UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

The Attitudinal Entropy (AE) Framework as a General Theory of Individual Attitudes

Dalege, J.; Borsboom, D.; van Harreveld, F.; van der Maas, H.L.J.

DOI
10.1080/1047840X.2018.1537246

Publication date
2018

Document Version
Final published version

Published in
Psychological Inquiry

License
CC BY-NC-ND

Link to publication

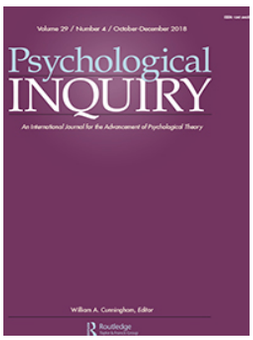
Citation for published version (APA):
Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2018). The Attitudinal Entropy (AE) Framework as a General Theory of Individual Attitudes. *Psychological Inquiry*, *29*(4), 175-193. https://doi.org/10.1080/1047840X.2018.1537246

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 Mar 2023



Routledge
Taylor & Francis Group

Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

ISSN: 1047-840X (Print) 1532-7965 (Online) Journal homepage: https://www.tandfonline.com/loi/hpli20

The Attitudinal Entropy (AE) Framework as a General Theory of Individual Attitudes

Jonas Dalege, Denny Borsboom, Frenk van Harreveld & Han L. J. van der Maas

To cite this article: Jonas Dalege, Denny Borsboom, Frenk van Harreveld & Han L. J. van der Maas (2018) The Attitudinal Entropy (AE) Framework as a General Theory of Individual Attitudes, Psychological Inquiry, 29:4, 175-193, DOI: 10.1080/1047840X.2018.1537246

To link to this article: https://doi.org/10.1080/1047840X.2018.1537246

© 2018 Taylor & Francis Group, LLC

Published online: 16 Jan 2019.

Submit your article to this journal

Article views: 356

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hpli20

PSYCHOLOGICAL INQUIRY
2018, VOL. 29, NO. 4, 175–193
https://doi.org/10.1080/1047840X.2018.1537246

Routledge
Taylor & Francis Group

TARGET ARTICLE

Check for updates

The Attitudinal Entropy (AE) Framework as a General Theory of Individual Attitudes

Jonas Dalege, Denny Borsboom, Frenk van Harreveld, and Han L. J. van der Maas

Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT

This article introduces the Attitudinal Entropy (AE) framework, which builds on the Causal Attitude Network model that conceptualizes attitudes as Ising networks. The AE framework rests on three propositions. First, attitude inconsistency and instability are two related indications of attitudinal entropy, a measure of randomness derived from thermodynamics. Second, energy of attitude configurations serves as a local processing strategy to reduce the global entropy of attitude networks. Third, directing attention to and thinking about attitude objects reduces attitudinal entropy. We first discuss several determinants of attitudinal entropy reduction and show that several findings in the attitude literature, such as the mere thought effect on attitude polarization and the effects of heuristic versus systematic processing of arguments, follow from the AE framework. Second, we discuss the AE framework's implications for ambivalence and cognitive dissonance.

KEYWORDS

Attention; attitude; entropy; network models; statistical mechanics; thought

A century of research on attitudes has produced an impressive amount of empirical findings and identified an abundance of concepts and processes related to attitudes. An important next step toward a thorough understanding of attitudes would be a theoretical framework able to explain these empirical findings from few first principles. The aim of this article is to develop such a framework. To do so, we make use of analogical modeling (Haig, 2005): We use statistical mechanics as a starting point for our framework because of its advanced theory and because our earlier analysis has shown that a formalized measurement model of attitudes can be based on statistical mechanics principles (Dalege et al., 2016) and show that an analogous theoretical approach to attitude can explain a wide variety of empirical phenomena.

Statistical mechanics revolves around three fundamental properties of a system—entropy (a measure of the system's randomness), energy, and temperature.[1] To investigate whether statistical mechanics represents a fruitful starting point for a general theory of attitudes, we search for analogies of these fundamental properties and test whether the consequences of these analogies match empirical findings in the attitude literature. Based on this approach we derive the Attitudinal Entropy (AE) framework, which rests on three propositions. First, inconsistency and instability of an attitude represents attitudinal entropy and is therefore the natural state of an attitude. Second, the energy of the attitudinal representation serves as a local processing possibility to evaluate the global entropy of an attitude. Third, attention and thought directed at the attitude object have an analogous effect on the attitudinal representation as (inverse) temperature has on thermodynamic behavior—heightened attention and thought make attitudinal representations low in energy more likely and therefore reduce the entropy of the attitude.

The structure of this article is as follows. First, we discuss the main tenets of the AE framework. Second, we discuss determinants of reduction of attitudinal entropy and show that several findings in the attitude literature, such as individual vs. group effects of implicitly measured attitudes, the mere thought effect, and systematic vs. heuristic processing, follow from these determinants. Third, we discuss ambivalence (e.g., Priester & Petty, 1996) and cognitive dissonance (Festinger, 1957) from the perspective of the AE framework. Throughout the subsequent sections we model several established phenomena in the attitude literature to show that the AE framework indeed holds promise in explaining several phenomena with few first principles. We also identify several predictions that can be derived from the AE framework in each discussion of a given phenomena to illustrate the predictive power of the AE framework and to define an empirical agenda for future research. We close by discussing potential neural substrates of the AE framework's propositions, the AE framework's relation to other broad models of attitude, and several open questions that need to be addressed to further develop the AE framework.

CONTACT Jonas Dalege j.dalege@uva.nl Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129, 1018 WT, Amsterdam, The Netherlands.

Color versions of one or more of the figures in this article can be found online at www.tandfonline.com/hpli

[1]Temperature, strictly speaking, might not be regarded as a fundamental property, because it can be derived from the relation between entropy and energy. However, for our current purposes it is beneficial to treat temperature as a fundamental property.

© 2018 Taylor & Francis Group, LLC

The AE Framework

In this section, we discuss the meaning of attitudinal entropy and its implications for the dynamics of attitudes. We first discuss micro- and macrostates of attitudes and then turn to the meaning of attitudinal entropy. Based on these definitions, we derive the AE framework.

Attitudinal Micro- and Macrostates

The first question that needs to be addressed before we can define attitudinal entropy is what constitutes microstates and macrostates of an attitude. In statistical mechanics, a microstate refers to the microscopic configuration of a given system (e.g., the position of each oxygen molecule in the room you are sitting in), and a macrostate refers to the macroscopic behavior of a given system (e.g., whether all the oxygen molecules are centered in one corner or whether they are evenly dispersed throughout the room). In line with several theories on attitudinal structure (e.g., Dalege et al., 2016; Eagly & Chaiken, 2007; Fishbein & Ajzen, 1975; Rosenberg, Hovland, McGuire, Abelson, & Brehm, 1960), we define the microstate of an attitude *as the configuration of the relevant beliefs, feelings, and behaviors toward an attitude object (i.e., attitude elements)*. As an example take the attitude toward snakes. The microstate of this attitude can be represented like this: Attitude Element 1 (e.g., snakes maintain ecological order) is positive, Attitude Element 2 (e.g., snakes are scary) is negative, Attitude Element 3 (e.g., I run away when I see a snake) is negative, and so forth. The macrostate of an attitude is then defined as *the combination of all attitude elements (e.g., how many attitude elements are negative and how many are positive)*. Based on several theories on the integration of attitude elements into a global evaluation (e.g., Anderson, 1971; Cacioppo, Petty, & Green, 1989; Cunningham & Zelazo, 2007; Fazio, 1995; Zanna & Rempel, 1988), we assume that the global evaluation of an attitude object is strongly related to the macrostate of an attitude, in that it represents a context-depended weighted sum score of the attitude elements. Thus, we propose the following three definitions:

Definition 1: The configuration of the attitude elements constitutes the microstate of the attitude.

