Networks of Beliefs: An Integrative Theory of Individual- and Social-Level Belief Dynamics

Jonas Dalege, Mirta Galesic, and Henrik Olsson

Santa Fe Institute

Author Note

Abstract

We present a novel theory of belief dynamics that explains the interplay between internal beliefs in people's minds and beliefs of others in their external social environments. The Networks of Belief theory goes beyond existing theories of belief dynamics in three ways. First, it provides an explicit connection between belief networks in individual minds and belief dynamics on social networks. The connection, absent from previous theories, is established through people's social beliefs, or perceived beliefs of others. Second, the theory recognizes that the correspondence between social beliefs and others' actual beliefs can be imperfect, because social beliefs are affected by personal beliefs as well as by the actual beliefs of others. Past theories of belief dynamics on social networks do not distinguish between perceived and actual beliefs of others. Third, the theory explains diverse belief dynamics phenomena parsimoniously through the differences in attention and the resulting felt dissonances in personal, social, and external parts of belief networks. We implement our theoretical assumptions in a computational model within a statistical physics framework and derive novel model predictions. We find support for our theoretical assumptions and model predictions in two large survey studies (N_1=973, N_2=669). We then derive novel insights about diverse phenomena related to belief dynamics, including group consensus and polarization, group radicalization, minority influence, and different empirically observed belief distributions. We discuss how the theory goes beyond different existing models of belief dynamics and outline promising directions for future research.

Networks of Beliefs: An Integrative Theory of Individual- and Social-Level Belief Dynamics

Why do people's beliefs sometimes become more extreme over time and in other instances stay quite moderate? When are these beliefs shaped by other related issues, and when are they shaped by one's social environment? Why do people sometimes project their personal beliefs onto others, and at other times perceive others' beliefs with high accuracy? How can a single theory of belief change answer these questions and explain a plethora of phenomena including polarization, radicalization, minority influence, and real-world belief patterns? These and other questions about belief formation and change have been investigated in a wide range of disciplines, from psychology (Ajzen, 1991; Latane & Wolf, 1981; Vallacher, Read, & Nowak, 2017; Dalege et al., 2018), sociology (Friedkin & Johnsen, 1990; Proskurnikov & Tempo, 2017), cultural evolution (Boyd & Richerson, 1985; Hoppitt & Laland, 2013), and economics (Acemoglu & Ozdaglar, 2011; Golub & Sadler, 2016), to statistical physics (Castellano et al., 2009; Pentland, 2014) and applied mathematics (Hickok et al., 2022). This large body of work has provided a plethora of empirical findings and theories about the structure, formation, and change of beliefs.

However, there is a disconnect between lines of work that study belief change at the level of a single individual mind (the internal level) and at the level of social networks (the external level). At the internal level, modeling individual belief networks has been a promising framework for explaining empirical findings on how and why people change their beliefs (Dalege et al., 2016; Monroe & Read, 2008; Shultz & Lepper,1996; Van Overwalle & Siebler, 2005). At the external level, belief change has been fruitfully studied within the context of social networks, whereby interactions with others' actual or perceived beliefs can influence one's own beliefs (Christakis & Fowler, 2010; DeGroot, 1974; Festinger, 1954; French, 1956; Friedkin &

Johnsen, 1990; Harary, 1959).

The internal and external level of beliefs have been studied largely in isolation from each other. Models of internal belief networks generally disregard external social networks, although a number of classic social-cognitive theories recognize the importance of social environments in changing individual beliefs and behaviors (Ajzen, 1991; Cialdini & Trost, 1998; Festinger, 1954; Fishbein & Ajzen, 1977; Petty & Cacioppo, 1986). On the other side, models of belief dynamics on external social networks, typically developed outside psychology in fields from sociology, economics, computational social science, to statistical physics (Albert & Barabasi, 2002; Easley & Kleinberg, 2010; Friedkin & Johnsen, 1990; Jackson, 2008; Newman, 2003; Watts, 2004) largely disregard the richness of individual belief structures and individual differences in perceptions of others' beliefs (Galesic, Bruine de Bruin, et al., 2021).

To acquire a comprehensive understanding of belief dynamics, we need to integrate findings and models at the internal and external levels, ideally using theoretically driven quantitative models that are simple but empirically testable (Galesic, Olsson, et al., 2021). In this paper, we develop such an integrative theory of belief dynamics, the Networks of Belief (NB) theory. We build on the Attitudinal Entropy (AE, Dalege et al., 2018) framework and the Hierarchical Ising Opinion Model (HIOM, van der Maas, Dalege, & Waldorp, 2020). Like these theories, we conceptualize beliefs as nodes in attitude networks within a statistical physics framework which enables rigorous investigations and predictions of belief dynamics.

The NB theory goes beyond past theories in three ways. First, it makes a direct connection between internal beliefs and external social worlds, through people's social beliefs or perceived beliefs of others. This is a crucial step that enables connecting models and findings on internal networks, largely developed in psychology, and belief dynamics models operating solely

on the level of external networks, of which multitudes exist in fields from computational social science to statistical physics.

Our second contribution is explaining how social beliefs can become more or less accurate representations of the actual beliefs of others, and how that affects the resulting belief dynamics. In this way we bridge the gap between "objectively" measured beliefs in social networks that are the major type of data used to study belief dynamics, and "subjective" perception of these beliefs that are what eventually matters for social influence (Thomas & Swaine Thomas, 1928).

Our third contribution is explicit modeling of attention to different sources of dissonance between these different beliefs in internal and external networks. We postulate and show empirically that differences in attention to different parts of the internal and external networks, and the resulting differences in felt dissonance regarding those parts, lead to changes in consistency of these network parts. This in turn leads to predictable patterns in people's beliefs and to well-established effects such as polarization, radicalization, and minority influence.

In what follows we first outline the main premises of the NB theory, describe the relationships between the central psycho-social constructs in the NB theory and statistical physics constructs, present our formal implementation of the theory together with illustrative simulations and predictions, and report empirical tests of our theoretical predictions. We then demonstrate that the NB theory can account for pervasive phenomena in the belief dynamics literature such as group polarization, group radicalization, and minority influence, as well as for empirical trends in real-real world data such as the patterns of political self-identification at a given moment and its increasing polarization over time. We end with a thorough discussion of how the NB theory contributes to the existing models of belief dynamics in different disciplines.

Main Premises

The Networks of Belief theory rests on three premises about the underlying psycho-social constructs. First, we assume that beliefs and their relations can be represented as network structures. Second, we assume that individuals aim to reduce dissonance within their belief networks and increase correspondence between perceived and actual beliefs of others around them. Third, dissonance between beliefs only affects belief change when it is *felt*, and this depends on whether individuals pay attention to their beliefs and their relationships.
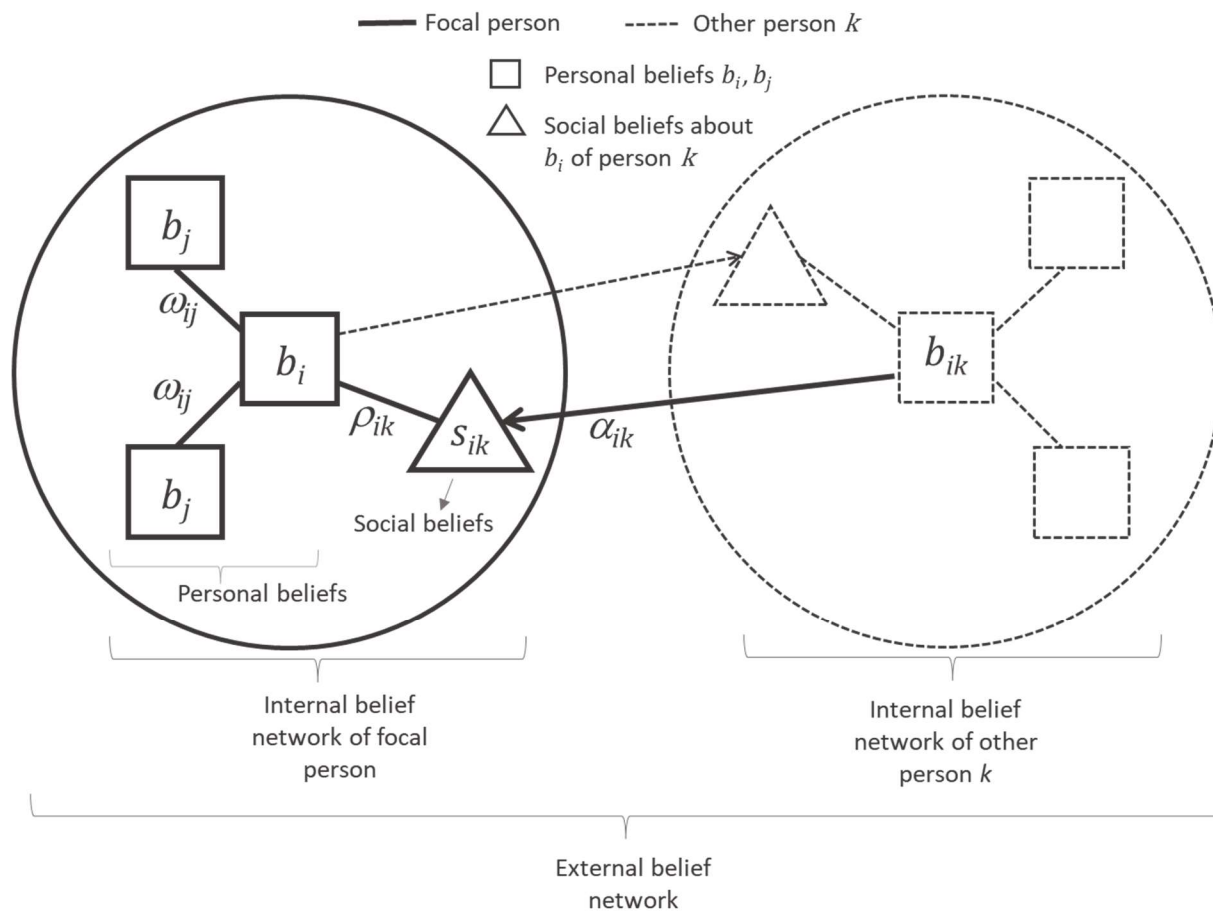
Belief Networks

The first main premise of the NB theory is that beliefs can be represented as nodes in a network, with edges representing mutual influence of beliefs that different individuals have about various issues and about each other. Going beyond previous theories, the NB theory assumes that beliefs of interconnected individuals form two distinct classes of networks that interact with each other (see Figure 1).

One network represents each individual's *internal network*, which consists of *personal beliefs* and *social beliefs*. *Personal beliefs* represent beliefs related to an issue, such as various facts, preferences, and more general moral and other values that one considers relevant (e.g., whether one believes that one should get vaccinated might depend on one's beliefs about the efficacy and safety of the vaccine and whether getting vaccinated aligns with one's religious and political beliefs). *Social beliefs* represent perceived beliefs of other individuals one is connected to.

A second, *external network,* describes the connections between individuals. Each connection between two individuals consists of the relation between individual A's *social beliefs* about individual B's *actual beliefs* and of the relation between individual B's *social beliefs* about

individual A's *actual beliefs*.

In sum, the two networks, internal and external, are connected through individuals' social

beliefs, which are in turn affected by both their own personal beliefs and by the actual beliefs of

others in their social circle. This assumption is based on findings that individuals perceive their

social environment relatively accurately (Galesic, Olsson, & Rieskamp, 2018; Nisbett & Kunda,

1985), while also showing ego projection in some instances (Gagne & Lydon, 2004; Goel,

Mason, & Watts, 2010).

Figure 1. NB theory premises illustrated on a minimal external network of two people, seen from

the perspective of the focal person on the left. The focal person has an internal belief network

that consists of personal (squares) and social (triangles) beliefs. The personal beliefs include a

belief about a certain issue (focal belief b_i, e.g., about safety of vaccines) and other related beliefs (b_j, e.g., different moral and health beliefs). The personal beliefs are connected by edges ω_{ij} that represent the mutual influence between the personal beliefs. The focal person also has a social belief, s_{ik}, which is their perceived belief of person k about the issue b_i (e.g., whether a friend believes vaccines are safe). The mutual influence between the social belief and the personal belief about issue b_i is represented by the edge ρ_{ik}. The influence of the actual belief b_{ik} of person k on the social belief of the focal person is represented by the directed edge α_{ik}. Note: Subscripts for the focal person and the edges between the other beliefs b_j are omitted for simplicity. The other person k has the same belief structure as the focal person.

Dissonance Reduction

The second main premise of the NB theory is that people have a need to *reduce dissonance* they might feel because of inconsistency between their different personal and social beliefs (Festinger, 1957; Gawronski, 2012; Gawronski & Strack, 2012; Heider, 1946, 1958), as well as because of a lack of correspondence between their social beliefs (perceptions of others) and what the others actually believe (Dhami & Olsson, 2008; Hammond, 1965). We assume three types of dissonance. Inconsistency between personal beliefs (e.g., believing that vaccines are effective at preventing diseases but also dangerous are inconsistent, because they have opposing implications for one's decision to get vaccinated) can cause *personal dissonance*. Inconsistency between personal and social beliefs (e.g., believing that vaccines are effective but perceiving some or all friends as believing otherwise) can cause *social dissonance*. Finally, lack of correspondence between social beliefs (perceived beliefs of others) and others' actual beliefs can give rise to *external dissonance*. Note that in contrast to personal and social dissonance,

external dissonance requires that one is able to observe lack of correspondence between one's social beliefs and others' actual beliefs (such as when a friend behaves in stark contrast to one's social beliefs).

The NB theory is the first to explicitly differentiate between personal, social, and external dissonances. Other models implement belief change processes through dissonance reduction, but they do not explicitly differentiate dissonances due to the lack of consistency of personal and social beliefs (social dissonance) and the lack of correspondence of social and actual others' beliefs (external dissonance; for related models, see Rodriguez, Bollen, & Ahn, 2016; Ellinas, Allan, & Johansson, 2017; Schweighofer, Schweitzer, & Garcia, 2020; van der Maas, Dalege, & Waldorp, 2020; more in the Discussion). To fully understand when and why individuals change their beliefs, we need to understand how these dissonances together lead to different social phenomena (e.g., see the section on Minority influence).

In line with social psychological research on dissonance, ambivalence, and related phenomena (Festinger, 1957; Newby-Clark et al., 2002), we distinguish between *potential and felt dissonance*. Potential dissonance refers to a lack of consistency and correspondence in one's beliefs, while felt dissonance refers to the psychological discomfort arising when attending to potential dissonance. The concept of felt dissonance is similar to the concept of felt ambivalence (Priester & Petty, 1996; van Harreveld et al., 2009; Dalege et al., 2018). As pointed out by several authors (e.g., Newby-Clark, McGregor, & Zanna, 2002; Priester & Petty, 1996; van Harreveld, van der Pligt, & de Liver, 2009), potential dissonance due to incongruent beliefs does not have to result in felt dissonance.

Attention to Dissonance

The third main premise of the NB theory is that the extent to which potential dissonance

translates into felt dissonance is moderated by the amount of *attention* to different inconsistencies between beliefs. This attention will depend on the overall importance of the issue, as well as on individual and cultural differences in sensitivity to potential dissonance and its sources. For example, people differ in how much they desire closure in their internal and external networks, as reflected in their propensity to seek different opinions, to like or dislike questions that can be answered in different ways, or to be irritated when one person disagrees with everyone else in a group (Webster & Kruglanski, 1994). Culture can also play a role, with for example Japanese but not European American participants showing more personal dissonance reduction in the presence of relevant others (Kitayama et al., 2004).

Individuals can direct different amounts of attention to personal, social, or external dissonance. For example, when considering who to vote for, some individuals might seek consistency between their beliefs about a certain candidate and their other personal beliefs such as beliefs about moral and economic issues. Other individuals might be more interested in achieving consistency between their beliefs about the candidate and what they perceive others around them think about the candidate. Still others might feel that it is important to know accurately what others think, that is, achieve correspondence between their social beliefs and others' actual beliefs. The attention to dissonance in different parts of one's internal and external network can also depend on the issue at hand. For example, individuals might care more about their personal dissonance when evaluating a political candidate, about their social dissonance when choosing what to wear, and about the correspondence dissonance when deciding whether to get vaccinated. Directing more attention to different parts of one's belief network results in a higher impact of those parts of the network, which will be more likely to motivate a change in one's beliefs according to the felt dissonance. Importantly, these different dissonances can be

measured and modeled, providing a nuanced picture of when and why different people change (or do not change) beliefs about different issues.

Relating Psycho-Social and Statistical Physics Constructs

The premises of the NB theory need to be translated into a computational model that allows to investigate the theory's implied dynamics and to derive testable predictions. There are many frameworks that can be used to develop such computational models (Borsboom, van der Maas, Dalege, Kievit, & Haig, 2021; Page, 2018). We choose a statistical physics framework, in which beliefs are represented as spins, potential dissonance is represented as energy, and attention is represented as temperature. We find the statistical physics framework useful for two main reasons. First, statistical physics models are uniquely well-suited to describe higher-level cognitive and social dynamics emerging from lower-level dynamics in a parsimonious way (Castellano et al., 2009; Rodriguez et al., 2016; Galesic et al., 2021; van der Maas et al., 2006). They enable us to model how belief dynamics emerges at the level of internal networks (from interactions between personal and social beliefs within individuals) and at the level of external networks (from interactions between individuals). Second, we can build on already existing statistical physics models at the internal level (Dalege et al., 2018; Schweighofer, et al., 2019; van der Maas et al., 2020) and at the external level (Castellano et al., 2009; Pham et al., 2022; Redner, 2019). We use the structure of a broad statistical physics framework to specify our particular theory of belief dynamics based on the premises described above. Of course, as every analogy in science (Gigerenzer, 1991), the statistical physics analogy has limitations, and it is important to not force statistical physics assumptions on human psychology and sociality when they are not applicable (more in Discussion). Similarly, we acknowledge that other factors than

dissonance reduction play a role in the formation and change of beliefs such as information integration (Anderson, 1972; see Discussion).

In Table 1, we describe the relationships between the psycho-social constructs in our main premises and the corresponding statistical physics constructs. The formal implementation of the NB theory rests on three core assumptions about the relationship between psycho-social and statistical physics constructs. These three assumptions map onto and enable modeling the three premises described in the previous section (Main Premises).

Beliefs as Spins

First, we assume that belief nodes can be modeled analogous to what is traditionally referred to as *spins* in statistical physics systems. In physics, the term spin refers to a variable that can take two or more states. Beliefs are assumed to range from very positive to very negative, a person can for example have a highly positive opinion about a politician A, but a weak negative opinion about a politician B. Conceptualizing beliefs as spins represents these magnitudes of beliefs as values ranging from -1 to 1, with higher values indicating stronger endorsement of the given belief.

Furthermore, like in physical systems where interacting particles are considered to be coupled, we can assume that beliefs within a single individual as well as beliefs of different individuals can be coupled, that is, they can influence each other. Couplings, or edges, can range in strength, representing different amounts of influence between one's own and between own and others' beliefs. Some beliefs are fairly independent, while other beliefs have a strong influence on each other. For example, one's beliefs about the effectiveness and safety of COVID-19 vaccines are probably more strongly related than one's beliefs about the effectiveness of COVID-19 vaccines and the safety of flu vaccines. The sign of a given coupling represents

whether two beliefs have an inhibitory or excitatory influence on each other. For example, the beliefs that vaccines are effective and safe might have an excitatory influence on each other, while the beliefs that vaccines are effective and unnatural might have an inhibitory influence on each other.

