: Developments in Theory and Practice*. Thousand Oaks, CA: Sage Publications, pp. 17–38, 2002.
- [66] B. D. Belleau, "Cyclical fashion movement: Women's day dresses: 1860-1980," *Cloth. Text. Res. J.*, vol. 5, no. 2, pp. 15–20, 1987.
- [67] J. Stimson, *Public Opinion in America: Moods, Cycles, and Swings*. Routledge, 2018.
- [68] A. S. Bizyaeva, "Nonlinear dynamics of multi-agent multi-option belief and opinion formation," Ph.D. dissertation, Princeton University, 2022.
- [69] M. Golubitsky and I. Stewart, *The Symmetry Perspective*, 1st ed., ser. Progress in Mathematics. New-York: Birkhäuser Basel, 2002, vol. 200.
- [70] D. Li and Q. Huang, "The signed graphs with symmetric spectra," *arXiv:2304.06864 [math.CO]*, 2023.