Definition 2: The number of positive versus negative attitude elements constitutes the macrostate of an attitude.

Definition 3: A situation-depended weighted sum score constitutes the global evaluation of an attitude object.

Attitudinal Entropy

Entropy is a concept originating from thermodynamics, where it was originally defined as energy that is lost when energy is transformed (e.g., from chemical to kinetic energy). Take as an example the situation when you walk up a steep hill. To do this, your body has to transform chemical energy in the form of calories to kinetic energy so that your legs move up the hill. However, during this transformation of energy, some energy is inevitably lost that is not put to work, constituting heat loss or entropy. This notion also lies at the heart of the second law of thermodynamics, which states that entropy of an isolated system *always* increases.

Although the concept of entropy originated in classical thermodynamics, its application to statistical mechanics resulted in a broader use of entropy as a general measure of disorder or uncertainty in a system. The physicist Ludwig Boltzmann (1877) developed the statistical mechanics definition of entropy, which holds that a macrostate that can be realized by many microstates has higher entropy than a macrostate that can be realized by few microstates (see Figure 1a). As an example, take the distribution of oxygen molecules in the room you are sitting in right now. Luckily, the macrostate of the oxygen molecules being distributed evenly throughout the room can be realized by many more microstates (thus having higher entropy) than the macrostate of the oxygen molecules clustering at one position in the room.[2] As an intuitive example of why the likelihood of a macrostate depends on its Boltzmann entropy, imagine a simple slot machine with three fields that can show a lemon, a peach, or a banana. The macrostate "win" (i.e., all fields showing the same fruit) can then be realized by three microstates (e.g., three lemons). The macrostate "lose" (i.e., the fields show at least two different fruits), on the other hand, can be realized by 24 (3^3–3) microstates. Although we already see with this simple example that high-entropy states are more likely than low-entropy states, the effect becomes increasingly pronounced with the size of the system increasing (up to the point where the high-entropy state is essentially the only possible state as is the case for the distribution of oxygen molecules in a room).

Applying the Boltzmann entropy to the domain of attitudes implies that inconsistent attitudes have higher entropy than consistent attitudes. To illustrate this, consider an attitude consisting of 10 attitude elements. A perfectly univalent attitude can be realized only by two different microstates (i.e., all attitude elements being either positive or negative). So the attitude of a snake enthusiast (i.e., judging snakes as entirely positive) can be realized only by one microstate. A perfectly ambivalent or neutral attitude, in contrast, can be realized by 252 microstates. So judging snakes as positive on some aspects and negative on others can be realized by a large number of microstates. This leads to the following first proposition of the AE framework:

Proposition I.1: Inconsistency of an attitude *is* the Boltzmann entropy of the attitude.

It is important to note here that the Boltzmann entropy concerns the entropy of a single given macrostate (e.g., five attitude elements are in a positive state and five attitude elements are in a negative state). The entropy of a system, on the other hand, is described by the Gibbs entropy (Jaynes, 1965). Gibbs entropy depends on the likelihood of the different microstates of a system. As Figure 1b illustrates, Gibbs entropy is at maximum when all microstates are

[2]Otherwise you might get crushed by all oxygen molecules distributed at the position of the room you are in, or you might suffocate because all oxygen molecules are at a different position than you.

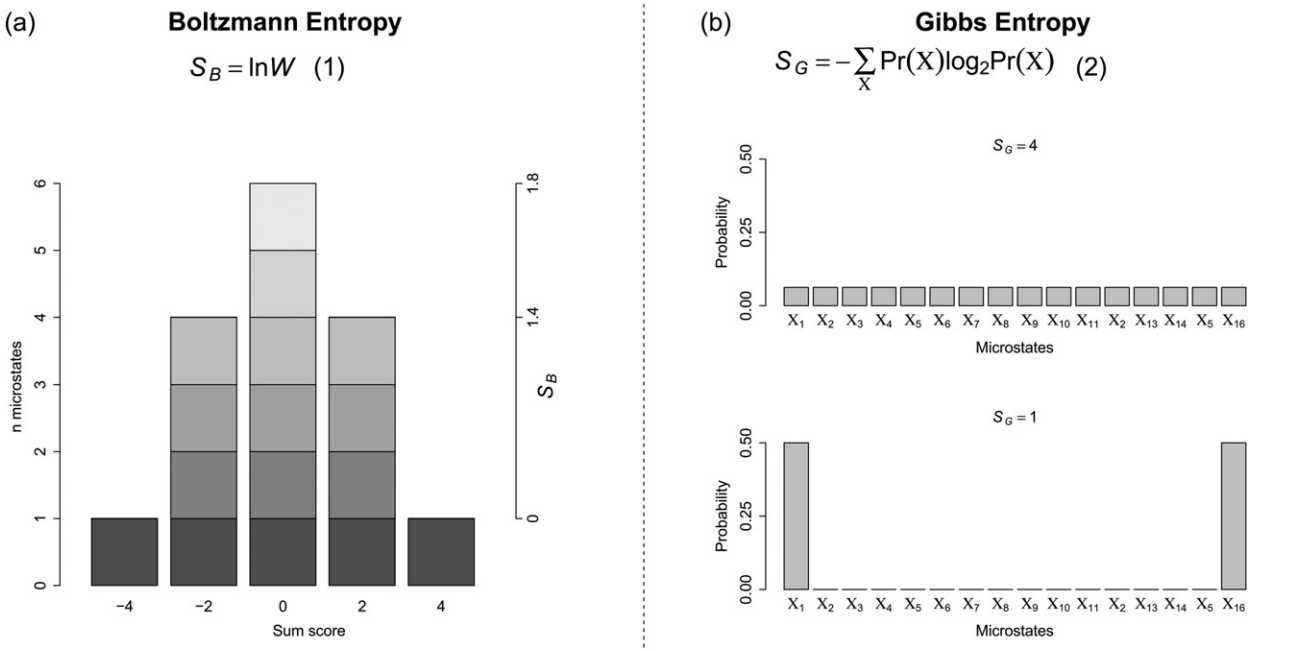


Figure 1. Illustrations of the Boltzmann and Gibbs entropies.
Note. In (1) *W* refers to the number of microstates that can realize the given macrostate. In (2) X refers to all possible states of a given system.

equally likely—implying that the system's behavior is completely random—and it is at minimum when only a single configuration is possible, implying that the system's behavior is completely ordered. As an example of Gibbs entropy, take the movement of water molecules. Under high temperature, water molecules move randomly (i.e., water is in a gas state); this indicates high Gibbs entropy, because the configuration (i.e., positions) of the water molecules is consistently changing (i.e., all microstates are roughly equally likely). In contrast, under low temperature the water molecules cannot move (i.e., water is in a solid state); this reflects low Gibbs entropy, because the configuration of the water molecules is stable (i.e., the current microstate is much more likely than all other microstates). Someone who consistently changes her attitude toward snakes would therefore have a high-entropy attitude toward snakes, whereas both a snake enthusiast and phobic have low-entropy attitudes toward snakes. The Gibbs entropy, therefore, measures the inherent stability of a system, which leads to the following proposition:

> **Proposition I.2:** The Gibbs entropy of the attitude network reflects the attitude's stability.

From Proposition I.1, it follows that the natural state of an attitude is neutral or ambivalent and that consistent attitudes should be rare. However, this is clearly not the case; even though individuals are often exposed to ambiguous information, they often arrive at consistent representations of the information (e.g., Holyoak & Simon, 1999; Simon & Spiller, 2016). So why are attitudes often consistent, whereas playing slot machines generally results in losing your money? The answer to this is that attitude elements are not independent of one another (to be explained next), and because of this dependency, attitudes can assume low-entropy macrostates. However, for a system to remain in a low-entropy state (i.e., low Gibbs entropy), force has to be put on this system, and it is our view that one of the main functions of focusing our attention on (or thinking about) an attitude object is to put such force on the attitude system and obtain (or maintain) a consistent attitude that is low in entropy.