Besides influencing each other, belief nodes can also be affected by an outside influence such as new information (e.g., scientific findings about an issue), a media source (e.g., arguing for or against an issue), a transformative event (e.g., surprising elections or a war), fitness benefits or drawbacks of different beliefs (e.g., anti-vaccination views), and others. More general beliefs such as those about moral values and cultural norms can also affect beliefs about specific issues, without being affected in turn, at least not in the short run. Theoretically, all these influences can be represented in a network model, but practically it is not feasible to measure or include all of them. Therefore, to model those more durable exogenous influences we use the statistical physics analogy of *local fields* whose positive (negative) values lead to a higher probability to (not) endorse a given belief. These local fields influence other belief nodes in the network but do not get influenced in turn.

Potential Dissonance as Energy

Second, we formalize potential dissonance as *energy*. In statistical physics systems, energy is high when systems include particles that are strongly connected but misaligned. Similarly, potential dissonance is a function of both misalignment of beliefs and the strength of the influence between beliefs. Formally, if two beliefs have a strong excitatory influence between each other but are misaligned, energy will be higher than if two beliefs, which have weak excitatory influence between each other, are misaligned.

To implement our assumption that there are three types of dissonances (personal, social,

and external, see section Main Premises), we assume three different contributions to energy –
one for the misalignment between different personal beliefs, one for the misalignment between
personal and social beliefs, and one for the misalignment between social beliefs and actual
beliefs of others. That potential dissonance can be formalized as energy, and that people can
differentiate between the dissonances in different parts of their belief networks, are core
assumptions of the theory that have not been investigated previously.

To investigate the validity of our assumption that there are three separate dissonances, we
conducted two empirical studies (see Appendix A for details). We investigated the separability of
the three dissonances empirically by testing whether each potential dissonance, measured by
incongruence of beliefs multiplied by the estimated connection between them, predicts its
associated felt dissonance. We found that felt personal dissonance is reliably best predicted by
personal potential dissonance, felt social dissonance is reliably best predicted by social potential
dissonance, and that felt external dissonance is best predicted, albeit not reliably, by potential
external dissonance (see Appendix B for details).

Table 1

Main Psycho-Social Constructs and Their Associated Statistical Physics Constructs

Psycho-Social construct	Statistical physics construct
Beliefs (personal or social)	Spins (b_i, s_{ik}, for personal and social beliefs), taking any value between -1 and 1.
Influence between beliefs	Couplings, or edges, between personal beliefs (ω_{ij}), personal and social beliefs (ρ_{ik}), and social and actual others' beliefs (α_{ik}), taking any negative value (inhibitory influence) or positive value (excitatory influence), with 0 representing absence of influence between beliefs.
Exogenous influences	Local field (τ_i), taking any value, with 0 representing absence of exogenous influence, and positive (negative) values representing positive (negative) exogenous influence.
Potential dissonances between different personal beliefs, between personal and social beliefs, and between social beliefs and actual beliefs of others	Energy (H), given by the misalignment of coupled spins with each other and the local field. Misalignment between different personal beliefs is represented by personal energy (H_{pers}), misalignment between personal and social beliefs is represented by social energy (H_{soc}), and misalignment between social beliefs and actual beliefs of others is represented by external energy (H_{ext}).

Attention directed at personal, social, and external dissonance	Inverse temperature (β), taking any value from 0 to ∞, determines the noisiness of the belief updating process. Low β (high temperature) results in high noise, while high β (low temperature) results in low noise. In the NB theory, attention to the three different sources of dissonance are represented by three inverse temperatures, one for dissonance between personal beliefs (β_{pers}), one for dissonance between personal and social beliefs (β_{soc}), and one for dissonance between social beliefs and actual beliefs of others (β_{ext}). The relative size of β for the different networks determines the relative importance of the dissonances in these networks for the belief updating process.
Felt dissonance of internal (personal and social) and external beliefs	Interactive effect between energy (H) and inverse temperature (β), represented by D, with high energy and low temperature producing the highest felt dissonance.

Attention as Temperature

Third, the moderating factor of attention is implemented by using the statistical physics concept of (inverse) *temperature*. Temperature represents the "noisiness" of the belief updating process. High temperature results in more randomness in the belief updating process so that beliefs change in a way that does not necessarily reduce potential dissonance, while low temperature results in predictable belief change in a way that reduces dissonance. The NB theory

therefore implies that belief change will be related to dissonance only when a person pays attention to the dissonance, while lack of attention results in beliefs changing in an almost random fashion.

To implement our assumption that individuals direct different amounts of attention to inconsistencies among their personal and social beliefs, as well as to the lack of correspondence between their social beliefs and actual beliefs of people around them, the NB theory features three temperatures – one for directing attention to the misalignment between different personal beliefs, one for directing attention to the misalignment between personal and social beliefs, and one for directing attention to the misalignment between social beliefs and actual beliefs of others. High temperature (or low attention) of one part in the network results in high noisiness for this network, implying low impact of this network on belief change. Low temperature (or high attention) of one part in the network results in low noisiness, implying high impact of this network on belief change. Implementing the NB theory in this way leads to the consequence that attention is necessary for dissonance reduction. With enough attention, potential dissonance translates into felt dissonance, which in turn typically leads to lower (potential) dissonance. In some cases, it can be difficult to lower the dissonance, for example when one's social environment is against vaccines but one's personal beliefs are highly positive towards vaccines. Such situations are probably the ones in which individuals feel the most dissonance for a prolonged time.

To test the separability of the different attention parameters, we again used the empirical studies described in Appendix A. As described in detail in Appendix C, we first tested whether questions about the importance of reducing different dissonances formed three separated factors. We found indeed separate factors for the importance of reducing personal dissonance, for the

importance of reducing social dissonance, and for the importance of reducing external

dissonance. We then proceeded to test whether these different factors uniquely predict their

associated dissonances. We found that personal potential dissonances and social potential

dissonances are reliably predicted by the importance of reducing personal dissonance and by the

importance of reducing social dissonance, respectively. Additionally, we found that external

potential dissonance was predicted by the importance of reducing external dissonance, albeit not

reliably.

Formal Implementation

In this section, we formalize the NB theory to study the interactions of internal and

external belief network dynamics and derive quantitative empirical predictions. We first present

the general implementation of the NB theory connecting different parts of the theory and then

introduce each part separately. We illustrate each part using simulations, discuss the main

predictions following from the formalization, and present empirical tests of our assumptions and

predictions. In Appendix D, we discuss several ways to estimate and measures the core

constructs of the NB theory in empirical data.

We define the probability that an individual updates their personal belief b_i from its

current state to a new state b_i' as:

$$P(b_i \rightarrow b_i') = \frac{\frac{1}{1+e^{\Delta(D_{pers})}}}{\sum_m \frac{1}{1+e^{\Delta(D_{pers})}}}, \qquad [1]$$

where b_i' is one of m possible states ranging from -1 to 1, updated using an adapted form

of Glauber (1963) dynamics for ordinal belief states, and $\Delta(D_{pers})$ represents the difference in

felt dissonance regarding the current state b_i and a potential new state b_i':

$$D_{pers} = \beta_{pers} H_{pers} + \beta_{soc} H_{soc}, \qquad [2]$$

where H_{pers} is potential dissonance between personal beliefs (defined by Equation 3),

H_{soc} is potential dissonances between personal and social beliefs (defined by Equation 4), and

β_{pers} and β_{soc} represent the attention paid to each of the dissonances. We assume that the

correlation between these attention parameters is positive but lower than 1 (for empirical

validation, see Appendix C). The ratio between these parameters determines the amount of

weight given to different sources of dissonance. If, for example, one values the consistency of

their different personal belief smore than consistency of their personal and social beliefs, then

β_{pers} would be higher than β_{soc}.

For the potential dissonance between personal beliefs, we extend the formulation from

the AE framework (Dalege et al., 2016, 2018; see also van der Maas, Dalege & Waldorp, 2020).

In the AE framework, the dynamics of belief networks is modeled using the Ising (1925) model

and beliefs are modeled as binary. Here, we extend this formulation to a version of Potts model

that allows for an approximation of continuous beliefs (measured on a scale from -1,

representing complete disagreement, to 1, representing complete agreement). The potential

dissonance between personal beliefs is given by:

$$H_{pers} = -\sum_i \tau_i b_i - \sum_{ij} \omega_{ij} b_i b_j, \qquad [3]$$

where b_i represents the current state of given personal belief i on a continuum between -1 and 1,

τ_i represents the exogenous influence on b_i, and the edge weights ω_{ij} represent the influence

between b_i and a different personal belief b_j. If the influence between two beliefs is excitatory,

their dissonance is high when they are in different states and low when they are in similar states.

Conversely, if the influence between two beliefs is inhibitory, their dissonance is high when they

are in similar states and low when they are in different states. Note that both the values of ω_{ij}

and τ_i are always relative to β_{pers}. For example, a value of 1 for both ω_{ij} and τ_i and a value of

0.1 for β_{pers} would result in the exact same dynamics as a value of 10 for both ω_{ij} and τ_i and a

value of 1 for β_{pers}. The same holds for the edges and attention parameters in the other energy

functions. Similarly, ω_{ij} and τ_i need to be in a similar range to both be able to affect the

dynamics of the personal beliefs.

In addition to the potential personal dissonance, personal beliefs are also affected by the

potential social dissonance, given by:

$$H_{soc} = -\sum \rho_{ik}\, b_i s_{ik}, \tag{4}$$

where the ρ_{ik} represents the influence between personal belief b_i and an associated social belief s_i

one has about a person k (in other words, one's perceived belief of the person k's belief). Note

that the social beliefs have no local field (such as τ_i in Equation 3), because their dispositions are

determined implicitly by the local fields acting on personal beliefs of others in one's social

environment. We thus assume that social beliefs are *only* affected by one's own beliefs and

actual beliefs in one's external network, given by the following equation:

$$P(s_i \to s_i') = \left.\frac{1}{1+e^{\Delta(D_{soc})}}\middle/ \sum_m \frac{1}{1+e^{\Delta(D_{soc})}}\right., \tag{5}$$

where $\Delta(D_{soc})$ represents the difference in felt dissonance between the current state s_i

and a new state s_i', stemming from potential dissonances between personal and social beliefs

(H_{soc}; defined by Equation 4) and potential dissonances between social and actual beliefs of

connected individuals (H_{ext}; defined by Equation 7), each weighted by an attention parameter β:

$$D_{soc} = \beta_{soc}H_{soc} + \beta_{ext}H_{ext}, \tag{6}$$

where β_{soc} and β_{ext} represent the attention paid to each of the dissonances. We again

assume that the correlation between these attention parameters is positive but lower than 1 (for

empirical validation, see Appendix C). The ratio between these parameters determines the

amount of weight given to different sources of dissonance. If, for example, one is motivated to

hold accurate social beliefs and discusses beliefs with others a lot, it is likely that their β_{ext} will

be higher than β_{soc}.

We define the external potential dissonance as:

$$H_{ext} = -\sum_i \sum_k \alpha_k s_{ik} b_{ik},$$ [7]

where s_{ik} is the individual's social belief about belief i of person k, b_{ik} is the actual belief i of the

connected person k, and α_k represents the influence of the actual belief b_{ik} of person k on the

social belief s_{ik}. In contrast to the edges in the internal network ω_{ij} and ρ_{ik}, α_k is directed as it is

assumed that only actual beliefs can influence social beliefs directly, but that social beliefs can

influence actual beliefs only indirectly. The dynamics of each person k's beliefs b_{ik} is modeled

by Equations 1 and 2.

Because the dissonance between two beliefs in our implementation is expressed as

multiplication of belief states ($b_i b_j$ in Equation 3, $b_i s_{ik}$ in Equation 4, and $s_{ik} b_{ik}$ in Equation 5),

the lowest dissonance is achieved when two beliefs have the same (respectively different) and

extreme states when they are positively (respectively negatively) connected. We think that it is

reasonable to assume that more extreme values result in lower energies, because extreme values

represent a more unambiguous belief, which probably serves the need for consistency more than

a more ambiguous belief. Additionally, in Appendix E we investigate several other distance

measures and show that multiplication is empirically the best suited distance measure for our

theory.

Note that different potential dissonances will have different sizes, reflecting the sum over

different numbers of pairs of personal beliefs (Equation 3), personal and social beliefs (Equation

4), and social and actual beliefs (Equation 7). This formulation as sums allows that a person who

has relatively more social contacts than beliefs related to an issue might feel higher social or external dissonance than a person who has relatively more personal beliefs about an issue.

Illustrations of the Networks of Beliefs Theory's Dynamics

To illustrate the belief dynamics implied by the different attention parameters, we ran several illustrative simulations varying one attention parameter while keeping the other two attention parameters constant. We turn to simulations of the interactions between the different attention parameters in the section Explaining Established Belief Dynamics Phenomena. Codes for all simulations can be found at

https://osf.io/n58h6/?view_only=0da2fb3267574e4fa53978bc4a36ba32.

General Simulation Setup

In all simulations reported in this paper, each individual has a fully-connected personal belief network of ten nodes, as illustrated by the square nodes in panel (a) of Figure 2. One of the personal beliefs of each individual represents a focal belief that is communicated in the external network. Consequently, each individual has social beliefs about the focal belief of each of the other individuals in the network, represented by the triangle nodes in panel (a) of Figure 2. For the illustrative simulations shown in Figure 2, we used a fully-connected ten-node external network, shown in panel (j) of Figure 2. Note that each circle in this panel corresponds to one of the internal networks shown in panel (a).

We set the edge weights in all simulations to $\omega = .4$ for the influence between personal beliefs, $\rho = 1$ for the influence between personal and social beliefs, and $\alpha = 1.4$. These values lead to moderate correlations between beliefs in all parts of the belief network ($r \cong .3$) at moderate levels of attention ($\beta = 1$; see Appendix F for details). Each belief (both personal and social) can have seven different states: -1, -.66, -.33, 0, .33, .66, and 1. Simulations with other

numbers of belief states produce equivalent patterns of results. We calculate probabilities of different belief states b_i' and s_i' according to Equation 1 and Equation 5 and assign a new belief state according to these probabilities. In each iteration, all beliefs are updated in turn and in random order. Every simulation run consists of 100 iterations unless specified otherwise. Unless otherwise noted, beliefs were initialized randomly and all τ were set to 0. For the simulations in Figure 2, we compared the belief dynamics with the attention parameter of interest set to either 0.5 or 2, keeping the other attention parameters constant. For each illustrative simulation, we ran each attention condition 1,000 times.

Implications of attention directed at personal dissonance. For the implications of varying attention directed at personal dissonance, we investigate the mean distribution of personal beliefs within each individual, as well as the dynamics of the mean beliefs over time. Panels (b) and (c) in Figure 2 show the distributions of mean belief states at the end of each of the 100 runs for each individual, for each level of attention. These results suggest that low attention leads to a normal distribution of mean end beliefs, with most networks ending up in a neutral state (equivalent to how they started). In contrast, high attention leads to a bimodal distribution with most networks ending up in an extreme state. To better understand the dynamics over time, panels (d) and (e) in Figure 2 show the results of five randomly selected simulation runs over time, assuming either low or high attention. Low attention leads to networks fluctuating around the neutral middle point, while high attention leads to more stable networks that settle at more extreme belief states.
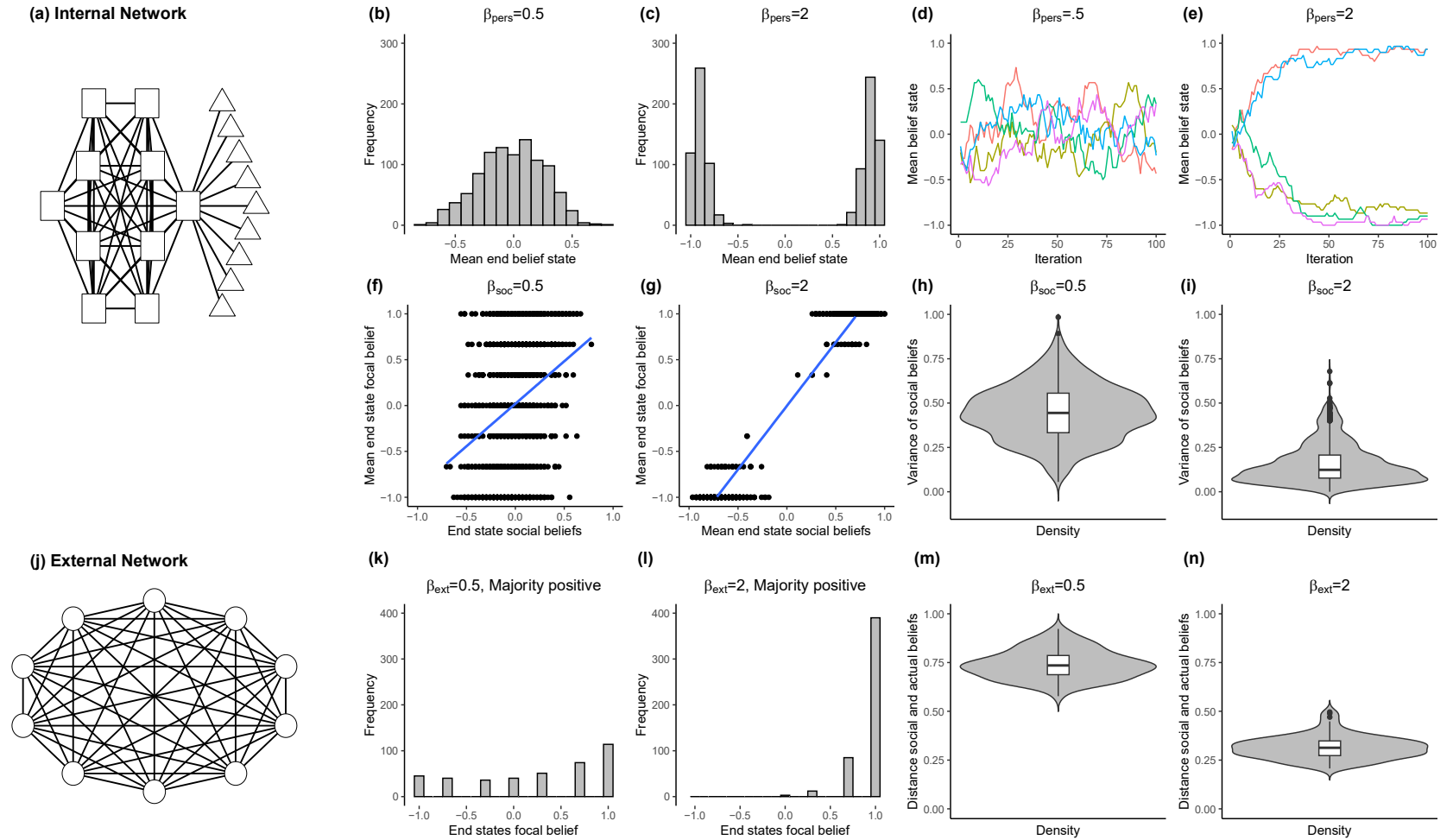
Figure 2. Illustrations of the networks of beliefs theory's dynamics. Panel (a) shows the internal network of every individual (squares represent personal beliefs and triangles represent social beliefs) used in the simulations. Panels (b)-(e) show dynamics for varying

attention to personal dissonance, either low (β_{pers}=.5, in (b) and (d)) or high (β_{pers}=2, in (c) and (e)). Panels (b) and (c) show distributions of mean belief states at the end of simulation runs, and (d) and (e) the whole dynamics of five randomly selected runs. Panels (f)-(i) show dynamics for varying attention to social dissonance, either low (β_{soc}=.5, in (f) and (h)) or high (β_{soc}=2, in (g) and (i)). Panels (f) and (g) show scatter plots of the relation between the mean end social beliefs and the mean end focal belief and panels (h) and (i) show density plots of the variance of the social beliefs at the end of simulation runs. Panel (j) shows the external network used in the simulations, with each circle representing one individual equivalent to that shown in panel (a). Panels (k)-(n) show dynamics for varying attention to external dissonance, either low (β_{ext}=.5, in (k) and (m)) or high attention (β_{ext}=2, in (l) and (n)). Panels (k) and (l) show histograms of the end states of the focal beliefs of groups which ended with a majority of positive belief states. Bars indicate frequency of different end states of focal beliefs across 10 individuals. Panels (m) and (n) show density plots of the distance between social beliefs and actual others' beliefs.

These simulation results lead to two basic implications of the NB theory. First, if individuals pay much attention to their potential personal dissonance, the theory predicts a bimodal distribution of belief averages, while it predicts a normal distribution of belief averages when individuals pay little attention to this dissonance. Second, paying attention to potential personal dissonance should lead to a higher stability and resistance of beliefs.

We tested these implications and found support for them (for details see Appendix G). First, importance of reducing personal dissonance led to more bimodal distributions of average beliefs, replicating earlier findings that involvement leads to more bimodal distributions of beliefs. Second, individuals who attached high importance to reducing their personal dissonance had lower dissonance in their personal beliefs, which predicted less change in response to an educational intervention.

Implications of attention directed at social dissonance. For the implications of varying attention directed at social dissonance, we focus on the correlation between social and personal beliefs and the variance in social beliefs within each individual. First, as can be seen in panels (f) and (g) in Figure 2, higher attention leads to higher consistency of the focal personal belief and the social beliefs. Second, as can be seen in panels (h) and (i) in Figure 2, higher attention leads to lower variance of the social beliefs at the end of a simulation run. Both of these effects result from the social beliefs becoming more interdependent with the focal personal belief.

These simulation results lead to two further basic implications of the NB theory. First, the theory predicts that those individuals will have social beliefs that are more consistent with their personal beliefs. This effect can occur either because these individuals project their beliefs more to others, because they change their personal beliefs to be aligned with their social beliefs, or due

to a mix of both these processes. Second, if individuals pay much attention to the potential

dissonance between their personal beliefs and their social beliefs, the theory predicts that

variance of their social beliefs will decrease. When one's social network is heterogeneous

(contacts have different beliefs), this effect will resemble ego-projection.

Again, we tested these implications and found support for them (for details see Appendix

G). We find that individuals who attach high importance to reducing social dissonance indeed

show higher consistency between their personal and social beliefs and lower variance in their

social beliefs.

Implications of attention directed at external dissonance. For the implications of

varying attention directed at external dissonance, we focus on how likely it is that group

radicalization occurs and how closely social beliefs are aligned with actual beliefs of other

individuals. First, as can be seen in panels (k) and (l) of Figure 2, higher attention to external

dissonance leads to group radicalization in the focal beliefs. The distribution of end focal beliefs

becomes skewed towards the extreme end of the initial average view (here we show only the

groups where the majority has positive beliefs), and more so when attention to external

dissonance is high. Second, as can be seen in panels (m) and (n) of Figure 2, social beliefs

(perceptions of actual others' focal beliefs) are closer to the actual beliefs when attention to

external dissonance is high.

These simulation results lead to two further basic implications of the NB theory. First, if

people pay a lot of attention to the dissonance between their social beliefs (what they believe

others think) and what others actually believe, group radicalization is more likely to occur and

individuals are more likely to agree with each other. Second, individuals are predicted to have

more accurate perceptions of their peers' beliefs if a belief is discussed and/or expressed often in

a group, as this will presumably increase group members' awareness of and attention to any lack of correspondence between their perceptions of others' beliefs (social beliefs) and what others actually believe.

As before, we tested these implications and found support for them (for details see Appendix G). We find that both importance of reducing external dissonance and frequency of discussion are positively related to holding more extreme beliefs and also more accurate social beliefs.

Explaining Established Belief Dynamics Phenomena

In this section we discuss how the NB theory can explain established phenomena in belief dynamics, including group polarization, group radicalization, and minority influence. First a note on the terminology. Some authors use the term group polarization to describe extreme differences in beliefs of different segments of the population (e.g., Iyengar & Westwood, 2015; Yardi & Boyd, 2010; Axelrod et al., 2021), or strong dislike among different segments (Mason, 2016). Here we will use the term "group polarization" in this sense. Other authors (e.g., Isenberg, 1986; Friedkin 1999; Sunstein, 1999) use the term "polarization" to denote radicalization of a whole group towards the same more extreme position, usually occurring after a group discussion or a mere exposure to arguments in line with an existing view. We will therefore call this type of polarization "group radicalization". Finally, we will discuss separately the phenomena of minority influence on group polarization and radicalization.

Consensus and Polarization

A number of different models have been developed to explain consensus and polarization in societies (Flache et al., 2017; Levin et al., 2021). Long run consensus is associated mostly with unconditional social influence models. These models assume that if individuals are

connected by an edge, they will always influence each other towards reducing their belief differences. For nominal beliefs, examples include voter models (Clifford & Sudbury, 1973; Holley & Ligget, 1975; Redner, 2019), while for continuous beliefs the most prominent ones are averaging models (DeGroot, 1974; Friedkin & Johnsen 1990, 2011).

Group polarization has been extensively studied with conditional influence models that assume a threshold or a tolerance level that determines if an individual will be influenced by a neighbor (Axelrod, 1997; Deffuant et al., 2000; Hegselmann & Krause 2002; van der Maas, Dalege, & Waldorp, 2020). Some models also include repulsive social influences, whereby the influence weights can be negative or positive depending on the difference in beliefs between individuals (Flache & Macy, 2011; Huet & Deffuant, 2010).

Several models go beyond this threshold dynamics of influence weights and include other psycho-social mechanisms that can produce similar patterns. For example, in Nowak et al.'s dynamic version of Latane's (1981) theory of social impact, group polarization is explained through an interplay of two individual parameters: persuasiveness - the likelihood that one will influence those who have a different point of view, and supportiveness - the likelihood that one will help those with the same point of view to resist change. In Leonard et al. (2021; see also Franci et al., 2021), it is explained through a dynamic interaction of group-level self-reinforcement and reflective partisanship, which determine the extent of within and between group influences. Finally, some models explain polarization through the process of balancing the relationships between beliefs and between individuals holding different beliefs (Pham et al., 2020; 2022; Rodriguez et al., 2016; Schweighofer et al., 2020; see Discussion).

In the NB theory, we aim to explain consensus and polarization that occurs in structured networks where people are already grouped in clusters due to reasons other than their beliefs, e.g.

geography, workplaces, family groups etc. These structures are ubiquitous in most human collectives. We ask, if such a clustered society starts with equally distributed, largely moderate beliefs, in what circumstances does it become polarized? We posit that consensus and polarization can emerge simply because of different levels of attention people pay to their personal, social, and external dissonances. For example, people's attention to these dissonances can change after prominent events (e.g., surprising elections or court rulings about issues such as abortion and gay marriage). As we show next, when people pay a lot of attention to their personal dissonance (how their own beliefs align to each other), the society remains rather diverse (heterophilous). However, when they pay more attention to their social and external dissonances (how their social beliefs align with their other beliefs and others' beliefs), either complete (radical) consensus or polarization can occur.

To investigate this, we run a simulation with 100 individuals, who each have an internal fully connected network ($\omega = 1$) of ten personal belief nodes, initialized at random belief states. One of these internal nodes functions as the focal belief in this simulation. This belief is connected, via social beliefs, to focal beliefs of other individuals (see Figure 1). We generate a social network using stochastic block model with two clusters of 50 individuals. We set the probability that any individual is connected to another individual in the same cluster to .4 for both clusters, and the probability that an individual is connected to an individual from the other cluster to .01. We set each type of attention to dissonance (β_{pers}, β_{soc}, and β_{ext}) to either 0.5 (low attention) or 2 (high attention). For each of the resulting eight combination of attention parameters, we simulated 100 runs. In each simulation run, nodes were first randomly initialized and then each updated over 100 iterations.

We use both qualitative and quantitative measures to understand how different levels of attention to dissonance affect the likelihood of consensus and polarization. On the qualitative level, we visualize the networks representative of those occurring at the end of simulation runs in different conditions. Panels (a) to (e) of Figure 3 show five types of networks that we observe in our simulations, showing radical polarization (RP) where the two social network clusters end up having completely opposing beliefs, radical consensus (RC) where the two clusters end up having the same beliefs, moderate polarization (MP) where individuals in the two clusters have different but moderate beliefs, moderate heterophily (MH) where the two clusters have similar distribution of mostly moderate beliefs, and extreme heterophily (EH) where the two clusters have similar distribution of mostly extreme beliefs. Abbreviations of different networks in panels (f) to (i) of Figure 3 show in which conditions they appear.

On the quantitative level, we use two measures of consensus and polarization. One is variance of focal beliefs at the end of simulation runs in different conditions, shown in panels (f) and (g) of Figure 3. High values of variance show that network members hold very different opinions (as in the network EH described above) while low values of variance indicate that the opinions are not very different from each other (as in MF or MR networks). Note that in some conditions RP (extreme variance) and RC (no variance) networks are equally likely outcomes, producing a moderate average variance. For this reason, we also use another quantitative measure, modularity, showing how likely is that connected individuals hold similar focal beliefs (panels (h) and (i); Newman, 2006). High values of modularity indicate that connected individuals hold similar beliefs (corresponding to RP, RC, and MP networks above), while low values of modularity indicate that connected individuals are not more likely to hold similar

beliefs than unconnected individuals (corresponding to MH and EH networks). To calculate

modularity, we binarized the focal beliefs into positive and negative.

Taken together, the patterns in Figure 3 suggest four main insights. First, when attention

to both personal and to social dissonance is low ($\beta_{pers} = \beta_{soc} = 0.5$), beliefs almost always

remain moderate and the final state resembles either moderate polarization (MP, where the two

clusters have different, but moderate beliefs) or moderate heterophily (MH, where there is little

difference between the two clusters), independently of attention to external dissonance. In other

words, when individuals pay little attention to either how much their personal beliefs align to

each other, or how much their personal beliefs align with the perceived beliefs of their social

contacts, they will likely end up holding moderate beliefs. If in addition they pay little attention

whether their perceived beliefs about social contacts are accurate ($\beta_{ext} = 0.5$), there will be a

range of moderate beliefs throughout the network (MH). If they do pay attention to accuracy

($\beta_{ext} = 0.5$), the result will be a moderately polarized network (MP).

Second, when attention to personal dissonance is high and attention to external

dissonance is low ($\beta_{pers} = 2$ and $\beta_{ext} = 0.5$) extreme heterophily (EH) is likely to emerge almost

independently of the level of attention to social dissonance (β_{soc}). In this case, individuals try to

align their personal beliefs and are not pressured to align with their actual social environments.

This allows their beliefs to radicalize unrestricted by others' beliefs.

Third, when attention to social dissonance is high and attention to external dissonance is

low ($\beta_{soc} = 2$ and $\beta_{ext} = 0.5$), extreme heterophily (EH) is again likely to emerge, independently

of the level of attention to personal dissonance (β_{pers}). In this case, individuals try to align their

personal beliefs and perceived beliefs of their social contacts, unfettered by concerns about what

their social contacts actually believe. These dynamics allows for radicalization of personal

beliefs independently of actual social environments.

Fourth, when attention to social dissonance and external dissonance are both high ($\beta_{soc} = \beta_{ext} = 2$), either radical polarization or radical consensus will emerge (RP or RC) with equal

probability, depending on minor biases in the initial random configuration of beliefs. If both

clusters randomly initialize with the same tendency in beliefs, we observe extreme consensus; if

the clusters randomly initialize with different tendencies in beliefs, we observe extreme

polarization. In this condition, individuals are strongly motivated to conform to their social

environments, radicalizing their personal beliefs in the process because the overall dissonance is

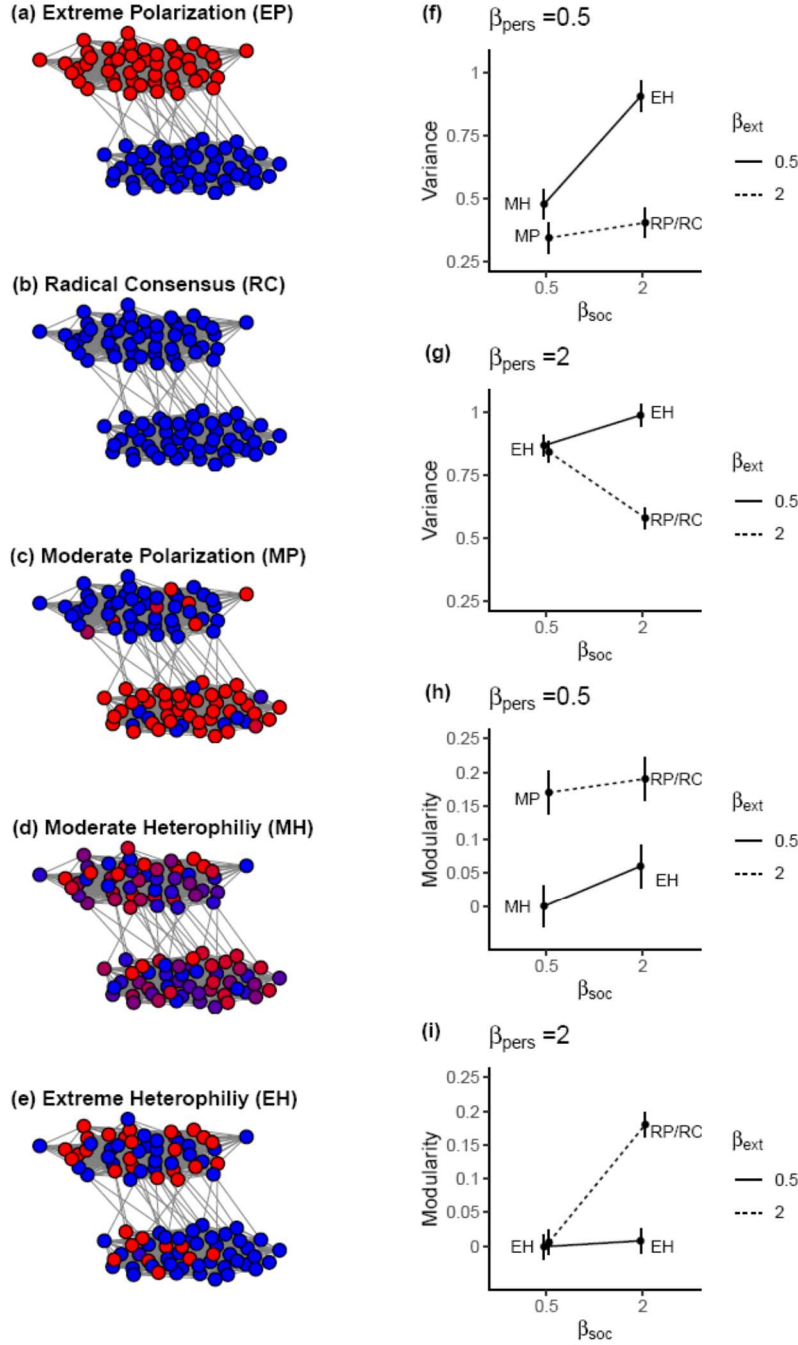lowest when all beliefs in one's network are extreme.

Figure 3. Conditions for different patterns of consensus and polarization. Panels (a) to (e) show representative illustrations of five types of networks that occur at the end of simulations assuming different combinations of high and low attention to personal, social, and external dissonances (β_{pers}, β_{soc}, and β_{ext}, respectively). Panels (f) and (g) show variance of focal beliefs, and panels (h) and (i) their modularity at the end of simulation runs in different

conditions. Each point in panels (f)-(i) is marked by the abbreviation of the network type that is most likely to occur in that case. Error bars indicate 95% confidence intervals. In network illustrations, circles represent individuals with their personal and social beliefs (equivalent to the circles in Figure 1), and edges represent the directed influence of individuals' actual beliefs to the social belies of their contacts (each equivalent to the two edges connecting the individual circles in Figure 1).

In sum, the NB theory can explain patterns of both consensus and polarization. Moderate consensus - clustered around a neutral opinion - is likely to occur when individuals do not pay too much attention to personal and social dissonances. On the other hand, either extreme consensus or extreme polarization can occur when individuals do not pay too much attention to personal dissonance and attend strongly to their social and external dissonances. Finally, if individuals pay a lot of attention to personal dissonance, individuals tend to hold extreme beliefs that are independent of the social network, resulting in extreme heterophily.

These simulations also point to ways in which radical polarization might be reduced. One is reducing attention to social dissonance, or more generally the need to align personal beliefs with the perceived beliefs of our social contacts. As shown for example in panels (f) and (g) of Figure 3, reducing β_{soc} from 2 to 0.5 leads to occurrence of moderate or extreme heterophily instead of radical consensus/ polarization. Note that reducing β_{soc} while keeping β_{ext} high (attention to accuracy of social beliefs) will generally lead to accurate social beliefs. Taken together, our results suggest that promoting originality and non-conformism could be sufficient to reduce unfavorable societal outcomes such as complete polarization or (seemingly less damaging but potentially dangerous in the long run) radical consensus.

Group Radicalization

Group radicalization refers to the phenomenon that group discussion can lead to a more extreme average group opinion than it was before the discussion. For instance, in Sunstein et al. (2007), liberal groups that discussed issues such as affirmative action or civil unions adopted more extreme beliefs about those issues than they started with, although such radicalization does not always occur (Isenberg, 1986). A variety of accounts have been proposed to explain this phenomenon. Some involve motivational processes such as a desire to be 'better' than the group average (Isenberg, 1986) and to be distinct from an outgroup (Turner, 1985). Other accounts postulate sampling processes that lead to higher likelihood of sharing arguments that are in line with majority the view (Burnstein & Vinokur, 1973). Yet other accounts propose that the core mechanism is re-alignment of internal belief networks, so that beliefs about an issue align with one's pre-existing and more stable internal preferences or core values such as political or religious orientation. Group discussion, or even just thinking about an issue (Dalege et al., 2018), can make people more aware of the alignment (or lack of thereof) of their belief about a particular issue and their core values. The threshold models described in the previous section, and their extensions, can also produce group-radicalization. For example, depending on the threshold difference in beliefs, influence weights between individuals can become positive or negative, leading these models to naturally settle in bi-polar belief states (Flache & Macy, 2011; Huet & Deffuant, 2010; Jager & Amblard, 2005)

The NB theory predicts group radicalization through a novel process where attention to different dissonances produce the phenomenon. To show this, we simulate a group of ten individuals who are all equally-strongly connected to each other (all $\alpha = 1$). All these individuals have an internal network of ten beliefs. Each of these networks is fully connected (all $\omega = 1$).

Each simulation run consisted of two phases of 100 iterations each, with the first phase modeling

the period before the group discussion, and the second phase the period after the group

discussion. In the first phase, both social dissonance β_{soc} and external dissonance β_{ext} were set

to 0, representing that individuals are not communicating with each other yet. In the second

phase, both β_{soc} and β_{ext} were set to either 0.5 or 2, representing less and more engaged group

discussion. In addition, attention to personal dissonance β_{pers} was also set to either 0.5 or 2,

representing less or more pressure to align one's personal beliefs with each other. Belief states at

the beginning of each simulation were initialized at random.

The results of these simulations are shown in Figure 4. As before, we examine the results

both through network visualizations and quantitative indicators. Panels (a)-(d) show four

different distributions of beliefs in networks observed at the end of our simulations. In the No

radicalization moderate (NRM) network, members have a range of relatively moderate beliefs. In

the no radicalization extreme (NRE) network, members have mostly extreme, but diverse beliefs.