Entropy reduction is a crucial aspect of life because a key characteristic of any living organism is that it must maintain order in their own system (Schrödinger, 1944). According to Kauffman (1993), the ability to reduce entropy is the most important selection criterion for evolution. This implies that the ability to reduce entropy is one of the central hallmarks of any living organism. We think that a similar argument can be made for the human mind, so that one of the central objectives of the human mind is to reduce its entropy (cf. Hirsh, Mar, & Peterson, 2012). It is straightforward that only attitudes low in entropy fulfill the functions typically associated with attitudes, such as to organize knowledge, increase utility, and express values (Katz, 1960; Smith, Bruner, & White, 1956). All these functions require attitudes to be in predictable, stable, and consistent states, and therefore attitudes are much more likely to fulfill their functions when they are low in entropy (e.g., only a low-entropy attitude toward snakes can clearly imply that you should run when you are near one). Linking the need for entropy reduction to cognitive consistency also echoes the fundamental and widespread assumption in research on attitudes that individuals have an inherent preference for cognitive consistency (e.g., Festinger, 1957; Gawronski & Strack, 2012; Heider, 1946, 1958; Monroe & Read, 2008; Shultz & Lepper, 1996).

The Causal Attitude Network Model

To formalize the ideas presented here, we build on the Causal Attitude Network (CAN) model (Dalege et al., 2016), which treats attitude elements as nodes in a network that are

connected by pairwise interactions. The complexity of the attitudinal representation is reflected by the size of the network (i.e., number of nodes). The CAN model is based on psychometric network models (e.g., Cramer, Waldorp, van der Maas, & Borsboom, 2010; van der Maas et al., 2006) and on constraint-satisfaction models of attitudes (e.g., Kunda & Thagard, 1996; Monroe & Read, 2008; Shultz & Lepper, 1996). The central assumption of the CAN model is that dynamics of attitude networks can be described in an idealized way by the Ising (1925) model, which originated from statistical mechanics. Although the Ising model is an extremely parsimonious model, its behavior is exceptionally rich. Due to these qualities, the Ising model has been applied to many different fields of research, such as magnetization (e.g., Ising, 1925), kinetic energy (e.g., Fredrickson & Andersen, 1984), predator–prey dynamics (e.g., Kim, Liu, Um, & Lee, 2005), neuroscience (e.g., Fraiman, Balenzuela, Foss, & Chialvo, 2009), clinical psychology (e.g., Cramer et al., 2016), and population dynamics (e.g., Galam, Gefen, & Shapir, 1982).

The Ising model describes the dynamics of networks by using the fact that systems strive toward low-energy configurations (see Figure 2 for an illustration of a simple Ising model). The energy of a configuration is determined by two classes of parameters. The first class constitutes the thresholds of the nodes, which determine the disposition of a given node to be "on" or "off" (denoted as τ_i). A node with a positive (negative) threshold requires less energy when it is "on" ("off"). In the original Ising model, thresholds represent the external field that influences the spins of the magnet. Similarly, in attitude networks, thresholds represent external information regarding the attitude object. These thresholds therefore represent the disposition of a given attitude element to be endorsed or not. A positive threshold represents a disposition of a given node to be "on" (e.g., a positive thresholds of judging snakes as dangerous indicates that one is inclined to judge snakes as dangerous holding all other information in the attitude network constant). A negative threshold represents a disposition of a given node to be "off" (e.g., a negative threshold of judging snakes as beautiful indicates that one is inclined to judge snakes as *not* beautiful). The magnitude of thresholds can also vary and the higher the magnitude, the stronger the disposition of the node to be "on" or "off". In the Ising model shown in Figure 2, two nodes have the disposition to be "on" (indicated by green thresholds, see the online article for the color version of the figure) and two nodes have the disposition to be "off" (indicated by red thresholds, see the online article for the color version of the figure).

The second class of parameters constitutes weights of edges between nodes, representing the strength of interaction between nodes (denoted as ω_i). Two nodes that have positive weights between them require less (more) energy when they assume the same (different) state, representing preference for consistency. A positive weight represents an exhibitory interaction (e.g., feeling afraid of snakes because you also judge them as dangerous), and a negative weight represents an inhibitory interaction (e.g., not judging snakes as beautiful because you judge them as dangerous). The magnitude of weights can also vary, and the higher the magnitude, the stronger the interaction. The CAN model assumes that weights between attitude elements generally arise based on inferences that support evaluative consistency. In the Ising model shown in Figure 2 all nodes are positively connected (indicated by green edges, see the online article for the color version of the figure). This Ising model thus represents a simple attitude network consisting of, for example, four positive beliefs (e.g., believing that snakes maintain ecological order and are safe, beautiful, and smooth). Note that in the current article we focus on the situation in which edges between attitude elements are already present. How we can model the development of edges in attitude networks is currently investigated in our laboratory. The starting point for this investigation is to combine the AE framework with connectionist models of attitudes, which assume that Hebbian learning underlies development of attitudinal structures (e.g., Monroe & Read, 2008).

Thresholds and weights determine a given configuration's energy (denoted by H). It is our view that, in contrast to the physical application of the Ising model, energy does not reflect an existing physical property. Calculation of energy is needed because it enables the mental system to arrive at a low-entropy state by evaluating locally which elements need to be changed. By evaluating several attitude elements in turn, the mental system is able to create a global low-entropy state without evaluating the global state directly (which would probably be too complex from a computational point of view). This leads to the following proposition:

> **Proposition II:** Energy of the attitudinal representation serves as a local processing possibility to evaluate the global Boltzmann entropy of an attitude. Attitude elements are likely to change when the opposite state has lower energy.

The extent to which a configuration's energy results in the configuration with lower energy being more likely than a configuration with higher energy depends on the dependence parameter β (representing temperature in the original Ising model). The higher the dependence parameter, the more the probability of a configuration depends on its energy. Because of this, the dependence parameter directly scales the Gibbs entropy of a given Ising model (e.g., Kindermann & Snell, 1980), implying that increasing the dependence parameter results in attitude networks being more ordered and stable. For example, the Ising model with dependence at 0 at the top of Figure 2 has also maximum Gibbs entropy, because all configurations are equally likely. In contrast, the Ising model with high dependence at the bottom of Figure 2 has lower Gibbs entropy, because the completely consistent configurations are much more likely than the inconsistent configurations. A system low in Gibbs entropy thus creates the possibility of macrostates having low Boltzmann entropy, but as long as the system is not at minimum Gibbs entropy, macrostates with high Boltzmann entropy are still possible.

The probability formula allows us to calculate the distribution of configurations we would expect if we measure an infinite number of individuals holding an attitude that can be described by a given Ising model. For describing the