In Moderate radicalization (MR) and Extreme radicalization (ER) almost all or all members,

respectively, have the same extreme beliefs.

For more nuanced results, panels (e)-(h) provide absolute means of focal beliefs at the

end of the two phases of our simulation, for different conditions. The higher those means, the

more radicalized the group. Panels (e) and (f) show that, as expected, without group discussion

beliefs in all conditions show no group radicalization, with different individuals still holding

different beliefs after 100 iterations. The only difference between conditions is that a stronger

attention to the consistency of personal beliefs (β_{pers} = 2 vs. 0.5) leads to more extreme personal

beliefs (panel (f) vs. (e) in Figure 4). After group discussion, however, beliefs can radicalize but

not necessarily. When attention to personal dissonance is low (β_{pers} = 0.5), high attention to

social dissonance ($\beta_{soc} = 2$), is sufficient to foster moderate (MR) or extreme (MR)

radicalization independently of external dissonance. In other words, when people try to align

their personal beliefs with perceived beliefs of others, while disregarding the consistency

between their personal beliefs, this will lead to radicalization even if their perceptions do not

correspond to the actual beliefs of others. When attention to personal dissonance is high ($\beta_{pers} =$

2), high attention to social dissonance ($\beta_{soc} = 2$), is not sufficient to overcome the need to align

personal beliefs with each other. In that case, high attention to external dissonance ($\beta_{ext} = 2$) is

also needed to produce at least moderate radicalization.
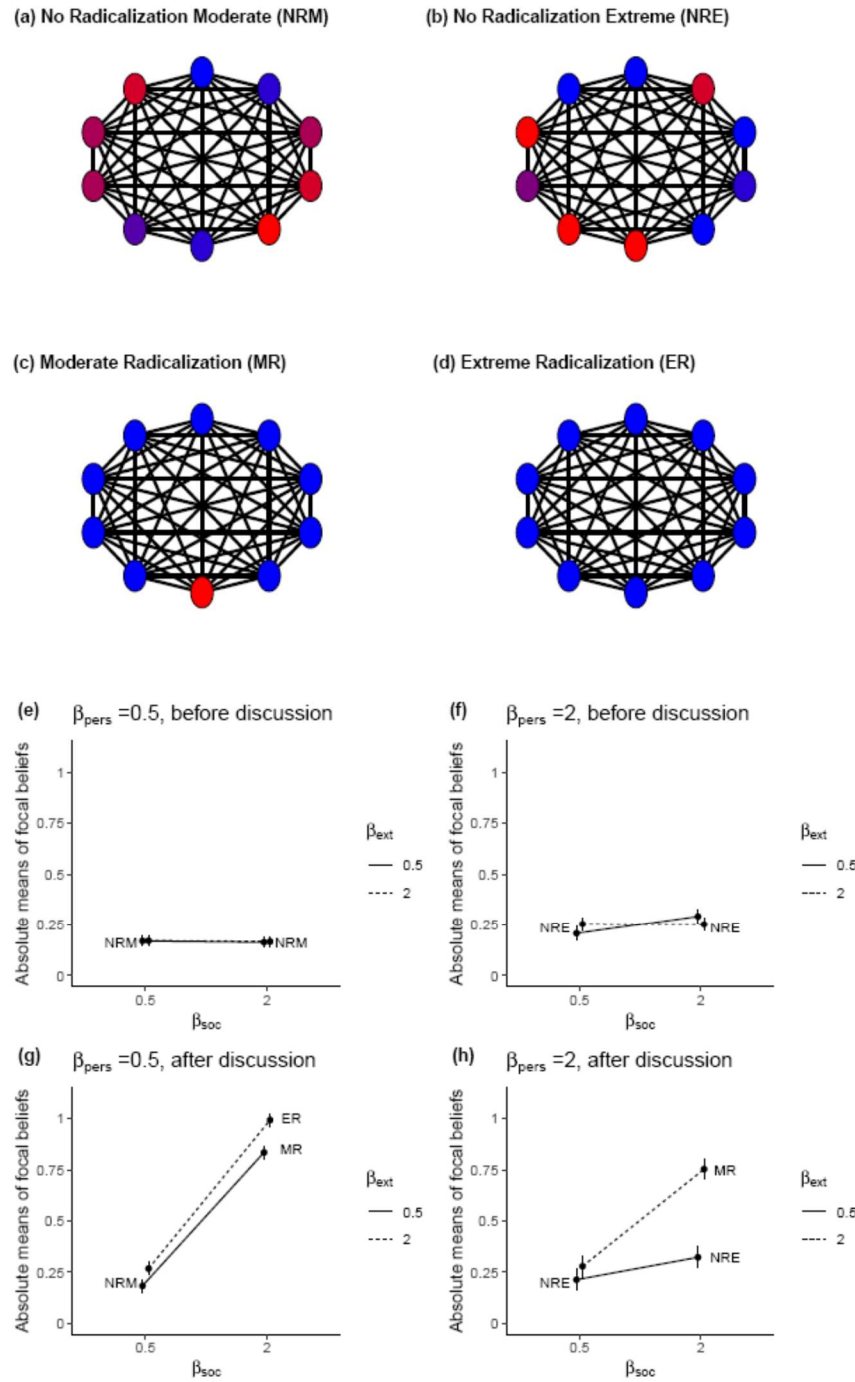
Figure 4. Conditions for group radicalization before and after discussion. Panels (a)-(d) show

representative illustrations of four types of networks that occur at the end of simulations

assuming different combinations of attention parameters. Panels (e) and (f) show absolute means

of focal beliefs after 100 runs before group discussion, and panels (g) and (h) show the same at

the end of additional 100 runs after group discussion. Error bars indicate 95% confidence intervals. In network illustrations, circles represent individuals with their personal and social beliefs (equivalent to the circles in Figure 1), and edges represent the directed influence of individuals' actual beliefs to the social belies of their contacts (each equivalent to the two edges connecting the individual circles in Figure 1). Blue nodes indicate negative beliefs and red nodes indicate positive beliefs, with higher color saturation corresponding to higher extremity of beliefs.

In sum, high attention to consistency of personal beliefs (high β_{pers}) is in most cases sufficient to protect against group radicalization, except when individuals also pay a lot of attention to how well their personal beliefs align with their social beliefs, and how their social beliefs align with the actual beliefs of others. More intense group discussions and in particular those that focus on (dis)agreement between individuals might raise the attention to both of these latter sources of dissonance (social and external), making group radicalization more likely. These results also suggest ways to reduce group radicalization: by reducing attention to social dissonance, for example by encouraging non-conformity and originality during group discussions.

Minority Influence

Minority influence refers to the process by which a small group of individuals with dissenting beliefs and behaviors can gradually influence and change the beliefs and behaviors of the majority (Moscovici et al., 1969; for reviews see e.g., Martin & Hewstone, 2009). A plethora of factors can affect how successful a minority is in influencing a majority (Crano, 2010; David & Turner, 2001; Quiamzade et al., 2010; Smith & Tindale, 2010). Minorities can directly

influence focal beliefs of the majority, or indirectly influence related beliefs even when focal beliefs of the majority remain unchanged.

Here, we focus on modeling the conditions for direct minority influence. We ran a simulation with 30 individuals embedded in a network with two clusters generated by a stochastic block model. The majority cluster consisted of 20 individuals whose belief states were initialized randomly in the range from -1 to 0. The minority cluster consisted of 10 individuals whose belief states were initialized randomly in the range from 0 to 1. The probability that any individual in the majority cluster was connected to any other individual in the majority cluster was set to .33, while the probability that any individual in the minority cluster was connected to any other individual in the minority cluster was set to .67. These settings resulted in an average degree of 6 within both clusters taking into account only the connections within each cluster. The probability of a link between the minority and majority clusters was set to .1. We also investigated stochastic block model networks with more sparsely connected clusters, as well as random networks, but the patterns of results were similar.

We varied the attention parameters independently for the majority and minority clusters. For ease of interpretation, we varied the attention to social and external dissonances together, setting both β_{soc} and β_{ext} to either 0.5 or 2. The attention to personal dissonance β_{pers} was set independently to either 0.5 or 2. For each combination of attention parameters, in both majority and minority clusters, we simulated 100 runs. Belief states at the beginning of each simulation were initialized at random.

To investigate under which circumstances we observe majority influence, minority influence, or no influence of either group on the other, we study both visual representations of networks typically observed at the end of our simulations, and mean focal beliefs at those end

points. As shown in panels (a)-(c) of Figure 5, we observe three representative end networks: Minority influence (MinI) where minority belief overcomes the whole network, Majority influence (MajI) where majority influence wins, and Polarization (P) where both groups become radicalized and extremely different from each other.

Furthermore, panel (d) of Figure 5 shows mean end focal beliefs for different combinations of attention parameters of the majority group, averaging over the attention parameters of the minority group. Panel (e) shows equivalent results for the attention parameters of the minority group. Note that in this setup, where as described above the majority comprises ⅔ of the population and has an average initial belief of -.5, and the minority comprises ⅓ of the population and has an average initial belief of .5, the average belief without any influence is expected to be -.17, as indicated by the red dashed line in panels (d)-(e) of Figure 5. Mean end focal belief that is around -.17 indicates no influence of either group, a mean that is clearly below indicates majority influence, and a mean that is clearly above indicates minority influence.

The results in Figure 5 suggest that a strong minority influence can occur only when majority group pays little attention to the dissonances in personal, social, and external belief networks (whenever $\beta_{pers} = \beta_{soc} = \beta_{ext} = 0.5$; top left point in panel (d)). When the majority attends to either of these dissonances, majority influence is more likely to occur. Independently of majority's levels of attention, minority group can at least avoid majority influence (if not influence the majority itself) by paying a lot of attention the consistency of own personal beliefs (top two points in panel (e)). In this case, group polarization occurs with the two groups holding radically different beliefs.
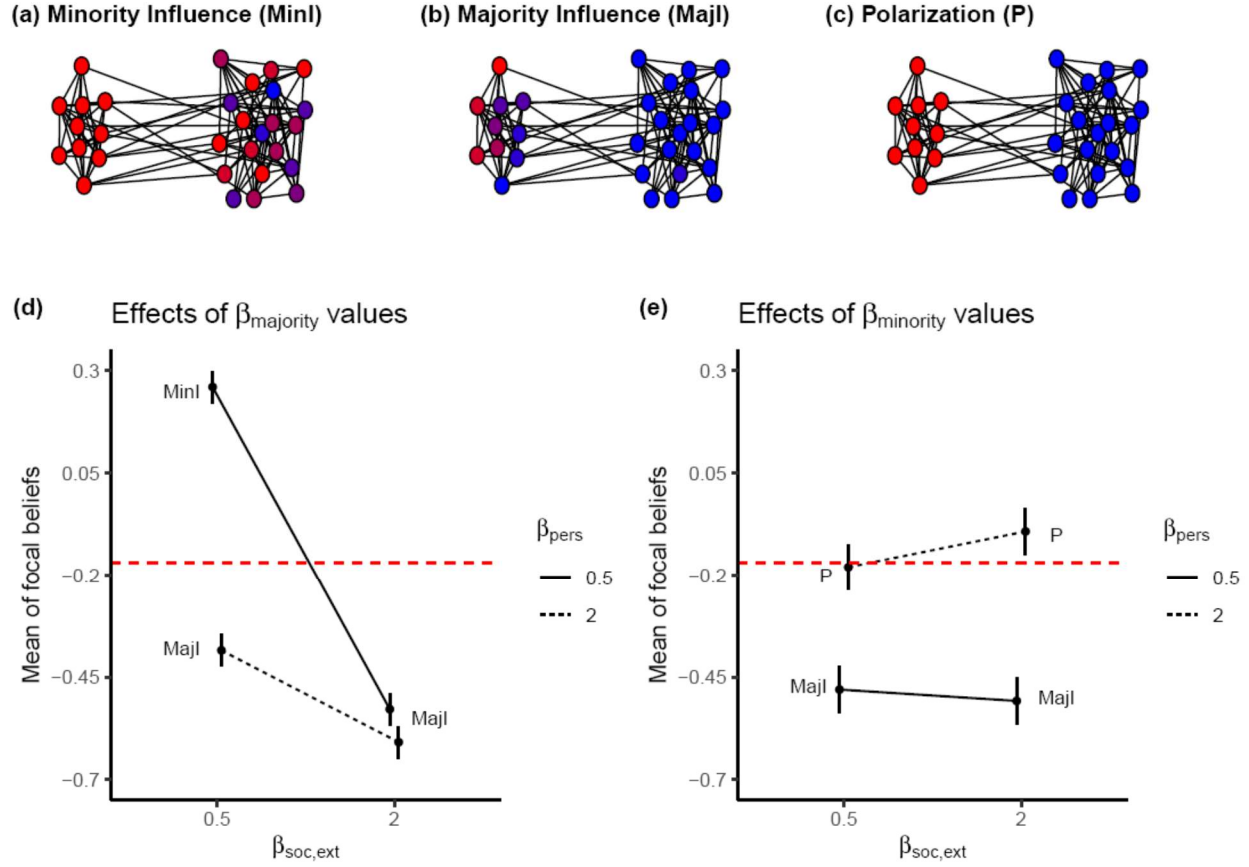
Figure 5. Conditions for minority and majority influence. Panels (a)-(c) show representative

illustrations of three types of networks that occur at the end of simulations assuming different

combinations of attention parameters. Panel (d) shows effects of majority β values, averaging

over minority β values, and panel (e) shows effects of minority β values averaging over majority

β. Red dashed lines indicate no majority or minority influence. In network illustrations, circles

represent individuals with their personal and social beliefs (equivalent to the circles in Figure 1),

and edges represent the directed influence of individuals' actual beliefs to the social belies of

their contacts (each equivalent to the two edges connecting the individual circles in Figure 1).

Blue nodes indicate negative beliefs and red nodes indicate positive beliefs, with higher color

saturation corresponding to higher extremity of beliefs.

Taken together, our findings suggest that the minority will have the most influence on the rest of the network if they strive to keep their personal beliefs consistent, and if the majority is distracted and does not pay attention to their personal, social, or external dissonance. These predictions fit well with empirical findings showing that minorities must hold consistent beliefs to influence majorities (Moscovici, Lage, & Naffrechoux, 1969; Wood, Lundgren, Ouellette, Busceme, & Blackstone, 1994). At the same time, the majority will have the most influence on the minority if they pay attention to those different sources of dissonance and strive for consistency of beliefs, independently of what the minority does.

Comparing Model Predictions with Real-World Trends

In this section, we discuss whether the NB theory can reproduce belief dynamics found in real-world data. We first model increasing polarization in the United States over time and then turn to cross-sectional distributions of beliefs in different European countries.

Empirical Trends in Group Polarization

A trend observed in many real-world settings is, as mentioned before, an increased polarization of societies into groups holding extreme and opposing beliefs. For example, as Figure 6 shows, evaluations of Democrat and Republican political candidates, as recorded in American National Election Studies, have become more and more extreme over the last several decades. Notably, interest in political campaigns has also increased in the same period.

The NB theory explains group polarization and related phenomena through the interplay of attentions to personal, social, and internal dissonances (see section Consensus and Polarization). In particular, when people pay a lot of attention to the consistency of their personal and social beliefs, as well as to correspondence of others' perceived and actual beliefs, either radical consensus or polarization can occur, depending on small biases in initial conditions. If the

two groups are even a little bit biased towards opposing positions to begin with, an increased

attention to these dissonances will likely lead to radical polarization.

It is reasonable to assume that the increased interest in political campaigns relates to

increased attention to political beliefs of others one knows and to the dissonance between own

and others' perceived and actual beliefs. According to the NB theory, this would then naturally

lead to increased radicalization of political beliefs within initially moderate groups of Democrats

and Republicans, and consequently to group polarization.
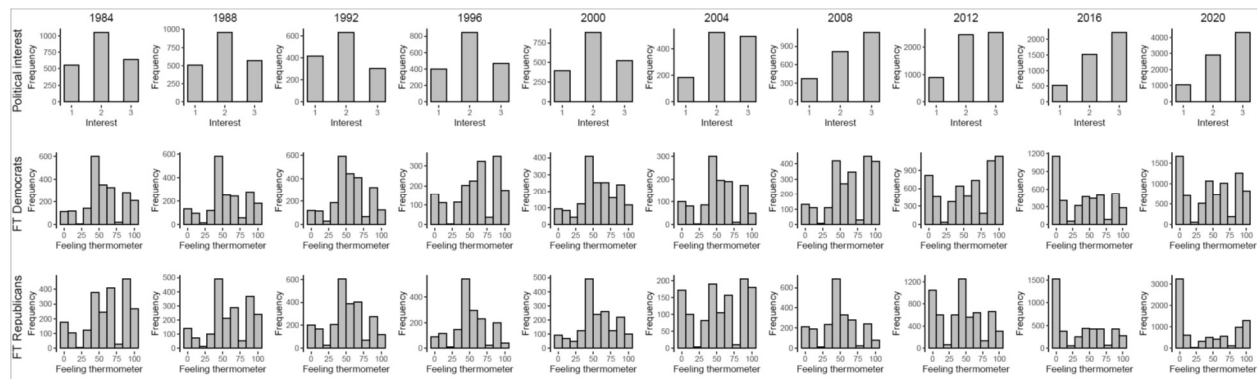
Figure 6. Trends in answers to questions about political interest (top row, question "Would you

say that you have been [not much interested-somewhat interested-very much interested] in the

political campaigns so far this year?") and evaluations of Democrat and Republican political

candidates that year ("How would you rate candidate [on the feeling thermometer]?"). Data and

full text of questions are available at electionstudies.org. Note: FT = feeling thermometer.

To investigate this prediction in more detail, we ran a simulation where we varied

network structure (from two initially disconnected clusters to two strongly interrelated clusters,

top to bottom rows of Figure 7) and increased all attention parameters gradually over time (from

$\beta_{pers} = \beta_{soc} = \beta_{ext} = 0.25$ to 1.5 over five time steps). Each network included 30 people, with 15

people in each cluster. The clusters differed in their exogenous influence, the first cluster had negative exogenous influence, $\tau = -.5$ and the second cluster had positive exogenous influence, $\tau = .5$. These differing exogenous influences could, for example, represent one cluster consisting of people living in a rural area, making them more predisposed to identify as conservatives and the other cluster consisting of people living in an urban area, making them more predisposed to identify as liberals. Within-cluster probability of connection was .35 and between-cluster probability was either 0, .01, .1, or .35. Initial beliefs in each cluster were set randomly. We ran 100 runs with eight iterations in each run.

Figure 7 shows the results for each time step. The results suggest that, starting from random initial beliefs, group polarization can occur purely because of the increase in people's attention to dissonance. This is likely to happen whenever there are sparse or no connections between the two clusters (the first two rows in Figure 7). When the two clusters are moderately related (connectivity 0.1, third row in Figure 7), both polarization or consensus are equally likely. Finally, when the two clusters are tightly related, consensus will occur in the direction of whatever the initial bias was.