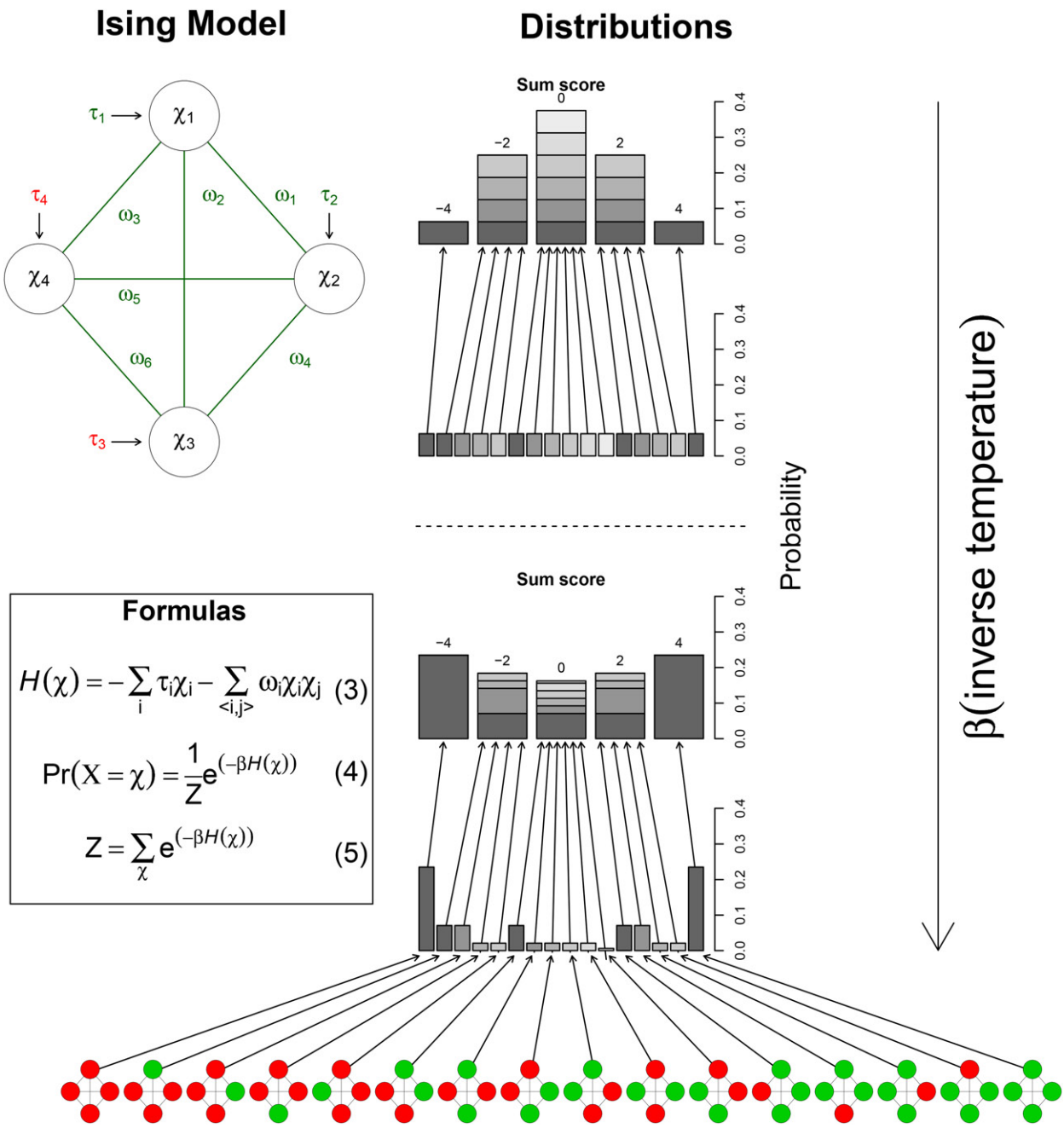


Figure 2. Illustration of the Ising model.
Note. In (3) $H(x)$ represents the Hamiltonian energy of the configuration of k distinct nodes $1, \ldots, i, j, \ldots k$, that engage in pairwise interactions and the variables x_i and x_j represent the states $(-1, +1)$ of nodes i and j, respectively. The parameter τ_i represents the threshold of node i and the parameter ω_i represents the interaction weight between nodes i and j. In (4) $\Pr(X = x)$ represents the probability of a given network configuration and β represents the dependence parameter of the Ising model. In (5) Z represents the standardization factor, which ensures that the probabilities add up to 1. The Distributions part of the figure shows the probability distributions of two Ising models for the sum scores of the nodes (upper distributions) and the individual configurations (lower distributions). The bottom of the figure shows all possible states of the four-node network, with green (red) nodes indicating that the node is "on" ("off").

dynamics of a given individual's attitude, we can use time-dependent dynamics called Glauber dynamics (Glauber, 1963). The basic workings of Glauber dynamics on Ising models are that at each iteration we (a) calculate the energy of the current configuration, (b) pick a random node and calculate the energy of this neighboring configuration when this node is "flipped" (e.g., when this node changes from on to off), (c) determine the probability of the node actually flipping by using the difference in energy, and (d) flip the node with this probability (see Figure 3 for an illustration and formula). For attitude dynamics, this implies that increasing an attitude's consistency can be described by such dynamics. For example, if one believes that snakes are safe while one also feels scared of them and always screams when one sees a snake, the probability that one changes his or her belief that snakes are safe is high. In the simulations we describe later, we make use of Glauber dynamics when we model individual-level dynamics.

Figure 3 illustrates the reason why the dependence parameter scales the Gibbs entropy of an Ising model. In the network with the dependence parameter at 0.5, the thresholds and weights have little influence on the network's

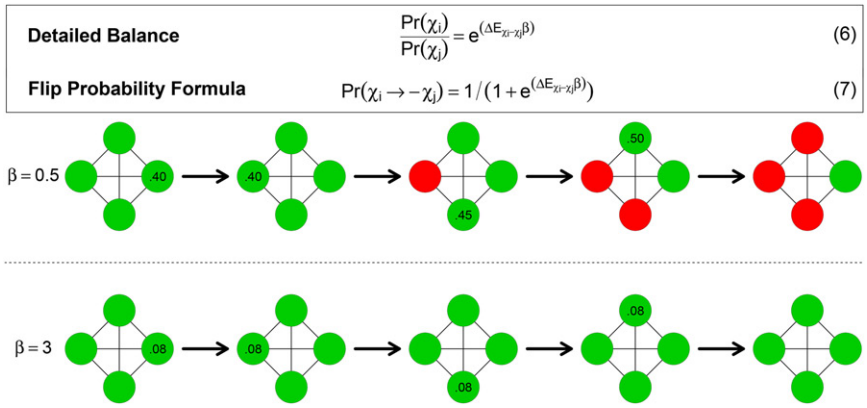


Figure 3. Glauber dynamics for two four-node networks under different dependence parameters.
Note. In (6) and (7) x_i and x_j represent the current state of a given node and its opposite state, respectively. In (6) the probability of a node remaining in its current state relative to the probability that it will flip is represented. Each network represents one iteration and the node with a given probability represents a node that was randomly picked to be flipped with the given probability.

dynamics and the network behaves essentially randomly. This situation therefore represents an attitude that is unstable and in which the different attitude elements are held with low certainty (e.g., the attitude of a person who does not care at all about snakes). With increases in the dependence parameter, the probability of the network configuration becomes increasingly dependent on the thresholds and the weights of the network. In the network with the dependence parameter at 3, the thresholds and weights have strong influence on the network's dynamics and the network behaves in accordance with these parameters. This situation therefore represents an attitude that is stable and in which the different attitude elements are held with high certainty (e.g., the attitude of a snake phobic or enthusiast). This underscores the necessity of attitude elements being dependent on one another to reduce attitudinal entropy. In this article, we argue that the dependence parameter in the Ising model constitutes a formalized representation of the effect of directing attention and thinking about attitude objects, leading to the third proposition of the AE framework:

> **Proposition III:** Focusing attention on the attitude object and thinking about the attitude object reduces the Gibbs entropy of attitudes by increasing the attitude network's dependence parameter. The higher the dependence parameter, the stronger the correspondence between energy and probability of a given attitude network configuration.

Based on our argument that individuals are motivated to reduce attitudinal entropy, we expect that a high level of attitudinal entropy causes psychological discomfort. However, because we assume that entropy of attitude networks cannot be directly evaluated, this influence is indirect. We expect that both measures of attitudinal entropy translate into psychological discomfort through proxies, which are easier to evaluate for the mental system. This leads to the following implications:

> **Implication I:** High Boltzmann entropy in combination with a high dependence parameter indirectly leads to psychological discomfort. The Boltzmann entropy is indirectly evaluated by the difference in energy of the current and neighboring configurations (i.e., configurations for which only one attitude element has to be flipped).