These results have implications for interventions that could be used to reduce the likelihood of group polarization. Paradoxically, reducing people's attention to political campaigns would help, but so would an increased level of intergroup contact between different groups. Note that such an increased contact could lead to adoption of whatever option was in slight majority initially.
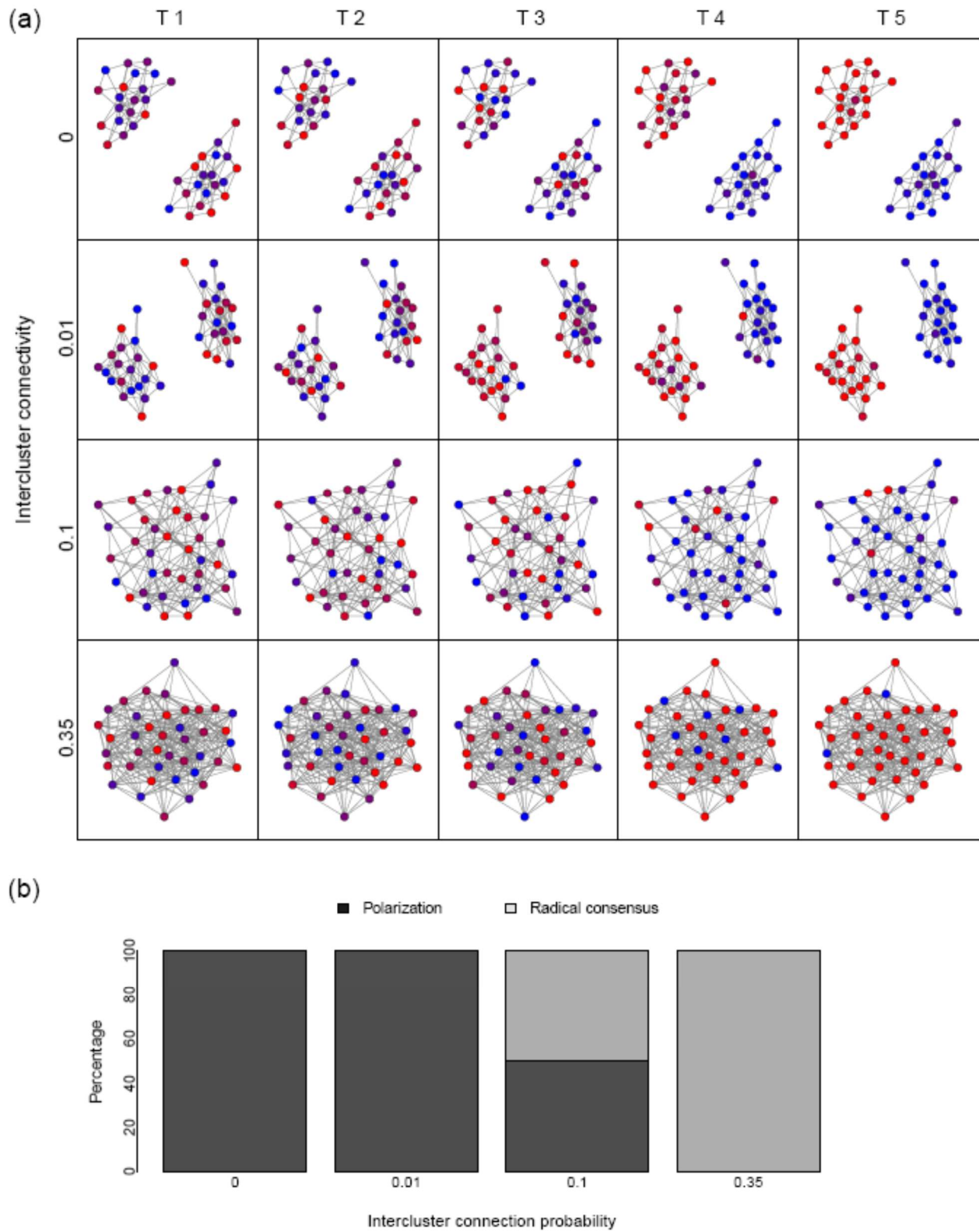
Figure 7. Simulation of belief dynamics occurring on networks with different levels of

intercluster connectivity (panel (a), top to bottom row: from no to high connectivity). At each

time step, attention to dissonance increases, resulting in different end patterns of beliefs. When intercluster connectivity is zero or low, the dynamics always leads to radical polarization (the first two bars in panel (b)). When it is medium (0.1), the dynamics ends up in radical consensus half of the time, and in radical polarization the other half, depending on small differences in initial belief distribution in the two clusters (the third bar in panel (b)). When the connectivity is high (0.35), the end result is always radical consensus (the fourth bar in panel (b)).

Cross-Sectional Distribution of Beliefs

We now look at how well the NB theory can explain specific distribution of beliefs observed at a single time point in different countries. Figure 8 shows data from four countries about respondents' own political placement and opinion about the European Union. Flache et al. (2017) argued that these data showed evidence of several belief dynamic processes. The central peak suggests assimilation to the central or moderate belief in the population, the non-extreme clusters just next to the center suggest similarity bias, and the (small) extreme peaks might suggest polarization produced by a mix of unconditional and repulsive processes.

The NB theory can reproduce this general pattern conditional on two empirically testable assumptions. First, most individuals pay little attention to the consistency of their personal beliefs about these topics, meaning that their attention parameters β_{pers} is low. Second, political placement represents an average of several more specific beliefs. To show that these assumptions can reproduce the empirically observed distributions, we ran a simulation with 100 individuals who each have a fully connected network consisting of 50 belief nodes, representing the many specific beliefs that give rise to the answer on the political placement and the EU questions. In this simulation, we focus only on dissonance arising from internal beliefs and omit the external

network by setting the attention parameters for both social and external dissonance to 0. We draw the values of the attention toward personal dissonances from exponential distributions with rates of either 2 (representing an exponential distribution with many extreme values), 5 (representing an exponential distribution with moderately many extreme values), or 10 (representing an exponential distribution with few extreme values). These distributions represent populations in which most individuals have low levels of attention to the consistency of their beliefs, with more or fewer individuals paying some attention. Exponential distribution is appropriate for such cases with a cutoff at 0 and a high upper bound.
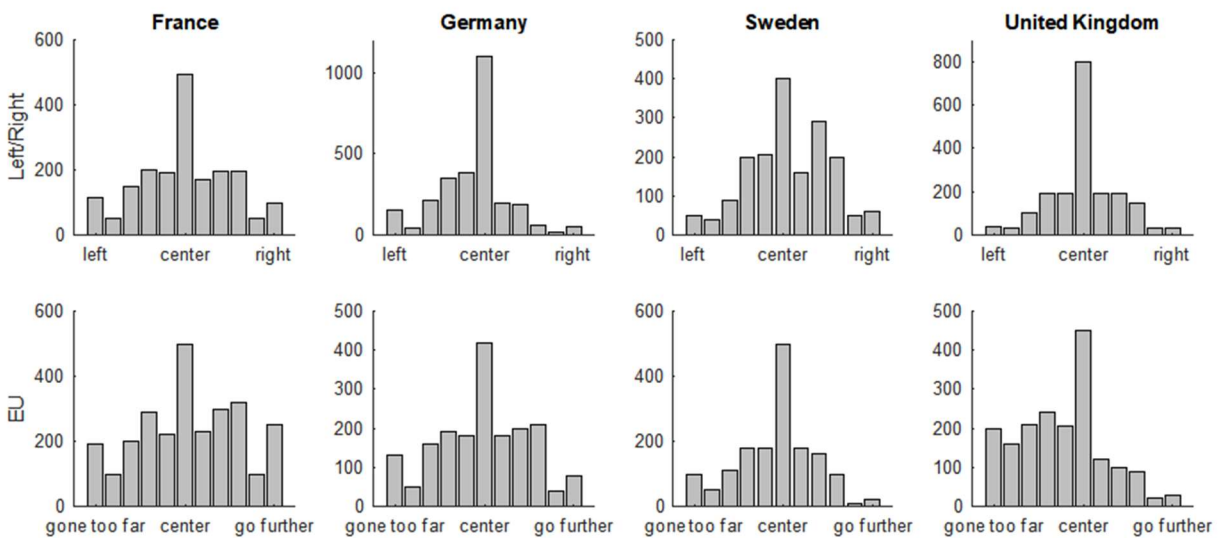
Figure 8. Empirical distributions of answers to questions about political position (top row, question "In politics people sometimes talk of "left" and "right". Where would you place yourself on this scale?") and the EU (bottom row, question "Now thinking about the European Union, some say European unification should go further. Others say it has already gone too far. What best describes your position?") in four European countries, from the 2012 European Social Survey (adapted from Figure 3 in Flache et al., 2017).

As shown in Figure 9, our simulations can reproduce some of the main patterns in the empirical answer distributions, depending on the attention distribution in the population (insets in Figure 9). The assumption that most people are not paying much attention to the consistency of their beliefs (the middle and right panel in Figure 9) produce predicted answer distributions that are closest to the empirical data. Attention distributions alone cannot explain biases towards one or the other end of the answer distributions noticeable in Figure 8. Those biases could be explained by assuming exogeneous influences on people's beliefs, such as political leaders and specific country-level events.
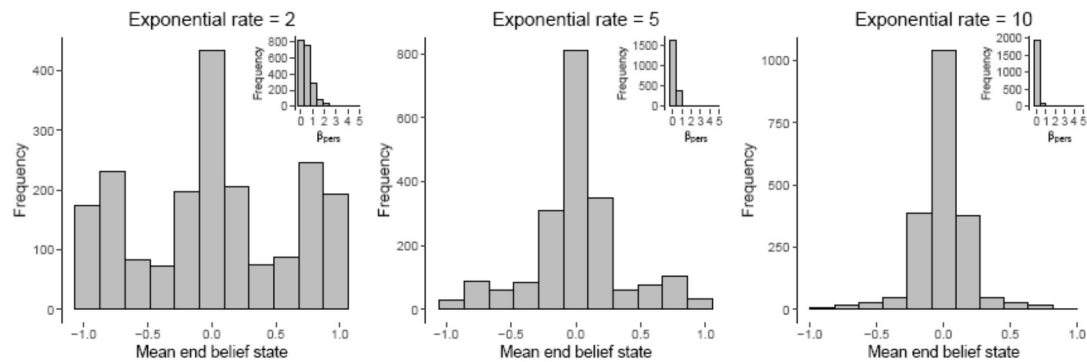
Figure 9. Predicted distribution of answers to questions shown in Figure 8. Insets show hypothesized distributions of attention parameters.

Discussion

In this paper, we develop an integrative theory of belief dynamics, the Networks of Belief (NB) theory based on three main premises about the underlying psycho-social constructs: 1) beliefs can be represented as nodes in a network, with edges representing mutual influence of beliefs that different individuals have about various issues and about each other, 2) people have a

need to reduce dissonance they might feel because of inconsistency between their different

beliefs, 3) the extent to which potential dissonance translates into felt dissonance is moderated by

the amount of attention to different inconsistencies between beliefs. Crucially, the theory

assumes that beliefs of interconnected individuals form two distinct but interacting networks, an

internal belief network consisting of an individual's personal and social beliefs, and an external

belief network consisting of different individuals.

We implement the NB theory as a computational model in a statistical physics framework

in which beliefs are represented as spins, potential dissonance is represented as energy, and

attention is represented as temperature. We test our theoretical assumptions in two large survey

studies and show that they are justified (see section Relating Psycho-Social and Statistical

Physics Constructs and Appendices A-C).

We then pose different predictions based on simulations of the internal and external

networks and find empirical support for them using survey data and survey experiments (see

Appendix G). Specifically, we predict that: 1) if individuals pay a lot of attention to their

potential personal dissonance, their average beliefs will show a bimodal distribution, but if they

pay little attention, their average beliefs will be distributed normally, 2) paying attention to

potential personal dissonance leads to more stable beliefs, 3) if individuals pay attention to the

potential dissonance between their personal beliefs and their social beliefs, variance of their

social beliefs will decrease, 4) those individuals will also have social beliefs that are more

consistent with their personal beliefs, 5) when people pay a lot of attention to the dissonance

between their social beliefs (what they believe others think) and what others actually believe,

group radicalization is more likely to occur and individuals are more likely to agree with each

other, and 6) individuals are predicted to have more accurate social beliefs if a belief is discussed

and/or expressed often in a group.

Finally, we show that the computational implementation of NB theory can account for established belief dynamics phenomena including consensus and polarization, group radicalization, and minority influence, providing new parsimonious explanations for these diverse phenomena. We also show that the model can reproduce real-world trends in group polarization over time, and in cross-sectional distribution of beliefs.

Relationship to Other Models

How does our model compare to the numerous previous accounts? Models of belief dynamics can be compared on many different dimensions. Here we focus on structural and process features of models: 1) Does the model represent one or more personal beliefs, 2) If it has more than one belief represented, does it represent them as independent beliefs, in a network, or does it assume a summary representation, 3) Does the model include social influence, and if so, does it include social influence from only one agent at the time or the whole network, 4) Are other's beliefs based on the factual beliefs, or are they represented subjectively, 5) What are the belief updating mechanisms for belief nodes, 6) What are the mechanisms for updating edge weights, 7) What are the mechanisms for updating the social network, 8) To what extent does the model been tested on empirical data. In this section, we will focus on comparing our model with models that are most similar to ours according to these features.

Our model represents several personal beliefs in a network (features 1 and 2 above) together with a subjective representation of others' beliefs (feature 4) and an external factual belief network that can influence the subjective representation of others' beliefs (feature 3). The updating mechanism for nodes follows the minimization of the weighted sum of the personal, social, and external energies (feature 5), which in turn relies on the weighted average of all

beliefs (personal beliefs have unique weights between beliefs while, social and external beliefs relies on one weighting parameter each, feature 7). The edge weights in our model can be derived by the relations in empirical data (6; i.e., partial correlations). In terms of tests on empirical data, we provide several tests of core assumptions (feature 8, see Appendices B and C). Given this feature list, this means that many prominent belief dynamic models will be left out in this comparison. These include models that only represent one belief or vectors of non-interactive beliefs (features 1 and 2), such as French's (1956) formal model of social power, Harary's (1959) generalization of French's model, and DeGroot's (1974) consensus formation model and others that followed in this and other traditions such as the bounded confidence models (Deffuant et al., 2000; Hegselmann & Krause, 2002; Weisbuch et al., 2002), vector models based on attraction or assimilation and rejection or repulsion mechanisms (Flache & Macy, 2011; Huet & Deffuant, 2010; Jager & Amblard, 2005), social influence network theory (e.g., Friedkin & Johnsen 1990; 2011), computational implementations of social impact theory (Nowak et al., 1990), models of the dissemination of culture (Axelrod, 1997), vector models that combine demographic and belief representations (Flache & Mäs, 2008), and several models inspired by statistical physics (e.g., Galesic & Stein, 2019; Pham et al. 2022; for reviews of models inspired by statistical physics see Castellano et al., 2009 and for a general overview of social influence models see Flache et al., 2017).

There are models that represent beliefs as networks (feature 2 above), but only include social influence implicitly (feature 3). For example, the AE framework (Dalege et al., 2016; 2018), which we use as the basis for the representation of the internal belief network in our model, exogenous or social influence can only be represented with the external field parameter (τ in Equation 3 above). A similar model with only non-specific exogenous influences is Brandt

and Sleegers (2021). An influential class of models that also represent beliefs as networks is

constraint satisfaction models. Shultz and Lepper (1996) formulated a constraint satisfaction

model that aims at reducing dissonance in a network of cognitions. In this model, there is no

explicit representation of social influence, but the general formulation of this and other constraint

satisfaction models allows for nodes that represent various external or social influences as they

are represented in the network. A more elaborate neural network model of constraint satisfaction

is Monroe and Read's (2008) attitudes as constraint satisfaction model. Again, there is no

explicit representation or mechanism of social influence or social network in this model. There

are also other neural network models that are not based on constraint satisfaction. For example,

Van Overwalle and Siebler's (2005) model is based on an autoassociative recurrent network that

focuses on assimilating new information instead of the settling of the network in a steady state.

There are not many models that both have a network representation of internal beliefs and

a representation of an external network (features 2 and 3). A model that shares the same basic

assumption about the internal network is the Hierarchical Ising Opinion Model (HIOM; van der

Maas, Dalege, & Waldorp, 2020), only without the social beliefs. In the HIOM, the internal

network of personal beliefs is described by a mean-field approach resulting in modelling these

networks as cusp catastrophes: Strongly involved individuals have resistant beliefs but change

them drastically if they do change, and weakly involved individuals have less resistant beliefs

and change them in a more gradual manner. These dynamics are also implied by the NB theory

as it is an extension of the AE framework (for an illustration of these dynamics implied by the

AE framework see Dalege, Haslbeck, & Marsman, 2022). The HIOM additionally assumes that

involvement increases when individuals interact. A crucial difference between the HIOM and the

NB theory lies in the representation of social beliefs - the HIOM does not differentiate between

actual beliefs and perceived beliefs of others (feature 4). The HIOM therefore cannot explain that

individuals differ in how much importance they attach to their personal versus social beliefs.

This difference also leads to a different coupling between individuals in the HIOM than our NB

theory. While the HIOM assumes that averages of more specific beliefs are communicated

between individuals, the NB theory assumes that beliefs are coupled between individuals through

perceived beliefs. Additionally, the HIOM treats beliefs as binary while the NB theory treats

beliefs as continuous. A potentially fruitful avenue for future research would be to integrate the

dynamics of increasing attention due to their interaction into the NB theory.

Another model that shares features with ours is the social cognitive-social belief model

Rodriguez et al. (2016). The aim of this model is to integrate insights from network models of

social influence and models that represent personal beliefs as networks (features 2 and 3 above),

and as such this is very much in line with the aims of our model. Another similarity to our model

is the reliance on a statistical physics formalism in the sense that the model relies on minimizing

an energy function that combines energies in the internal network with weights attached to each

of the personal and social components in the energy function. The main difference between our

model and the Rodriguez et al.'s model is that their nodes represent concepts and the signed

edges between them are the associative beliefs, while in our model the nodes are the beliefs, and

the edges determine the relation between the beliefs. In contrast to our energy function for the

personal beliefs that takes a weighted average over all pairs of connected nodes, in Rodriguez et

al.'s model the average is taken over the edge weights of all triads. Rodriguez et al. allow for the

possibility that the beliefs are not factual (feature 4), although they do not separate personal

beliefs from social beliefs. The external network in Rodriguez et al.'s model represents the

individuals' and the factual beliefs of these individuals. A comparison of mechanisms of edge

weight updating is not straightforward as the edge weights in Rodriguez et al.'s belief network are the beliefs, and as such they can be updated by accepting a social contact's belief if it decreases their individual energy (feature 6). In terms of empirical tests (feature 8), the simulations of Rodriguez et al.'s model are compared in broad terms to different patterns that have been observed in empirical data, such as how the existence of "zealots" shape belief dynamics and the entrenchment of fringe groups.

Another model that integrates insights from network models of social influence and models that represent personal beliefs as networks (features 2 and 3) is the one presented in Ellinas et al. (2017). In contrast to our model, Ellinas et al.'s model is not based on the statistical physics notion of minimizing energy and is focused on the probability of accepting associations from neighboring individuals (feature 7). The probability that an individual accepts an incoming association from a randomly selected other individual is a function of both the portion of the individual's neighbors that agree with that belief and the social rank difference between the receiver and its neighbors, with different weights allowed for each of these components. In Ellinas et al.'s model, the belief networks and the social rank of individuals are estimated from survey data. In contrast to our model, this model has a mechanism to change the value of edges between personal beliefs by accepting proposed link values by other neighboring individuals (feature 6). However, the node values are static and do not change (feature 5). In Ellinas et al. conduct theoretical explorations of different parameter values, but there are no empirical tests of assumptions or predictions (feature 8).

In contrast to Ellinas et al.'s (2017) model that does not have a mechanism for updating of beliefs, the weighted network balance model described in Schweighofer et al. (2020) also has a mechanism for updating node values (feature 5 above). The model is a computational extension

of the ideas in Heider's (1946) balance theory and Cartwright and Harary's (1956) structural balance theory and is particularly aimed at explaining hyperpolarization, which Schweighofer et al. (2020) defines as the simultaneous occurrence of belief extremeness and issue constraint for multidimensional beliefs. In contrast to our model, where we distinguish between internal and external belief networks, the weighted network balance model considers relations between two individuals A and B and a policy issue, D. The model is a combination of network representations and vector representations (features 1 and 2). The beliefs of any individual are represented with a vector that is all beliefs of an individual towards the policy issue. These interpersonal beliefs in the model do not need to reflect reality but might be purely subjective (feature 4). These vector representations of beliefs are then combined with signed geometrical means. These are then used to determine edge weights between the individuals the policy issue under consideration (feature 6). Beliefs are updated by moving them closer to a perfectly balanced opinion vector. In Schweighofer et al. (2020), the weighted network balance model is tested with data from American National Election Study (feature 8; ANES).