> **Implication II:** High Gibbs entropy in combination with a high dependence parameter indirectly leads to psychological discomfort. The Gibbs entropy is indirectly evaluated by the temporal stability of the attitude.

Levels of Attitudinal Entropy Reduction

In this section, we discuss different levels of attitudinal entropy reduction and research supporting these levels. Note that these levels do not represent distinctive categories but are assumed to lie on a dimension from weak entropy reduction to high entropy reduction (just as the dependence parameter in the Ising model is also a continuous variable). It is our view that thinking about an attitude object—or, more generally, paying attention to an attitude object—has the default effect of slightly increasing the dependency of the attitude network; as such, simply focusing attention on the attitude object represents the most basic level of dependency of the attitude network. Such a situation, for example, arises when an individual observes an attitude object. The dependence parameter increases when the individual is prompted to think about the attitude object, which would, for example, be the case when the individual responds to a questionnaire about an attitude object.

Increased levels of attitudinal entropy reduction may arise when individuals are for some reason prompted to think more elaborately about an attitude object and dependency of the attitude network is further increased when motivational factors come into play, representing intermediate levels of attitudinal entropy reduction. Examples of factors moderately increasing motivation to reduce attitudinal entropy are situations in which individuals are committed to an evaluation or in which they have to make a relatively unimportant decision.

Even more enhanced levels of attitudinal entropy reduction arise when individuals attach personal importance to their attitudes. Attitude importance is a widely researched topic and is a key determinant of attitude strength (Howe & Krosnick, 2017). Factors increasing attitude importance are the attitude's relevance to self-interests (e.g., attitude's relevance to important decisions), to personal values, and to social identification (Boninger, Krosnick, & Berent, 1995).

Crucially, attitude importance is strongly related to how much attention individuals devote to an attitude object (Krosnick, Boninger, Chuang, Berent, & Carnot, 1993), making it likely that, indeed, attitudes high in personal importance represent attitude networks with the highest dependence and therefore also the lowest entropy attitudes. Furthermore, as we discuss in a later section, strong attitudes show exactly the dynamics that would be expected from high dependence attitude networks.

To summarize, lower levels of attitudinal entropy reduction arise when individuals pay some attention to the attitude object or briefly think about it. Intermediate levels represent situations in which an individual is for some reason prompted to think about the attitude object in more detail (e.g., when an argument regarding the attitude object has to be evaluated) and when individuals have to base a decision on their attitude network or are committed to an evaluation. The final levels of attitudinal entropy reduction arise when an individual attaches high personal importance to an attitude object. In the remainder of this section, we show that several central findings in the attitude literature follow from the entropy reducing function of attention and thought.

Implicit Measures Are More Likely to Tap Attitudes in High-Entropy States

The dependence parameter of attitude networks increases when attention is directed at the attitude object, which implies that the measurement of attitudes influences the dependence parameter. Implicit measures of attitudes, such as the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) and the Affective Misattribution Task (Payne, Cheng, Govorun, & Stewart, 2005), limit attention directed at the attitude object by measuring attitudes without directly asking individuals to introspect. These measures are therefore more likely to tap attitudes in high-entropy states than explicit measures. Attitudes in high-entropy states are less internally consistent than attitudes in low-entropy states, which might contribute to the fact that implicit measures generally show both poor internal reliability and test–retest reliability (e.g., Bar-Anan & Nosek, 2014; Gawronski, Morrison, Phills, & Galdi, 2017; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). In contrast to the low individual temporal stability of scores on implicit measures of attitudes, mean effects on implicit measures are substantially more robust (Payne, Vuletich, & Lundberg, 2017). For example, children show similar scores on the IAT as adults (Baron & Banaji, 2006), and the IAT predicts behavior much better on a global level (e.g., police shootings of Blacks is strongly associated with prejudice assessed with the IAT on a regional level; Hehman, Flake, and Calanchini, 2018) than on an individual level, which is generally rather low (e.g., Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). A recent review identified these patterns as important puzzles in the literature on implicit measures of attitudes (Payne, Vuletich, & Lundberg, 2017). With the following simulations we show that these puzzles can be straightforwardly solved by assuming that implicit measures are more likely to tap attitudes in high-entropy (instable) states.

Simulation 1a: Implicit measures – low temporal stability, stable means. For this simulation, we investigated Glauber dynamics on a fully connected 10-node network with all edges set to .1. We varied the thresholds uniformly from −.2 to .7 with steps of .1, resulting in 10 different sets of thresholds. (We chose these thresholds so that the means differ from 0 and that there is sufficient variance for correlations to be meaningful.) The dependence parameter was set to .5 (representing that only some attention is directed at the attitude object) and the network was randomly initialized. We simulated 100 individuals for each set of thresholds, resulting in the total number of 1,000 individuals. For each individual we simulated 1,000 iterations. After 500 and 1,000 iterations, respectively, we measured the sum score of the nodes. The resulting scores at the first measurement and at the second measurement were only weakly correlated ($r = .24$, $p < .001$). The means of the first measurement ($M = 1.91$) and the second measurement ($M = 2.04$), in contrast, were virtually identical, $t(999) = 0.71$, $p = .479$, although the standard deviation at both the first measurement ($\sigma = 4.55$) and the second measurement ($\sigma = 4.45$) were substantial. Networks under a low dependence parameter thus show low individual temporal stability but stable means, which precisely matches the known behavior of implicit attitude measures.

Simulation 1b: Implicit measures – low behavior prediction on individual level, high behavior prediction on group level. For this simulation, we used mostly the same setup as in Simulation 1a, with the following adjustments. First, we added an 11th node that represented behavior. This node was also connected with weights of .1 to all other nodes but always had a threshold of 0 (so that all systematic variation in this node is caused by its connection to other nodes). Second, we ran a total number of 10,000 individuals to be able to create groups of sufficiently large size. Third, we created 10 groups that differed in their mean thresholds (see Table A1 in the appendix). The correlation between the first 10 nodes and the "behavior" node on an individual level was relatively weak ($r = .23$, $p < .001$). In contrast, the correlation on the group level was very strong ($r = .80$, $p = .006$). These patterns fit the finding that individual-level correlations between implicit measures of attitudes and behavior are relatively low and that group level correlations are considerably stronger.

The implication of the AE framework that implicit measures are more likely to tap attitudes in high-entropy states than explicit measures has fundamental implications for the research on implicit measures of attitudes. Although researchers in this domain have long acknowledged that implicit measures show low internal consistency, they have generally interpreted this as a measurement problem (e.g., Fazio & Olson, 2003b; Gawronski, LeBel, & Peters, 2007; LeBel & Paunonen, 2011; Nosek & Banaji, 2001). However, the AE

framework implies that the construct measured by implicit measures is *itself* more internally inconsistent than the construct measured by explicit measures because the former by their very nature direct less attention toward the attitude object than the latter. One consequence of this is that often the only way to make implicit measures more reliable is to make them more explicit, leading to the counterintuitive conclusion that a valid measurement of attitudes (or any system for that matter) in high-entropy states *must* be unreliable.

Prediction 1a. Manipulating the dependency in attitude networks (e.g., letting individuals think for some time about the attitude object) is expected to increase internal consistency and stability of implicit measures.

Prediction 1b. Scores on implicit measures assessing attitudes individuals regularly think about are expected to have higher internal consistency and stability than scores on implicit measures assessing attitudes individuals think only infrequently about.

Prediction 1c: Implicit and explicit measures should show the lowest convergence when the dependence of the attitude network is generally low.