 In sum, our comparison of NB theory and other models suggests three main contributions of the NB theory. It includes both internal beliefs and external social worlds, connected through people's social beliefs or perceived beliefs of others. It explains how these social beliefs can differ from the actual beliefs of others, which is often disregarded in models of belief dynamics. Third, while many of these models implement belief change processes through some form of dissonance, incoherence, or imbalance reduction in internal belief networks or in the relation between beliefs in social networks, they do not explicitly differentiate dissonances due to the consistency of personal and social beliefs and the correspondence of social and actual others' beliefs.

We also make a contribution by providing extensive empirical tests of the core assumptions of our theory as well as of specific empirical predictions. This is very rare, with most models being presented on the conceptual level without empirical tests (Castellano et al., 2009; Redner 2019). Going forward, given the plethora of models that can account for phenomena such as consensus, polarization, and radicalization, an estimate of model mimicry (Wagenmakers et al., 2004) in the area of belief dynamic models would be very helpful to tease out what representations and mechanisms are really needed to explain pertinent phenomena.

Limitations and Possible Extensions

The statistical physics framework that we used here is just one the many analogies that have been used to explore and understand human belief dynamics. Many other analogies have been used for this purpose, including epidemiological models, where transmission of belief is like a transmission of disease (Newman, 2003); percolation, where beliefs seep through a society like liquid through a substance (Duffie et al., 2010); balance, where beliefs and individuals align in a way that leads to most consistent relationships on the level of pairs and triads (Heider, 1958; Pham et al., 2020); expected utility, where beliefs change to maximize the product of value and likelihood of different cognitions (Ajzen, 1991; Fishbein & Ajzen, 1977); evolution, where beliefs evolve in the process of cultural learning (Richerson & Boyd, 2008); Bayesian networks, where networks of beliefs change in line with their conditional dependencies (Cook & Lewandowsky, 2016); forces, where belief change under combined influence of several distinct social forces (Latane, 1981); and networks, where systems of beliefs are conceptualized as networks that aim to minimize the overall energy (Dalege et al., 2016).

While analogies can be very useful to cope with novelty and uncertainty in science and life in general (Holyoak & Thagard, 1996, Dunbar, 1997), it is important to keep in mind that

they were "borrowed' from another domain that does not necessarily correspond to all of the

intricacies of the domain we wish to explain (Gigerenzer, 1991; Olson, 2019), in this case belief

dynamics. We need to recognize the unnecessary "baggage" of worldview, assumptions, and

methodologies that are transferred to the domain of belief dynamics along with the analogy.

In the case of statistical physics analogy, such potential baggage includes its assumptions

that beliefs are like spins, that there is no strategic behavior, no individual differences, no

emotions, and no institutions. Another important baggage is the assumption that people always

want to minimize dissonance among their beliefs, while it is clearly possible that people

sometimes can consciously maintain some dissonance between their beliefs. For example,

scientists are trained to forego changing their beliefs about scientific issue just because they do

not fit with their moral values.

More specifically, in NB theory we assume that the main mechanism for belief change is

dissonance reduction, and we implement it within a statistical physics framework that assumes

that people find belief states that minimizes energy. We want to make clear that we do not

believe that all belief change is caused by an optimization process of one criterion. For a full

understanding of belief dynamics, a comprehensive theory of belief dynamics should include

mechanisms that go beyond the goal of dissonance or energy reduction, although several

mechanisms of belief change can be incorporated in a framework that is assumed to minimize

energy. For example, another form of consistency criterion that has been applied to several belief

dynamic models is triangle balance (e.g., Pham et al., 2022; Rodriguez et al., 2016;

Schweighofer et al., 2020). That is, a triangle of nodes is balanced if it includes zero or two

negative links, otherwise it is unbalanced. This form of consistency mechanism could be

incorporated as a separate term in the energy function (see e.g., Pham et al., 2022). Going

beyond dissonance, balance, and minimization of energy, other mechanisms have been proposed as fundamental for belief change. For example, information integration in the form of summation and averaging can produce belief change (Anderson, 1971). For social and external beliefs, in this view, people are not trying to minimize dissonance or achieve balance, but might just be influenced by people around them without ever considering dissonance or balance.

Another limitation of our statistical physics implementation of NB theory is that it assumes primacy to edge weights in the sense that these influence node values but nodes does not influence edge weights. There is no mechanism in the model that can differentially change the edge weight between two nodes. The only mechanism that can change the values of all edge weight is the attention parameter β. That is, increasing or decreasing the value of β has the same effect as increasing or decreasing all edge weights. Possible extensions could include edge weight updating mechanisms from neural network models of belief dynamics (e.g., Monroe & Read, 2008).

Finally, in the current formulation, NB theory and its implementation is only concerned with beliefs and their dynamics and not the relation between beliefs, preferences, and behavior The beliefs are not at present connected to a preference to choose a specific option or act in certain way. In addition, we have not modeled the strategic considerations related to beliefs. For example, people might choose to signal some beliefs, but not others, and some beliefs might be signaled covertly (van der Does, Galesic, Dunivin, & Smaldino, 2022).

Conclusion

To our knowledge, NB theory is the first model of belief dynamics that connects internal and external belief networks. Unlike many other models of belief dynamics, it does not assume that people perceive others' beliefs accurately, but allows that these perceptions can be shaped

by both personal beliefs and others' actual beliefs. The theory can explain different phenomena in belief dynamics largely by one basic mechanism – attending to different parts of belief networks. We find empirical support for the different assumptions and predictions of the theory in two large studies. Some of the confirmed predictions reproduce previous findings, and others are unique to the NB theory, for example that paying attention to the dissonance between personal and social beliefs decreases the variance of social beliefs and increases their consistency with personal beliefs, and that group radicalization is more likely when people pay attention to the accuracy of their social beliefs. The theory also offers a novel and parsimonious explanation for group consensus, polarization, radicalization, and minority influence, and reproduces empirically observed patterns of beliefs in different studies. We hope that in addition to providing answers to crucial questions on the dynamics of beliefs, the NB theory will inspire further work on the integration of different research areas on belief dynamics and addressing some of the most pressing issues we currently face as a society.

References

Acemoglu, D., & Ozdaglar, A. (2011). Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, *1*, 3-49.

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211.

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica: Journal of the Econometric Society*, 1005-1027.

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, *74*(1), 47.

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological review*, *78*(3), 171.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, *41*(2), 203-226.

Axelrod, R., Daymude, J. J., & Forrest, S. (2021). Preventing extreme polarization of political attitudes. *Proceedings of the National Academy of Sciences*, *118*(50), e2102139118

Bhatia, N., & Bhatia, S. (2021). Changes in gender stereotypes over time: A computational analysis. *Psychology of Women Quarterly, 45(1)*, 106-125.

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science, 16(4)*, 756-766.

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., ... & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers, 1(1)*, 58.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. University of

    Chicago Press.

Brandt, M. J., & Sleegers, W. W. (2021). Evaluating belief system networks as a theory of

    political belief system dynamics. *Personality and Social Psychology Review*, *25*(2), 159-

    185.

Burnstein, E., & Vinokur, A. (1973). Testing two classes of theories about group induced shifts

    in individual choice. *Journal of experimental social psychology*, *9*(2), 123-137.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics.

    Reviews of modern physics, *81*(2), 591.

Card, D., Chang, S., Becker, C., Mendelsohn, J., Voigt, R., Boustan, L., ... & Jurafsky, D.

    (2022). Computational analysis of 140 years of US political speeches reveals more

    positive but increasingly polarized framing of immigration. *Proceedings of the National*

    Academy of Sciences, 119(31), e2120510119.

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory.

    Psychological review, *63*(5), 277.

Charlesworth, T. E., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-

    term change and stability from 2007 to 2016. *Psychological science, 30(2)*, 174-192.

Christakis, N. A., & Fowler, J. H. (2010). Social network sensors for early detection of

    contagious outbreaks. *PloS one*, *5*(9), e12948.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and

    compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), The handbook of social

    psychology (pp. 151–192). McGraw-Hill.

Clifford, P., & Sudbury, A. (1973). A model for spatial conflict. *Biometrika*, *60*(3), 581-588.

Converse, P. E. (1964). The Nature of Belief Systems in Mass Publics. In Ideology and

    Discontent, ed. David Apter. New York: Free Press.

Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief

    polarization using Bayesian networks. *Topics in cognitive science*, *8*(1), 160-179.

Crano, W. D. (2010). Majority and minority influence in attitude formation and attitude change:

    Context/categorization–leniency contract theory. *Minority influence and innovation:*

    Antecedents, processes and consequences, 53-77.

Dalege, J., Borsboom, D., Van Harreveld, F., Van den Berg, H., Conner, M., & Van der Maas,

    H. L. (2016). Toward a formalized account of attitudes: The Causal Attitude Network

    (CAN) model. *Psychological review*, *123*(1), 2.

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. (2018). The attitudinal

    entropy (AE) framework as a general theory of individual attitudes. *Psychological*

    Inquiry, *29*(4), 175-193.

Dalege, J., Haslbeck, J. M. B., & Marsman, M. (2022). Idealized modeling of psychological

    dynamics. In A. M. Isvoranu, S. Epskamp, L. Waldorp, & D. Borsboom (Eds.), Network

    Psychometrics with R.

Dalege, J., & van der Does, T. (2022). Using a cognitive network model of moral and social

    beliefs to explain belief change. *Science Advances*, *8*(33), eabm0137.

David, B., & Turner, J. C. (2001). Majority and minority influence: A single process self-

    categorization analysis. In C. K. W. De Dreu & N. K. De Vries (Eds.), *Group consensus*

    and minority influence: Implications for innovation (pp. 91–121). Blackwell Publishing.

Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting

    agents. *Advances in Complex Systems*, *3*(01n04), 87-98.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical association*, *69*(345), 118-121.

Dhami, M. K., & Olsson, H. (2008). Evolution of the interpersonal conflict paradigm. *Judgment and Decision Making*, *3*(7), 547-569.

Duffie, D., Giroux, G., & Manso, G. (2010). Information percolation. *American Economic Journal: Microeconomics, 2(1)*, 100-111.

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge university press.

Epskamp, S. (2020). Psychometric network models from time-series and panel data. *Psychometrika*, *85*(1), 206-231.

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of statistical software*, *48*, 1-18.

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617.

Ellinas, C., Allan, N., & Johansson, A. (2017). Dynamics of organizational culture: Individual beliefs vs. social conformity. *PloS one*, *12*(6), e0180193.

Festinger, L. A theory of cognitive dissonance. (Stanford University Press, 1957).

Festinger, L. A theory of social comparison processes. Hum. Relations 7, 117–140 (1954).

Fishbein, M., & Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research.

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J.

    (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial*

    Societies and Social Simulation, *20*(4).

Macy, M., & Flache, A. (2011). Social dynamics from the bottom up: Agent-based models of

    social interaction.

Flache, A., & Mäs, M. (2008). How to get the timing right. A computational model of the effects

    of the timing of contacts on team cohesion in demographically diverse teams.

    Computational and Mathematical Organization Theory, *14*(1), 23-51.

Franci, A., Bizyaeva, A., Park, S., & Leonard, N. E. (2021). Analysis and control of agreement

    and disagreement opinion cascades. *Swarm Intelligence, 15(1-2)*, 47-82.

French Jr, J. R. (1956). A formal theory of social power. *Psychological review*, *63*(3), 181.

Friedkin, N. E. (1999). Choice shift and group polarization. *American Sociological Review*, 856-

    875.

Friedkin, N. E., & Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical*

    Sociology, *15*(3-4), 193-206.

Friedkin, N. E., & Johnsen, E. C. (2011). *Social influence network theory: A sociological*

    examination of small group dynamics (Vol. 33). Cambridge University Press.

Gagné, F. M., & Lydon, J. E. (2004). Bias and accuracy in close relationships: An integrative

    review. *Personality and Social Psychology Review*, *8*(4), 322-338.

Galesic, M., Olsson, H., & Rieskamp, J. (2018). A sampling model of social judgment.

    Psychological Review, *125*(3), 363.

Galesic, M., Bruine de Bruin, W., Dalege, J., Feld, S. F., Kreuter, F., Olsson, H., Prelec, D., Stein, D. L., van der Does, T. (2021). Human social sensing is an untapped resource for computational social science. *Nature*, 595, 214-222.

Galesic, M., Olsson, H., Dalege, J., van der Does, T., & Stein, D. L. (2021). Integrating social and cognitive aspects of belief dynamics: Towards a unifying framework. *Journal of the Royal Society Interface*, *18*(176), 20200857.

Galesic, M., & Stein, D. L. (2019). Statistical physics models of belief dynamics: Theory and empirical tests. *Physica A: Statistical Mechanics and its Applications*, *519*, 275-294.

Gawronski, B. (2012). Back to the future of dissonance theory: Cognitive consistency as a core motive. *Social cognition*, *30*(6), 652-668.

Gawronski, B. & Strack, F. Cognitive consistency as a basic principle of social information processing. in Cognitive consistency: A fundamental principle in social cognition (eds. Gawronski, B. & Strack, F.) 1–16 (Guilford Press, 2012).

Gigerenzer, G. From tools to theories: A heuristic of discovery in cognitive psychology. Psychol. Rev. 98, 254–267 (1991).

Glauber, R. J. (1963). Time-dependent statistics of the Ising model. Journal of Mathematical Physics, 4(2), 294–307.

Goel, S., Mason, W., & Watts, D. J. (2010). Real and perceived attitude agreement in social networks. *Journal of personality and social psychology*, *99*(4), 611.

Golub, B. & Sadler, E. (2016). Learning in social networks. In Bramoulle, Y., Galeotti, A. & Rogers, B., The Oxford Handbook of the Economics of Networks. Oxford University Press.

Hammond, R. A. (2015). Considerations and best practices in agent-based modeling to inform policy. In Wallace, R. B., Geller, A. B., & Ogawa, V. A. (Eds.), *Assessing the use of agent-based models for tobacco regulation.* National Academies Press.

Harary, F. (1959). On the measurement of structural balance. *Behavioral Science*, *4*(4), 316-323.

Hegselmann, R. & Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of psychology*, *21*(1), 107-112.

Heider, F. (1958). *The psychology of interpersonal relations.* Wiley.

Hickok, A., Kureh, Y., Brooks, H. Z., Feng, M., & Porter, M. A. (2022). A bounded-confidence model of opinion dynamics on hypergraphs. SIAM Journal on Applied Dynamical Systems, 21(1), 1-32.

Holley, R. A. & Liggett, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *Annals of Probability. 3,* 643–663.

Holyoak, K. J., & Thagard, P. (1996). *Mental leaps: Analogy in creative thought*. MIT press.

Hoppitt, W., & Laland, K. N. (2013). Social learning. In *Social Learning*. Princeton University Press.

Huet, S., & Deffuant, G. (2010). Openness leads to opinion stability and narrowness to volatility. *Advances in Complex Systems*, *13*(03), 405-423.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology*, *50*(6), 1141.

Ising, E. (1925). Contribution to the theory of ferromagnetism. *Z. Phys*, *31*(1), 253-258.

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on

 group polarization. *American journal of political science*, *59*(3), 690-707.

Jackson, M. O. (2008). *Social and economic networks.* Princeton University Press.

Jager, W. & Amblard, F. (2005). Uniformity, bipolarization and pluriformity captured as generic

 stylized behavior with an agent-based simulation model of attitude change.

 Computational & Mathematical Organization Theory, 10(4), 295–303

Kitayama, S., Snibbe, A. C., Markus, H. R., & Suzuki, T. (2004). Is there any "free" choice? Self

 and dissonance in two cultures. *Psychological science*, *15*(8), 527-533.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A

 parallel-constraint-satisfaction theory. *Psychological review*, *103*(2), 284.

Latané, B. (1981). The psychology of social impact. *American psychologist*, *36*(4), 343.

Latané, B., & Wolf, S. (1981). The social impact of majorities and minorities. *Psychological

 Review*, *88*(5), 438.

Leonard, N. E., Lipsitz, K., Bizyaeva, A., Franci, A., & Lelkes, Y. (2021). The nonlinear

 feedback dynamics of asymmetric political polarization. *Proceedings of the National

 Academy of Sciences,* 118(50), e2102149118.

Levin, S. A., Milner, H. V., & Perrings, C. (2021). The dynamics of political polarization.

 Proceedings of the National Academy of Sciences, 118(50), e2116950118.

Macy, M. W., Kitts, J. A., Flache, A., & Benard, S. (2003). Polarization in dynamic networks: A

 Hopfield model of emergent structure.

Martin, R., & Hewstone, M. (Eds.). (2009). *Minority influence and innovation: Antecedents,

 processes and consequences*. Psychology Press.

Mason, L. (2016). A cross-cutting calm: How social sorting drives affective polarization. Public
  Opinion Quarterly, 80, 351–377.

Monroe, B. M., & Read, S. J. (2008). A general connectionist model of attitude structure and
  change: The ACS (Attitudes as Constraint Satisfaction) model. *Psychological review*,
  115(3), 733.

Moscovici, S., Lage, E., & Naffrechoux, M. (1969). Influence of a consistent minority on the
  responses of a majority in a color perception task. *Sociometry*, 365-380.

Newby-Clark, I. R., McGregor, I. & Zanna, M. P. Thinking and caring about cognitive
  inconsistency: When and for whom does attitudinal ambivalence feel uncomfortable? J.
  Pers. Soc. Psychol. 82, 157–166 (2002).

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review,* 45,
  167–256.

Newman, M. E. (2006). Modularity and community structure in networks. Proceedings of the
  national academy of sciences, 103(23), 8577-8582.

Nisbett, R. E., & Kunda, Z. (1985). Perception of social distributions. *Journal of Personality and
  Social Psychology*, *48*(2), 297.

Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: A dynamic
  theory of social impact. *Psychological review*, *97*(3), 362.

Olson, M. E., Arroyo-Santos, A., & Vergara-Silva, F. (2019). A user's guide to metaphors in
  ecology and evolution. *Trends in Ecology & Evolution,* 34(7), 605-615.

Olsson, H., Galesic, M., van der Does, T. (in prep). *Influence of Peers on Beliefs*.

Page, S. E. (2018). The model thinker: What you need to know to make data work for you. Basic
  Books.

Pentland, A. (2014). *Social physics: How good ideas spread-the lessons from a new science*.

    Penguin.

Petty, R. E. & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances*

    in Experimental Social Psychology, 19, 123–205.

Pham, T.M., Kondor, I., Hanel, R., & Thurner, S. (2020). The effect of social balance on social

    fragmentation. *Journal of the Royal Society Interface,* 17(172), 20200752.

Pham, T. M., Korbel, J., Hanel, R., & Thurner, S. (2022). Empirical social triad statistics can be

    explained with dyadic homophylic interactions. *Proceedings of the National Academy of*

    Sciences, 119(6), e2121103119.

Priester, J. R. & Petty, R. E. The gradual threshold model of ambivalence: Relating the positive

    and negative bases of attitudes to subjective ambivalence. J. Pers. Soc. Psychol. 71, 431–

    449 (1996).