The Mere Thought Effect as an Initial Level of Heightened Attitudinal Entropy Reduction

The mere thought effect on attitude polarization refers to the classic finding that briefly thinking about an attitude object without receiving external information results in more extreme evaluation of the attitude object (e.g., Tesser, 1978; Tesser & Conlee, 1975). Based on several studies on the mere thought effect, Tesser, Martin, and Mendola (1995) argued that (a) sufficiently complex cognitive schemas (defined as the number of dimensions an attitude object is rated on) are necessary (Tesser & Leone, 1977) and (b) the evaluative dimensions, on which the cognitive schema is based, need to be sufficiently interdependent for the mere thought effect to manifest itself (Millar & Tesser, 1986). In the following simulations, we show that the mere thought effect and its moderators naturally follow from the AE framework.

Simulation 2a: Basic mere thought effect. We calculated the probabilities of the sum scores of a fully connected 10-node network with the dependence parameter set to either 1 (representing merely asking individuals about their attitudes) or 1.5 (representing mere thought). All edge weights were set to .1 and all thresholds were set to 0. As can be seen in Figure 4a, increasing the dependence parameter leads to an increase in extreme sum scores, mimicking the basic mere thought effect.

Simulation 2b: Network size as formalization of a complex cognitive schema. As stated in the section on the CAN model, the size of networks reflects the complexity of the cognitive schema of the attitude object. We therefore expect that a network with few nodes will not show a strong increase in extreme sum scores when the dependence parameter is increased. To investigate this, we adapted Simulation 2a by decreasing the number of nodes to 4. As can be seen in Figure 4b, increasing the dependence parameter for such a small-size network does not lead to a substantial increase in extreme sum scores, mimicking the finding that complex cognitive schemas are necessary for the mere thought effect to manifest itself.

Simulation 2c: Magnitude of edge weights as a formalization of dependence between evaluative dimensions. Edge weights in Ising networks are a straightforward formalization of interdependence between evaluative dimensions. To investigate whether decreasing edge weights leads to lower increase in extreme scores, we adapted Simulation 2a by setting all edge weights to .05. As can be seen in Figure 4c, increasing the dependence parameter for such a weakly connected network does not lead to a substantial increase in extreme sum scores, mimicking the finding that interdependence of complex schemas is necessary for the mere thought effect to manifest itself.

Further support for the proposition that merely thinking about an attitude object represents an intermediate level of attitudinal entropy reduction comes from research on the coherence effect in judgment and decision making (e.g., Holyoak & Simon, 1999; Simon, Krawczyk, & Holyoak, 2004; Simon, Pham, Le, & Holyoak, 2001; Simon, Snow, & Read, 2004; Simon, Stenstrom, & Read, 2015). The coherence effect represents the general finding that when individuals are presented with ambiguous information about a given scenario (e.g., a legal case), individuals interpret this information in such a way that it allows for a coherent judgment or decision about the scenario. Of interest, such coherence shifts are also observed for dependency between emotions and beliefs regarding an attitude object (Simon et al., 2015) and in the complete absence of making a decision (Simon et al., 2001). However, there are some indications that having to make a decision heightens the coherence shift effect (Simon et al., 2001, 2004).

Prediction 2: Sizes of edge weights and size of attitude network predict the strength of the mere thought effect.

Prediction 3: Because the AE framework assumes that increasing dependency of attitude networks is a continuous process, the AE framework predicts that an opposite mere thought effect also exists, in the sense that when individuals are asked to very quickly answer attitude questions, attitudes are expected to be *less* polarized than when individuals are given more time to answer the questions. Note that the AE framework predicts that this would constitute a small effect.

Attitude Strength

The highest levels of attitudinal entropy reduction have implications for attitude strength. The macrobehavior of Ising networks is governed by the dependence of the network and can be described by the cusp catastrophe model (Sitnov, Sharma, Papadopoulos, & Vassiliadis, 2001). The cusp catastrophe model describes sudden versus smooth changes in a variable depending on two control variables, referred to as the normal variable and splitting variable, respectively (Gilmore, 1981; Thom, 1972; Zeeman, 1976). Depending on the value of the splitting factor, the influence

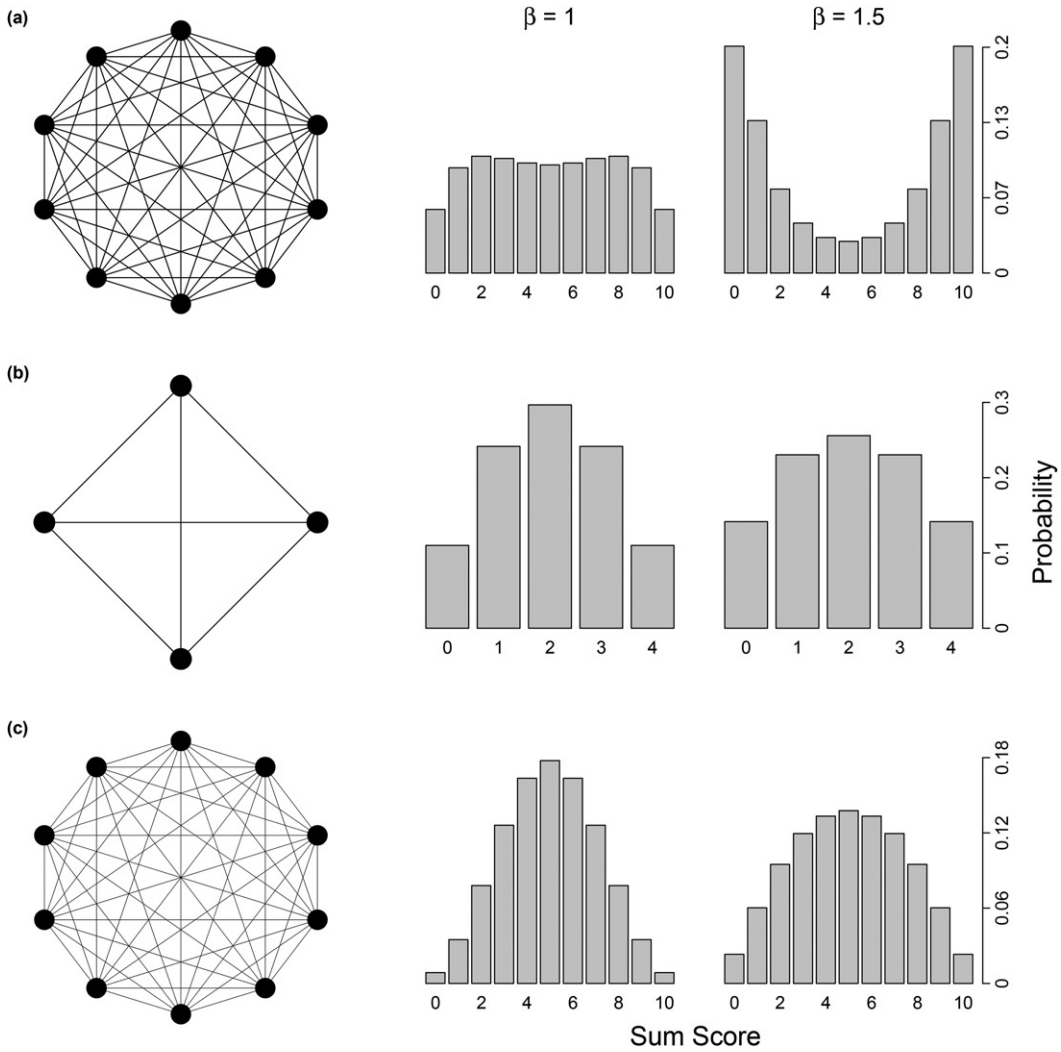


Figure 4. The mere thought effect on polarization modeled by the Attitudinal Entropy framework.
Note. Implied distributions of sum scores are shown under different independence parameters for (a) a 10-node network with all edge weights equal to .1, (b) a four-node network with all edge weights equal to .1, and (c) a 10-node network with edge weights equal to .05.

of the normal variable on the dependent variable is either gradual or discrete, implying that the so-called bifurcation area, in which sudden transitions happen in the dependent variable, is larger when the splitting factor is high (see Figure 5). As an illustration, take the freezing of water. In this case, temperature represents the normal variable and pressure represents the splitting variable: Under low pressure, water freezes and melts at the same temperature, whereas under high pressure, frozen water melts at a higher temperature than when liquid water freezes (and the other way around).