Proskurnikov, A. V., & Tempo, R. (2017). A tutorial on modeling and analysis of dynamic social

    networks. Part I. *Annual Reviews in Control*, *43*, 65-79.

Quiamzade, A., Mugny, G., Falomir-Pichastor, J. M., Butera, F., Martin, R., & Hewstone, M.

    (2010). The complexity of majority and minority influence processes. *Minority influence*

    and innovation: Antecedents, processes and consequences, 21-52.

Redner, S. Reality-inspired voter models: a mini-review. C. R. Phys. 20, 275–292 (2019).

Richerson, P. J., & Boyd, R. (2008). *Not by genes alone: How culture transformed human*

    evolution. University of Chicago Press.

Rodriguez, N., Bollen, J., & Ahn, Y. Y. (2016). Collective dynamics of belief evolution under

    cognitive coherence and social conformity. *PLoS one*, *11*(11), e0165910.

Schweighofer, S., Schweitzer, F., & Garcia, D. (2020). A weighted balance model of opinion

   hyperpolarization. *Journal of Artificial Societies and Social Simulation,* 23(3).

Schweighofer, S., Garcia, D., & Schweitzer, F. (2020). An agent-based model of multi-

   dimensional opinion dynamics and opinion alignment. *Chaos: An Interdisciplinary*

   Journal of Nonlinear Science, *30*(9), 093139.

Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction.

   Psychological review, *103*(2), 219.

Smith, C. M., & Tindale, R. S. (2010). Direct and indirect minority influence in groups. In R.

   Martin & M. Hewstone (Eds.) Minority influence and innovation: Antecedents, processes

   and consequences (pp. 263 – 284). New York: Psychology Press.

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological*

   review, *99*(1), 3.

Sunstein, C. R. (1999). The law of group polarization. *University of Chicago Law School, John*

   M. Olin Law & Economics Working Paper, (91).

Sunstein, C. R., Schkade, D., Ellman, L. M., & Sawicki, A. (2007). *Are judges political?: an*

   empirical analysis of the federal judiciary. Brookings Institution Press.

Thomas, W. I. & Swaine Thomas, D. (1928). *The child in America: Behaviour problems and*

   programs. Knopf.

Tomasello, M. Becoming Human: A Theory of Ontogeny (Belknap, 2019).

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of

   group behavior. In E. Lawler (Ed.), *Advances in group processes* (pp. 77–121).

   Greenwich, CT: JAI Press.

Vallacher, R. R., Read, S. J., & Nowak, A. (Eds.). (2017). *Computational social psychology*. Routledge.

van der Does, T., Stein, D. L., Fedoroff, N., & Galesic, M. Moral and Social Foundations of Beliefs about Scientific Issues: Predicting and Understanding Belief Change. doi: 10.31219/osf.io/zs7dq.

Van Der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. Psychological review, 113(4), 842.

van der Maas, H. L., Dalege, J., & Waldorp, L. (2020). The polarization within and across individuals: The hierarchical Ising opinion model. Journal of complex networks, 8(2), cnaa010.

van Harreveld, F., van der Pligt, J. & de Liver, Y. N. The agony of ambivalence and ways to resolve it: Introducing the MAID model. Personal. Soc. Psychol. Rev. 13, 45–61 (2009).

Van Overwalle, F., & Siebler, F. (2005). A connectionist model of attitude formation and change. *Personality and Social Psychology Review*, *9*(3), 231-274.

Wagenmakers, E. J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology*, *48*(1), 28-50.

Watts, D. J. (2004). Six degrees: The science of a connected age. Norton.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. Journal of Personality and Social Psychology, 67(6), 1049–1062.

Weisbuch, G., Deffuant, G., Amblard, F., & Nadal, J. P. (2002). Meet, discuss, and segregate!. *Complexity*, 7(3), 55-63.

Wood, W., Lundgren, S., Ouellette, J. A., Busceme, S., & Blackstone, T. (1994). Minority influence: a meta-analytic review of social influence processes. *Psychological bulletin*, *115*(3), 323.

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on twitter. Bulletin of science, technology & society, 30(5), 316-327.

Appendix A

Study Descriptions

Study 1 was a survey of 973 U.S. participants from Mechanical Turk, conducted from May 11-13, 2022. Among the participants, 55% were male, and 44.6% female; mean/min/max age was 40.6/10/78 years; 9.6% had high school or less, 30.8% some college, and 59.6% college degree. They answered questions (see full text in Supplementary Materials) about: 1) their personal beliefs about safety of GM food and related moral and political beliefs (personal belief nodes); 2) importance of consistency of their personal beliefs (attention to potential personal dissonance), 3) experience of dissonance in their personal beliefs (felt personal dissonance), 4) their social beliefs about three beliefs about GM safety held by each of five of their social contacts (social belief nodes); 5) importance of consistency of their personal and social beliefs (attention to potential social dissonance); 6) experience of dissonance between their personal and social beliefs (felt social dissonance); 7) importance of accuracy of social beliefs (attention to potential external dissonance); 8) frequency of contact with social contacts; 9) overall importance of the topic of GM food; 10) personal beliefs about safety of GM food after receiving a brief informational intervention (a quote from a National Academies of Sciences, Engineering, and Medicine report saying that there is no evidence GM food currently on the market is harmful when consumed); and 11) basic demographics. All questions except for the demographics were answered on 7-point scales with labeled extremes. The study was approved by University of New Mexico IRB #10819. This study was not preregistered.

Study 2 was a two-wave survey conducted on a probabilistic national sample of U.S. population from October 13-December 15, 2022, by the Center for Economic and Social Research at the University of Southern California. Participants, who were members of the USC's

Understanding America Panel, answered questions about their personal and social beliefs related to GM food, flu vaccination, and climate change. The social beliefs were about a friend they nominated themselves from a different household. This friend was invited to complete a short survey about their actual personal beliefs. Participants also answered questions about the attention to and feelings of social and external dissonance as well as questions about their interaction with their social contacts and their experience with scientists. They received an informational intervention which consisted of seeing their friend's actual answers as well as scientists' views on GM food, flu vaccination, and climate change. In total, 669 participants completed both survey waves, and each had a friend completing the short survey as well. Most of the results will be presented elsewhere (Olsson et al., in prep), and here we focus on the participants' answers about their 1) personal and 2) social beliefs; their 3) felt social and 4) felt external dissonance; and 5) their attention to potential external dissonance (see those questions in the Supplementary Materials). The study was approved by University of Southern California (BRANY SBER) IRB #22-065-1044. This study was not preregistered.

Appendix B

Energy as Potential Dissonance, and Separability of Energies: Validation Tests

A central assumption of the NB theory is that the statistical physics concept of energy provides a formalization of potential dissonance. If this is the case, then potential dissonance should be related to subjectively measured felt dissonance, provided that participants pay some attention to the potential dissonance. Furthermore, we assume that people can differentiate between potential dissonances in different parts of their belief networks. If this is correct, then felt personal dissonance should be more related to potential personal dissonance than to potential social dissonance, and vice versa for the felt social dissonance. Similarly, felt social dissonance should be more related to potential social dissonance than to potential external dissonance, and vice versa for the felt external dissonance.

To test the assumption about the relationships of potential and felt personal and social dissonances, we analyze questions about Study 1 participants' personal beliefs about GM food and about perceived beliefs of five of their social contacts. These questions enabled us to calculate potential personal and social dissonances for each participant, as described below. The survey also included measures of felt dissonances among personal and social beliefs, adapted from the felt ambivalence scale developed by Priester and Petty (1996; see Supplementary Materials for more details). For example, one of three items measuring personal dissonance was "I experience no conflict at all towards the issue of GM food" and for social dissonance "I experience no conflict at all between my beliefs and the beliefs of [*social contact]* towards the issue of GM food", on the scale from 1-7.

We checked whether the personal and social felt dissonances are indeed two different psychological constructs, by investigating whether the three items measuring each construct

loaded on two separate factors. To test whether the personal and social felt dissonance items

loaded on two separate factors, we fitted a confirmatory factor model with two correlated factors.

Items of the different scales were only allowed to load on their respective factor. This model did

not fit the data well, *RMSEA* = .15, *CFI* = .93. We therefore investigated the reasons for the bad

fit of the model and found that this was generally caused by higher correlations between the

negatively worded items of the personal and social felt dissonance scales than would be expected

by the model. We therefore allowed for correlated errors between these items. This resulted in a

model with good fit to the data, *RMSEA* = .08, *CFI* = .98. In this model, personal and social felt

dissonances were moderately correlated, *r* = .32, *p* < .001. We then tested whether a two-factor

model fits the data better than a one-factor model (we allowed for correlated errors between the

negatively-worded items in both these models). The one-factor model showed poor fit to the

data, *RMSEA* = .29, *CFI* = 73, and also showed worse fit than the two-factor model, *AIC* =

21,980.53 vs. *AIC* = 21,368.35 for the two-factor model, *BIC* = 22,043.98 vs. *BIC* = 21,436.68

for the two-factor model. The results thus indicate that personal and social felt dissonances form

two separate factors, which are moderately related. Both these findings are in line with the

assumptions of our theory.

We then proceeded to test whether personal and social felt dissonances relate to personal

and social potential dissonances. We expected that personal felt dissonances predict personal

energies and that they do so better than social dissonances. Conversely, we expected that social

dissonances predict social energies and that they do so better than personal dissonances. To

calculate potential personal and social dissonance, we first estimated edges between personal

beliefs (ω_{ij}) and between personal and social beliefs (ρ_{ik}), consisting of the average of one's

own beliefs and the average of the social beliefs for each of five social contacts. Edges represent

regularized partial correlations, which were estimated using the *EBICglasso* function in the R-package *qgraph* (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012; Epskamp & Fried, 2018). Using these networks, we calculated potential dissonance for each participant using Equations 3 and 4, and setting $\tau_i = 0$. We then added these energies to the factor model described above, and regressed each of them on personal and social felt dissonances. Personal energies were significantly and moderately to strongly predicted by personal felt dissonances, $\beta = .42$, $p < .001$, but not by social felt dissonances, $\beta = .05$, $p = .14$.   Social energies were significantly and strongly predicted by social dissonances, $\beta = .55$, $p < .001$, but only weakly by personal dissonances, $\beta = .12$, $p < .001$. To summarize, we found that personal and social felt dissonances load on two separate factors, these dissonances are moderately related to each other, and they are more related to their associated potential dissonances (energies) than to the energies stemming from the other part of the internal belief network. Taken together, these results indicate that indeed felt dissonances arise from lack of consistency between beliefs, and are separate for personal and social beliefs.

To test the assumption about the relationships of potential and felt social and external dissonances, we analyze questions about Study 2 participants' personal and social beliefs about climate change, GM food, and vaccination. Here, because of time constraints in the survey we measured only one personal belief for each topic (e.g. "What comes closer to your view on climate change? 1=There is solid evidence that the climate is NOT changing because of human activity, 7=There is solid evidence that the climate is changing because of human activity"). Similarly, we used just one question per topic to measure social felt dissonance (e.g., "I experience a lot of conflict between my beliefs about climate change and the beliefs of [friend].") and external felt dissonance (e.g., "I feel uneasy about the discrepancy between what I thought

[friend] believes about climate change and what [friend] actually believes."). We measure potential social dissonance as the absolute difference between personal beliefs and actual friend's beliefs, and external social dissonance as the absolute difference between social beliefs and actual friend's beliefs.

We use a mixed-effects regression models to investigate how social and external felt dissonance relates to social and external potential dissonances. In the model for each of the two felt dissonances, we also control for the other felt dissonance, as well as for the clustering of the three topics within participants. As expected, social potential dissonance was reliably positively related to social felt dissonance ($\beta = .07$, $p < .001$) but not to external felt dissonance ($\beta = -.01$, $p = .43$). And, external potential dissonance was positively, although not reliably related to external felt dissonance ($\beta = .03$, $p = .12$) and negatively and not reliably to social felt dissonance ($\beta = -.01$, $p = .59$).

Appendix C

Attention to Dissonance as Temperature, and Separability of Temperatures: Validation

Tests

Another central assumption of the NB theory is that there are three separate types of attention (or temperature in statistical physics terms), one each for the dissonance in personal, social, and external belief networks. To test this assumption, in Study 1 (see Appendix A) we also included questions tapping participants' subjective importance of consistent personal beliefs (a proxy for temperature of the personal part of the internal network, for example "It is important to me that my beliefs toward GM food are not in conflict with each other"), importance of consistent social beliefs (a proxy for temperature of the social part of the internal network, for example "It is important to me that my personal beliefs and the beliefs of *social contact* toward GM food are not in conflict with each other."), and importance of accurate social beliefs (a proxy for temperature of the external network, for example "It is important to me that I know what *social contact* thinks about GM food.").

We first test whether the importance items load on three separate factors corresponding to personal, social, and external belief networks. We first tested whether the importance items from Study 1 load on three separate factors. To do so, we fitted a confirmatory factor model with three correlated factors. Items of the different scales were only allowed to load on their respective factor. This model did not fit the data well, *RMSEA* = .20, *CFI* = .89. We therefore investigated the reasons for the bad fit of the model and found that this was generally caused by higher correlations between similarly worded items of the importance of consistent social beliefs scale and the importance of accurate social beliefs scale than would be expected by the model. We therefore allowed for correlated errors between these items. This resulted in a model with good

fit to the data, *RMSEA* = .07, *CFI* = .99. In this model, the importance of consistent personal beliefs was strongly correlated with the importance of consistent social beliefs, *r* = .53, *p* < .001, and with the importance of accurate social beliefs, *r* = .51, *p* < .001. The importance of consistent social beliefs was very strongly correlated with the importance of accurate social beliefs, *r* = .77, *p* < .001.

To test whether indeed three factors are needed to explain the different importances, we first compared the three-factor model with a two-factor in which the importance of consistent personal beliefs items loaded on the first factor and the importance of consistent social beliefs and the importance of accurate social beliefs items loaded on the second factor (both models included correlated errors between similarly worded items). The two-factor did not fit the data well, *RMSEA* = .27, *CFI* = .81, and also showed worse fit than the three-factor model, *AIC* = 27,348.60 vs. *AIC* = 25,800.25 for the three-factor model, *BIC* = 27455.97 vs. *BIC* = 25,917.37 for the three-factor model. We then proceeded to test whether a one-factor model fitted the data. This model did not fit the data well, *RMSEA* = .35, *CFI* = .66, and also showed worse fit than the three-factor model, *AIC* = 28,627.50, *BIC* = 28,729.99. We therefore conclude that three factors are needed to explain the correlational patterns in the data. This finding provides first support that there are indeed three separate forms of temperatures in belief networks.

Furthermore, we investigate whether higher attention to dissonance predicts a reduced potential dissonance (energy). This is what would be expected from the NB theory: when people pay attention to their potential dissonance (that is, when they experience felt dissonance), their beliefs will tend to become more consistent, reducing the potential dissonance. We find that attention to dissonance in a particular part of belief network relates negatively to the potential dissonance (energy) in that part of the network. For personal and social beliefs, we use measures

of importance of personal and social potential dissonances collected in Study 1 as proxies for attention, and find the expected negative relationship of attention and potential dissonances (β = -.21, p < .001 for the personal beliefs, and β = -.17, p = .002 for the social beliefs). For external beliefs, as proxies for attention we use measures of importance of external dissonance as well as of frequency of discussing different topics with a friend, collected in Study 2. We again find the expected negative relationship between external potential dissonance (the difference between perceived and actual beliefs of friends) and both proxies for attention (β = -.06, p = .002 for importance and β = -.24, p < .001 for frequency of discussion).

While in Study 1 we could not measure the energy in external networks (as we did not have data about the actual beliefs of participants' social contacts), we can predict that the importance of accurate social beliefs might *increase* energies of the social part of the internal network. The reason is that, assuming at least moderately heterogeneous social networks, higher accuracy of one's social beliefs would lead to higher inconsistency of one's social and personal beliefs. We find weak support for this hypothesis (β = .09, p = .10), possibly because social networks of our participants were homogeneous. Another reason might be that external temperature is more strongly affected by how often a belief is expressed and/or discussed in a social network.

Appendix D

Measurement of Theoretical Constructs

There are several different ways to measure or estimate the core constructs and predictions specified in the NB theory. First, the state of nodes in belief networks can be measured through survey questionnaires (Dalege et al., 2016; van der Does et al., 2022), by inferences from behavior (e.g., from voting records), or by inferring networks of related beliefs from textual corpora such as books, newspapers, congressional speeches, and social media (Bhatia & Bhatia, 2021; Card et al., 2022; Charlesworth & Banaji, 2019). Edges between belief nodes can be estimated using partial correlations or regression weights based on co-occurrences of beliefs within or between participants at a single time point or over time or by other methods for fitting parameters of network models (Borsboom et al., 2021). Additionally, edges could also be set by asking participants how strongly their beliefs are related. Second, energy can be estimated by combining the state of nodes with the estimated network structure and using the equations in the section Formal Implementation to calculate potential dissonance between beliefs in different parts of the belief network (cf., Dalege & van der Does, 2022). Third, temperature or attention to potential dissonance can be estimated by comparing belief networks of different groups (Epskamp, 2020), or by directly asking individuals about the importance they attach to their beliefs and the dissonance between them (as we do here).

Appendix E

The Choice of Distance Measures

In Figure E1, we present results comparing different distance measures.
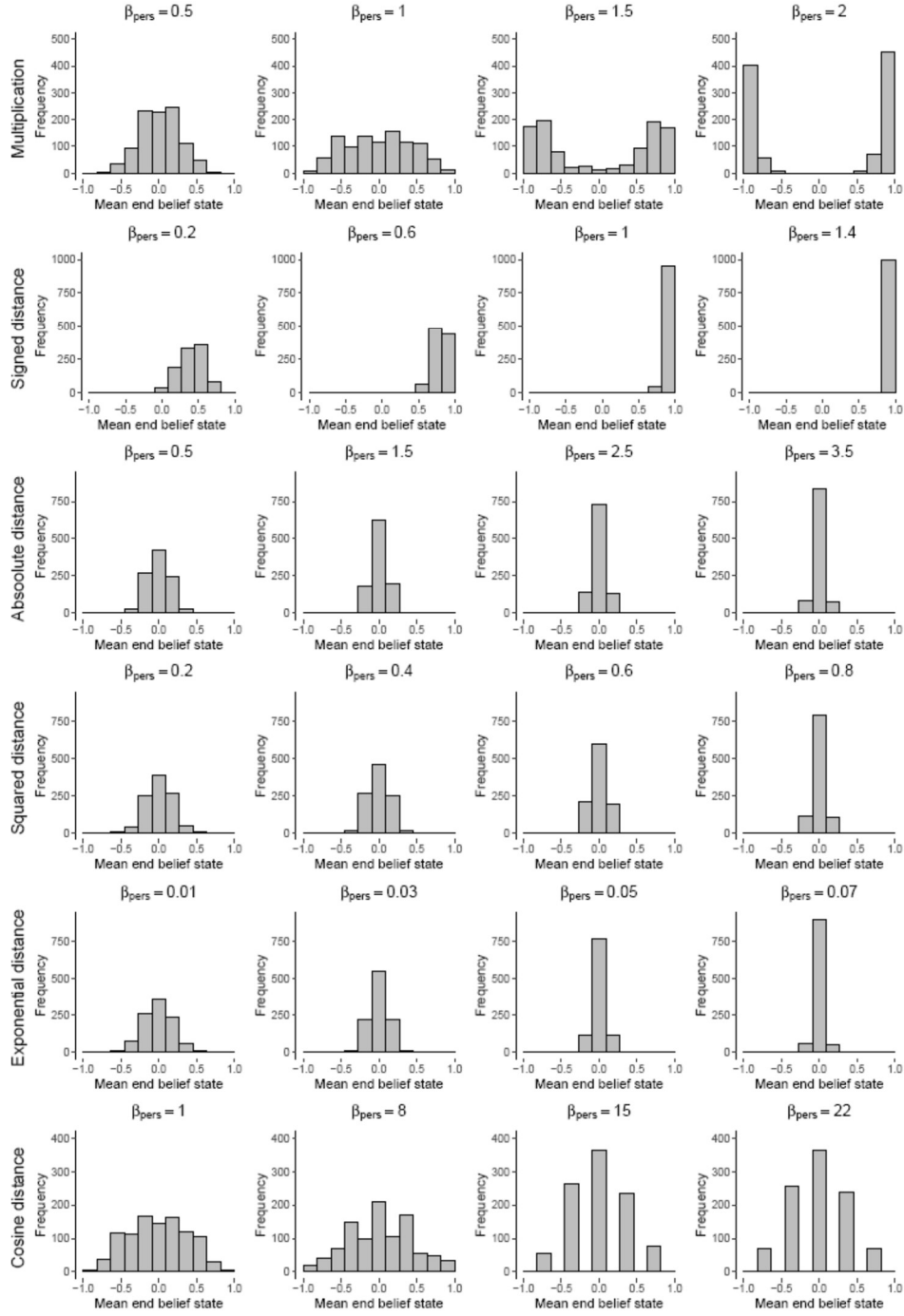
Figure E1. Steady state distributions of different implementations of energy minimization (distance measures) under different inverse temperatures. The simulations were run on a network of 10 belief nodes, using the settings described in the Formal Implementation section. The first row represents the implementation we use in this paper, relying on multiplication. The second row represents calculating energies using signed distance (Akerlof, 1997). The third row represents calculating energies using absolute distance (Akerlof, 1997). The fourth row represents calculating energy using squared distance (Deffuant et al., 2000). The fifth row represents calculating energy using exponential distance (Smith & Zarate, 1992). The sixth row represents calculating energy using cosine distance (Sirbu, Loreto, Servedio, & Tria, 2013). As can be seen, only the multiplication implementation can reproduce the increasing bimodality that we see in empirical data (see Figures 6 and G1).