In Ising networks the average of the thresholds functions as normal control variable, the dependence of the network as splitting variable, and the macrobehavior of the network as dependent variable (Sitnov et al., 2001). Because of this, networks high in dependency are stable and ordered and change happens suddenly, whereas networks low in dependency are fluctuating and random and change happens gradually. These observations link the AE framework to the catastrophe model of attitudes, which assumes that attitude change can be described by the cusp catastrophe model (Flay, 1978; Latané & Nowak, 1994; Zeeman, 1976). In the catastrophe model of attitudes, valenced information functions as the normal variable, attitude involvement or attitude importance functions as the splitting variable, and the global evaluation functions as the

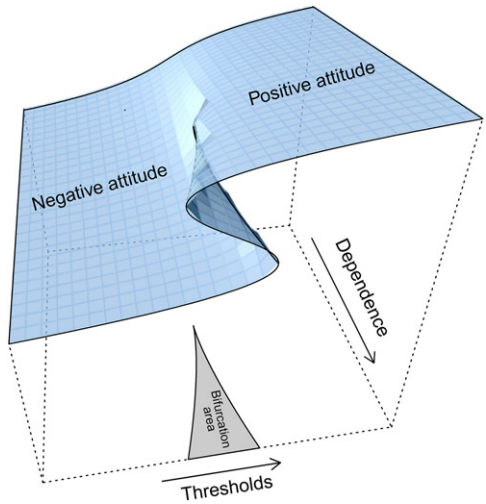


Figure 5. The cusp catastrophe model from the perspective of the Attitudinal Entropy framework.

dependent variable (see Figure 5). Several studies support the catastrophe model of attitudes by showing that important attitudes are more extreme than unimportant attitudes (e.g., Latané & Nowak, 1994; Liu & Latané, 1998) and by directly fitting the catastrophe model to data on attitudes (van der Maas, Kolstein, & van der Pligt, 2003). The CAN model can easily integrate the catastrophe model of attitudes and also provides a micro-level explanation of the postulates of the catastrophe model (Dalege et al., 2016). Thresholds in the CAN model directly relate to the valenced information a person receives regarding an attitude object and the macro-behavior of an attitude is strongly related to global evaluations of the attitude object. These similarities lead to the conclusion that important attitudes are based on attitude networks high in dependence.

Linking attitude importance to the dependence of attitude networks also has broader implications for attitude strength. As attitude importance is a central determinant of attitude strength (Howe & Krosnick, 2017), it becomes likely that strong attitudes represent high-dependence attitude networks. Indeed, the dynamics of strong attitudes are highly similar to the dynamics of strongly connected networks (Dalege et al., 2016), which in turn are similar to the behavior of low-dependence networks. Similar to strong attitudes (Krosnick & Petty, 1995), high-dependence networks are more stable and resistant (Kindermann & Snell, 1980). Increasing the dependence of attitude networks likely results in information being processed in accordance with the attitude, which represents another central feature of attitude strength.

Biased information processing is related to the phenomenon of hysteresis in the cusp catastrophe model. Hysteresis implies that the point at which a system moves to the opposite state depends on the direction of change (just as is the case for the melting and freezing of water under high pressure). The strength of the hysteresis effect in the cusp catastrophe model depends on the splitting variable—implying that attitude networks under high dependence should show strong hysteresis effects (i.e., the bifurcation area becomes broader). Changing such an attitude thus requires a disproportionate amount of persuasion compared to the amount of information the individual already received. In such a situation it would probably be more effective to first reduce the dependence parameter of the attitude network so that the individual is more "open" to change.

Attitude-behavior consistency, which represents the final central feature of attitude strength, is also more likely in high dependence attitude networks, because attitude elements are more dependent on one another. As the CAN model treats behavior as part of the attitude network, increasing dependence of attitude networks also increases the dependence of behavior on beliefs and feelings regarding the attitude object, and vice versa, implying higher attitude-behavior consistency (Dalege, Borsboom, van Harreveld, Waldorp, & van der Maas, 2017). This hypothesis is supported by findings indicating that attitude-behavior consistency depends on the stability of attitudes (Glasman & Albarracín, 2006), which represents a proxy of the dependence of the attitude network.

Prediction 4: The mere thought effect extends to the stability and resistance of attitudes. Thinking briefly about attitude objects is predicted to (temporally) increase the stability and resistance of attitudes.

Prediction 5: Persuasion is more effective for strong attitudes when the dependence parameter is lowered (e.g., by reducing attention directed at the attitude object) before the persuasion is employed, because lowering the dependence parameter reduces the hysteresis effect.

Prediction 6: Whether an attitude changes continuously or discretely depends on the dependence parameter of the attitude network.

Prediction 7: Reducing the dependence parameter of networks (e.g., by limiting cognitive capacity) results in less stable and less resistant attitudes.

Heuristic-Based versus Argument-Based Persuasion as Global versus Specific Threshold Changes

The elaboration likelihood model (Petty & Cacioppo, 1986) and the Heuristic Systematic Model (Chaiken, Liberman, & Eagly, 1989) are two hallmark dual process theories assuming that persuasion can be accomplished via two routes—one in which individuals change their attitudes based on heuristic cues (e.g., whether the source of the message is an expert) and one in which individuals change their attitudes based on a deeper processing of the quality of the persuasive arguments. Several studies have supported this idea and showed that individuals low in involvement are more likely to change their attitudes according to heuristic cues, whereas individuals high in involvement are more likely to change their attitudes according to argument quality (e.g., Petty & Cacioppo, 1984; Petty, Cacioppo, & Goldman, 1981; Petty, Cacioppo, & Schumann, 1983). From the perspective of the AE framework, heuristic-based persuasion represents a moderate global change in the attitude network's thresholds (i.e., change in the magnetic field in the language of the original Ising model), whereas argument-based persuasion represents a strong change of few specific thresholds, implying that moderate global change is more influential under low dependence and strong specific change is more influential under high dependence. We tested this hypothesis in the following simulation.

Simulation 3: Global versus specific threshold changes.

For this simulation, we again used a fully connected 10-node network with all edge weights set to .1. We investigated Glauber dynamics of this network using 1,000 iterations. In the first 500 iterations, all simulated individuals' thresholds were set to .2 (thus representing a positive initial attitude) and the network was randomly initialized. In the second 500 iterations, (a) thresholds remained at .2 (representing the no heuristic cue/weak arguments condition), (b) all thresholds changed to −.12 (representing the heuristic cue/weak arguments condition), (c) the first four thresholds changed to −.6 and the other thresholds remained at .2 (representing the no heuristic cue/strong arguments condition), or (d) the first four thresholds changed to −.72 and