Appendix F

Determining Edge Weights

To determine the edge weights used in all simulations reported in this paper, we initialized beliefs randomly and then allowed them to settle in the steady state achievable for $\beta = 1$. We then calculated correlations over runs between one of the personal beliefs (the focal belief) and all the other personal beliefs, between the focal belief and the social beliefs, and between social and actual beliefs. We searched through different values of edge weights until we found those that lead to a moderate size of all those three types of correlations. In this way we aimed to establish an intermediate case where all parts of the belief network have similar correlations given a moderate attention to potential dissonance, enabling us to compare effects of lower or higher attention under similar conditions for all parts of the belief network. In other words, attention to potential dissonance of $\beta = 1$ represents the baseline for each part of the belief

network and results in beliefs belonging to that part of the network being moderately correlated. A β lower than 1 will result in less correlated beliefs and a β higher than 1 will result in more strongly correlated beliefs. The resulting edge weights, used in all simulations, are $\omega = .4$ for the influence between personal beliefs, $\rho = 1$ for the influence between personal and social beliefs, and $\alpha = 1.4$ for the correspondence between social and actual beliefs.

Appendix G

Empirical Tests of Implications of Directing Attention to Personal, Social, or External

Dissonance

Implications of Directing Attention to Personal Dissonance

The first implication of directing attention to personal dissonance is that individuals, who pay a lot of attention to their personal dissonance will show a bimodal distribution of their mean beliefs, and those who pay little attention will show a normal distribution of the mean beliefs. In other words, those for whom consistency of personal beliefs is important will show more extreme beliefs on average than those for whom consistency is not that important. We tested this prediction in our Study 1, by analyzing distributions of the average of 14 personal beliefs about GM, separately for participants who reported lower or higher importance of consistency of their personal beliefs. As described in Appendix C, the importance was measured by three items, evaluated on a 7-point scale (see Supplementary Materials for the full text of questions). Results shown in the top row of Figure F1 suggest that participants who assigned very high importance to their personal consistency (and presumably directed a lot of attention to their potential personal dissonance) had indeed a more bimodal distribution of average beliefs than those who assigned lower importance to their personal dissonance.
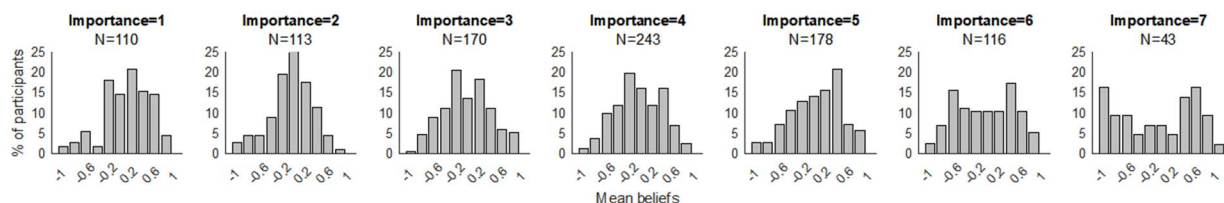
Figure G1. Distribution of average beliefs for participants in Study 1 who assigned different levels of importance to their potential personal dissonance (consistency of their personal beliefs).

Results from Study 1, N = 973.

The second implication of directing attention to personal dissonance is that directing attention to personal dissonance will lead to a higher stability and resistance of beliefs. This is because, according to the NB theory, more attention will lead to more consistent and extreme beliefs. Being in this extreme and consistent position will require an exceptionally strong intervention to change beliefs of those individuals. Therefore, we would expect little change for most people with high attention to personal dissonance (though if they change, they should change radically).

To examine this prediction, we analyzed the extent of change from before to after the educational intervention implemented in Study 1. As described before, participants were asked about their personal beliefs about safety of GM food (using three questions, about whether GM food is safe, harmful, and natural) at the beginning of the study, and then again at the end after they first received a brief statement from a recent National Academies of Sciences, Engineering, and Medicine report, stating that there is no evidence that GM is harmful. As a proxy for attention to personal dissonance, they were also asked about the importance of consistency of their personal beliefs.

We then fit a Structural Equation Model (SEM) to examine the relationship between importance of reducing personal dissonance (as rated by participants and here serving as a measure of attention to dissonance), potential dissonance between each participant's personal beliefs, and belief change after the educational intervention. The SEM model specifying these paths showed good fit to the data, *RMSEA* = .07, *CFI* = .99.

We found that attention to personal dissonance (measured as importance) predicts

potential dissonance of personal beliefs (see Appendix C). More crucially, we found that belief

change after the intervention is positively predicted by potential dissonance for the personal

beliefs, $\beta = 0.28$, $p < .001$. In other words, participants who thought that reducing personal

dissonance is important had less dissonance in their belief networks before the intervention and

have changed less after the intervention.

Implications of Directing Attention to Social Dissonance

The implications of directing attention to social dissonance are that (a) people who pay

more attention to social dissonance will have lower variance of social beliefs and (b) their

personal and social beliefs will be more consistent. To test these predictions, we calculate the

variance of social beliefs (across the five social contacts) that participants reported in Study 1, as

well as the average difference of each of the social beliefs and personal beliefs about the same

issues. As Figure G2 suggests, both the variance (left panel) and the difference of personal and

social beliefs (right panel) tend to be smaller for those participants who reported a higher

importance of social consistency (more attention to the social dissonance).
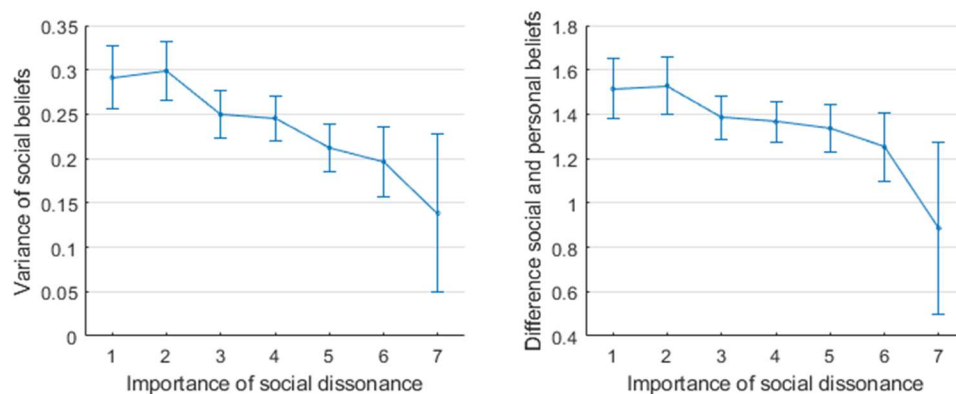
Figure G2. Top row: variance of social beliefs about five social contacts, for participants

assigning different levels of importance to their potential social dissonance (consistency of their

personal and social beliefs). Bottom row: average difference between personal and social beliefs

for the same groups of participants. Results from Study 1, N = 973.

Implications of Directing Attention to External Dissonance

The first implication of directing attention to external dissonance is that groups of people who pay a lot of attention to their external dissonance (that is, to the correspondence between their social beliefs and others' actual beliefs), will be more likely to show higher levels of radicalization. To investigate this, we analyze data about the extremity of participants' beliefs in the first wave of Study 2 and regress them on our two proxies of attention to external dissonance: the importance of accurately perceiving friend's beliefs and frequency of discussing a topic with the friend. In a mixed model adjusting for clustering of participants within, we find as expected that both of these proxies for attention were positively related to more radical beliefs (for importance, $\beta = .08$, $p < .001$, for frequency, $\beta = .08$, $p = .02$). These results remain robust after including interactions with topics and controlling for felt social dissonance.

The second implication of directing attention to external dissonance is that individuals should have more accurate perceptions of their peers' beliefs if a belief is discussed and/or expressed often in a group, was also confirmed. As mentioned in the Appendix C, we find that accuracy of perceived friend's beliefs is higher (i.e., external dissonance lower) for participants who report higher frequency of discussing these beliefs with the friend.

Supplementary Materials

Question Texts

Study 1

1) Personal beliefs

For each of the following considerations, select a number between 1 and 7 that most closely represents your opinion.

Would you say that…

GM food is not part of our tradition, or that it is part of our tradition?

1 Not part of our tradition 7 Part of our tradition

Growing GM food is harmful or beneficial to the environment?

1 Harmful to the environment   7 Beneficial to the environment

GM food negatively or positively affects your family?

1 Negatively affects my family 7 Positively affects my family

GM food is not approved by all the appropriate agencies, or that it is approved by all the appropriate agencies?

1 Not approved by all the appropriate agencies   7 Approved by all the appropriate government agencies

Some important information about GM food is not shared with the public, or that all important

1 Some important information is not shared 7 All important information is shared

Big biotechnology companies benefit more from GM food than farmers, or that both big biotechnology companies and farmers benefit from GM food?

1 Big biotechnology companies benefit more   7 Both big biotechnology companies and farmers benefit

Large-scale farmers benefit more from growing GM food than small-scale farmers, or that both large-scale and small-scale farmers benefit from growing GM food?

1 Large-scale farmers benefit more 7 Both large-scale and small-scale farmers benefit

It is very unnatural or very natural to eat GM food?

1 Very unnatural 7 Very natural

You are forced to eat GM food without your consent, or that you are free to choose whether to eat GM

God disapproves or approves of GM?

1 God disapproves 7 God approves

GM food is very harmful or beneficial to children?

1 Harmful to children   7 Beneficial to children

GM food negatively or positively affects our country?

1 Negatively affects our country   7 Positively affects our country

It is very unsafe or very safe to eat GM food?

1 Unsafe7 Safe

It is harmful or not harmful to eat GM food?

1 Harmful   7 Not harmful

Do you consider yourself liberal or conservative?

1 Very liberal   7 Very conservative

2) Importance of consistency of personal beliefs (attention to potential personal dissonance)

Please rate the following statements about your beliefs toward GM food.

It is important to me that my beliefs toward GM food are not in conflict with each other.

1 Strongly disagree   7 Strongly agree

It is important to me that all my beliefs toward GM food point in the same direction.

1 Strongly disagree   7 Strongly agree

It bothers me if I have mixed reactions toward GM food.

1 Strongly disagree   7 Strongly agree

3) Experience of dissonance in personal beliefs (felt personal dissonance)

Please rate the following statements about your opionions about GM food.

I experience no conflict at all towards the issue of GM food.

1 Strongly disagree   7 Strongly agree

I experience no indecision at all towards the issue of GM food.

1 Strongly disagree   7 Strongly agree

I have completely mixed reactions toward the issue of GM food.

1 Strongly disagree   7 Strongly agree

4) Social beliefs

Now, we would like you to think about five individuals who are most important to you when it comes to issues like GM food. These could be your friends, family, colleagues, or other acquaintances. We will call these your social contacts. Please fill in their Contacts below. On the next pages we will ask you questions about these individuals using these Contacts. It is therefore important that you use Contacts that you will remember while answering the questions.

Now think about the opinions of [Contact 1]. For each of the following considerations, select a number between 1 and 7 that most closely represents the opinion of [Contact 1].

Do you think [Contact 1] would say that…

It is very unsafe or very safe to eat GM food?

1 Very unsafe   7 Very safe

It is very harmful or not harmful at all to eat GM food?

1 Very harmful   7 Not harmful at all

It is very unnatural or very natural to eat GM food?

1 Very unnatural 7 Very natural

[repeated for all five social contacts]

5) Importance of consistency of personal and social beliefs (attention to potential social dissonance)

Now think about [Contact 1] and rate the following statements.

It is important to me that my personal beliefs and the beliefs of [Contact 1] toward GM food are not in conflict with each other.

1 Strongly disagree   7 Strongly agree

It is important to me that my personal beliefs and the beliefs of [Contact 1] toward GM food point in the same direction.

1 Strongly disagree   7 Strongly agree

It bothers me if [Contact 1] and I have mixed reactions toward GM food.

1 Strongly disagree   7 Strongly agree

[repeated for all five social contacts]

6) Experience of dissonance between personal and social beliefs (felt social dissonance)

Please rate the following statements about your opinions and the opinions of [Contacts 1-5] towards GM food.

I experience no conflict at all between my opinions and the opinions of [Contacts 1-5] towards the issue of GM food.

1 Strongly disagree   7 Strongly agree

The opinions that [Contacts 1-5] have about GM food and my opinions point all in the same direction.

1 Strongly disagree   7 Strongly agree

[Contacts 1-5] and I have completely different reactions towards the issue of GM food.

1 Strongly disagree   7 Strongly agree

7) Importance of accuracy of social beliefs (attention to potential external dissonance)

Now think about [Contact 1] and rate the following statements.

It is important to me that I know what [Contact 1] thinks about GM food.

1 Strongly disagree   7 Strongly agree

It is important to me that I accurately perceive the beliefs of [Contact 1] toward GM food point.

1 Strongly disagree   7 Strongly agree

It bothers me if [Contact 1] has different beliefs toward GM food than I thought.

1 Strongly disagree   7 Strongly agree

[repeated for all five social contacts]

8) Frequency of contact with social contacts

About how often do you have any kind of contact with [Contact 1]?

Several times a day - Once a day - Several times per week - Once a week - Several times per month - Several times per year - More rarely

[repeated for all five social contacts]

9) Overall importance of the topic of GM food

Please rate the following statements about your opinions about GM food.

I care a great deal about my opinions about GM food.

1 Strongly disagree   7 Strongly agree

I think a lot about issues related to GM food.

1 Strongly disagree   7 Strongly agree

The issue of GM food is very important to me.

1 Strongly disagree   7 Strongly agree

I am very certain in my different opinions about GM food.

1 Strongly disagree   7 Strongly agree

10) Personal beliefs after receiving a brief informational intervention

Here is some information about GM food. Please read it carefully.

There is no evidence that GM food currently on the market is harmful when consumed by people or farm animals.

This information comes from a summary of recent studies compiled by the National Academies of Science,

Engineering, and Medicine.

The production of GM crops imitates natural farming techniques. These studies explain how all food crops have

been modified from their wild relatives by farmers for thousands of years.

Please rate the following considerations again for what you personally think. For each of the following

considerations, select a number between 1 and 7 that most closely represents your opinion.

Would you say that…

It is very unsafe or very safe to eat GM food?

1 Very unsafe   7 Very safe

It is very harmful or not harmful at all to eat GM food?

1 Very harmful   7 Not harmful at all

It is very unnatural or very natural to eat GM food?

1 Very unnatural 7 Very natural

11) Basic demographics

Please answer just a few more questions about yourself.

What is your gender?

       Man – Woman - Other

What is your age?

___ years

What is your highest level of education?

Less than high school - High school or GED - Some College - Associate's degree – Bachelor - Graduate

Study 2

1) Personal beliefs

What comes closer to your view on climate change?

1 There is solid evidence that the climate is NOT changing because of human activity   7 There is solid evidence that the climate is changing because of human activity

Do you think it is generally unsafe or safe to eat genetically modified foods?

1 Generally unsafe   7 Generally safe

Thinking about childhood diseases, such as measles, mumps, rubella and polio, what comes closer to your view?

1 Parents should be allowed to choose to NOT vaccinate their children 7 All children should be required to vaccinate

2) Social beliefs

When answering the question What comes closer to your view on climate change?' what do you think [Contact] would answer?

1 There is solid evidence that the climate is NOT changing because of human activity   7 There is solid evidence that the climate is changing because of human activity

When answering the question 'Do you think it is generally unsafe or safe to eat genetically modified food?' what do you think [Contact] would answer?

1 Generally unsafe   7 Generally safe

When answering the question Thinking about childhood diseases, such as measles, mumps, rubella and polio, what comes closer to your view?' what do you think [Contact] would answer?

1 Parents should be allowed to choose to NOT vaccinate their children 7 All children should be required to vaccinate

3) Experience of dissonance between personal and social beliefs (felt social dissonance)

I experience a lot of conflict between my beliefs about climate change and the beliefs of [contact].

1 Strongly disagree   7 Strongly agree

I experience a lot of conflict between my beliefs about genetically modified foods and the beliefs of [contact].

1 Strongly disagree   7 Strongly agree

I experience a lot of conflict between my beliefs about childhood vaccination and the beliefs of [contact].

1 Strongly disagree   7 Strongly agree

4) Experience of dissonance between social and actual contact's beliefs (felt external dissonance)

I feel uneasy about the discrepancy between what I thought [Contact] believes about climate change and what [Contact] actually believes.

1 Strongly disagree   7 Strongly agree

I feel uneasy about the discrepancy between what I thought [Contact] believes about genetically modified foods climate change and what [Contact] actually believes.

1 Strongly disagree   7 Strongly agree

I feel uneasy about the discrepancy between what I thought [Contact] believes about childhood vaccination and what [Contact] actually believes.

1 Strongly disagree   7 Strongly agree

5) Importance of accuracy of social beliefs and frequency of discussion (attention to potential external dissonance)

It is important to me that I accurately perceive the belief of [Contact] about climate change.

1 Strongly disagree   7 Strongly agree

It is important to me that I accurately perceive the belief of [Contact] about genetically modified (GM) food.

1 Strongly disagree   7 Strongly agree

It is important to me that I accurately perceive the belief of [Contact] about childhood vaccination.

1 Strongly disagree   7 Strongly agree

About how often do you discuss climate change with [Contact] ?

1 Every day 2 A few times per week 3 A few times per month 4 A few times per year 5 More rarely

About how often do you discuss genetically modified (GM) food with [Contact] ?

1 Every day 2 A few times per week 3 A few times per month 4 A few times per year 5 More rarely

About how often do you discuss childhood vaccination with [Contact] ?

1 Every day 2 A few times per week 3 A few times per month 4 A few times per year 5 More rarely