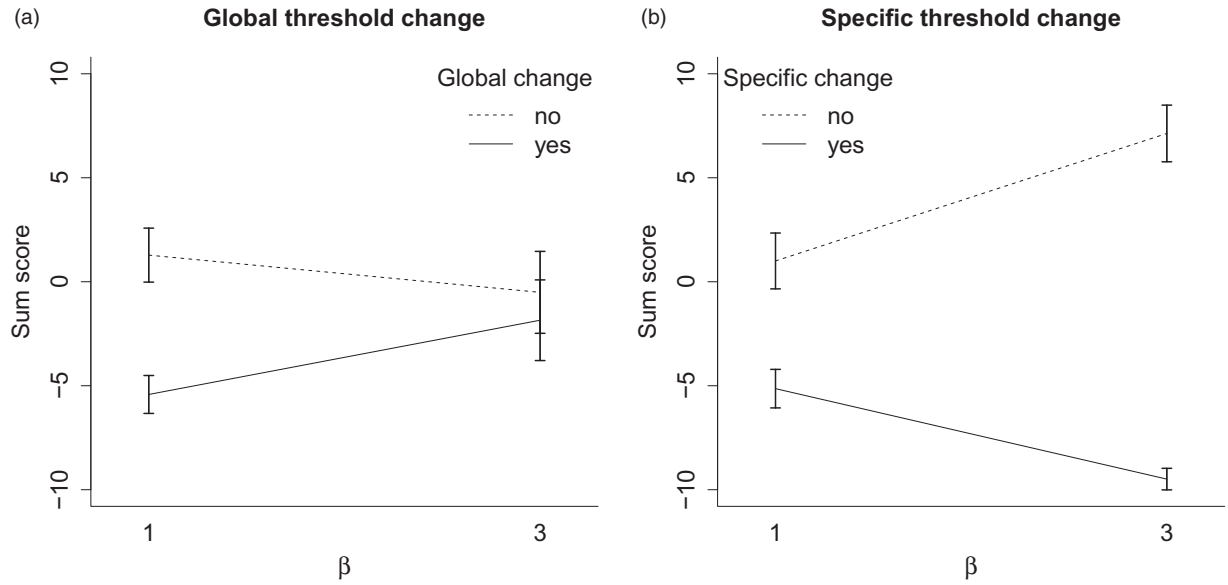


Figure 6. Effects of global (a) and specific threshold change (b) under moderate and high β.
Note. Error bars represent ±2 *SD*s around the mean.

the other thresholds changed to −.12 (representing the heuristic cue/strong arguments condition). The thresholds were chosen this way so that the mean change in the heuristic cue/weak arguments condition and the no heuristic cue/strong arguments condition was equal. Half of the simulated individuals' βs was set to 1 (representing low involvement) and the other half of the simulated individuals' βs was set to 3 (representing high involvement). We simulated 100 individuals for each experimental cell, resulting in 600 simulated individuals in total. The results showed a pattern reminiscent of the typical findings in the heuristic-based versus argument-based persuasion literature (e.g., Petty et al., 1981). The three-way interaction on the sum score of the attitude elements at the 1,000th iteration was significant, $F(1, 792) = 16.40$, $p = .001$, $\eta_p^2 = .02$, and the pattern of the results was in line with the hypotheses (see Figure 6). Conceptualizing heuristic cues as global moderate threshold change and strong arguments as specific strong thresholds change thus explains the basic result in the heuristic versus argument-based persuasion literature.[3]

Prediction 8a: Sufficiently strong heuristic cues lead to attitude change under both low and high involvement.

Prediction 8b: A large number of strong arguments lead to attitude change under both low and high involvement.

Prediction 9: As can be seen in Figure 6b, specific threshold change also affected networks with low β to a meaningful extent. This leads to the prediction that, given sufficient power (e.g., in a meta-analysis), an effect of strong arguments should also be detected under low involvement.

[3]We want to emphasize that although in all the other simulations presented here the findings are highly robust to changes in parameters, for the current simulation specific parameters had to be chosen to find the reported pattern of the results (e.g., when global threshold changes are chosen that are too high or when too many specific thresholds are targeted, differences between the β conditions become less meaningful). The AE framework therefore predicts that the effect of argument versus persuasion-based persuasion is limited to a specific range of stimuli (i.e., not too strong heuristic cues or not too many strong arguments).

Aversiveness of Attitudinal Entropy: Ambivalence and Cognitive Dissonance from the Perspective of the AE Framework

The first proposition of the AE framework holds that inconsistency of an attitude is attitudinal entropy. Research on ambivalence and cognitive dissonance underscores that consistency is a fundamental human need and that violations of this need cause psychological discomfort. Cognitive dissonance refers to aversive feelings caused by incongruent beliefs and behaviors vis-à-vis an attitude object, with most research on cognitive dissonance focusing on the effects of carrying out a behavior inconsistent with the beliefs an individual holds. A crucial distinction in the research on ambivalence is that between potential (or objective) and felt (or subjective) ambivalence (e.g., Newby-Clark, McGregor, & Zanna, 2002; Priester & Petty, 1996; van Harreveld, van der Pligt, & de Liver, 2009).[4] Potential ambivalence refers to the number of incongruent attitude elements, and felt ambivalence refers to the aversive feelings caused by these incongruent attitude elements. Crucially, the distinction between potential and felt ambivalence is made, because potential ambivalence *can*, but *not necessarily does*, result in felt ambivalence. In other words: Ambivalence can, but does not have to be, unpleasant. From the perspective of the AE framework, the question when potential ambivalence results in felt ambivalence becomes the question when attitudinal entropy results in psychological discomfort. In this section we discuss two possibilities of how the mental system indirectly evaluates attitudinal entropy and under which circumstances this results in psychological discomfort. This discussion is based on Implications I and II of the AE framework. Implication I holds that Boltzmann entropy is indirectly evaluated through the energies of neighboring

[4]We use the terms *potential* and *felt ambivalence* throughout the article because these terms fit our framework better than the recently more commonly used terms *objective* (or *structural*) and *subjective ambivalence*.

configurations. In the following subsection we show that this implication integrates the gradual threshold (GT) model of ambivalence (Priester & Petty, 1996) into the AE framework. Second, Implication II holds that Gibbs entropy is indirectly evaluated by the temporal stability of the attitude network's configuration.

Boltzmann Entropy as Ambivalence

An influential account of how potential ambivalence translates into felt ambivalence is the GT model (Priester & Petty, 1996). This model assumes a curvilinear relation between the number of conflicting evaluations (treated here as attitude elements) and felt ambivalence, in which felt ambivalence increases less as the number of conflicting attitude elements increases (e.g., the difference between holding no conflicting attitude element and holding one conflicting attitude element is larger than the difference between holding three conflicting attitude elements and holding four conflicting attitude elements). The specific formula of the GT model is the following:

$$Ambivalence = 5(C+1)^{p} - (D+1)^{1/C}, \quad (8)$$

where C refers to conflicting attitude elements (i.e., attitude elements that are incongruent to the majority of attitude elements) and D refers to the number of dominant attitude elements (i.e., attitude element that is consistent with the majority of attitude elements). The p determines the power function and was estimated by Priester and Petty (1996) to lie somewhere between .4 and .5. Although Priester and Petty explicitly stated that these specific values are exogenous to the GT model, most research based on the GT model uses a value between .4 and .5 to calculate expected felt ambivalence scores (e.g., Clark, Wegener, & Fabrigar, 2008; Refling, Calnan, Fabrigar, MacDonald, Johnson, & Smith, 2013). Clearly, a better understanding of the power function would provide us with more knowledge on when and why potential ambivalence results in felt ambivalence (high p values would indicate an almost linear relation, whereas low p values would indicate a steep relation). From the perspective of the AE framework, the number of conflicting attitude elements is given by the configuration of the attitude network. We therefore expect that felt ambivalence as modeled by the GT model indirectly reflects Boltzmann entropy of the configuration of the attitude network. As stated in Implication I of the AE framework, Boltzmann entropy is indirectly evaluated by the energy difference between the current configuration and its neighboring configurations and that the psychological discomfort caused by the energy difference is amplified by the dependence parameter. Based on this reasoning, we expect the dependence parameter to determine the steepness of the relation between potential and felt ambivalence.

Simulation 4: Felt ambivalence as Boltzmann entropy. For this simulation we again used a fully connected 10-node network with all edge weights set to .1 and all thresholds set to 0. We first calculated the GT model's implied ambivalence scores for each of the possible configurations and varied the power function parameter between .3, .5, and .7.

Using Equation 6, we then calculated the mean preference of each node to remain in its current state for each configuration. We then averaged these scores for each configuration with the same number of conflicting attitude elements (e.g., for each configurations in which all but one attitude elements are in the positive state). We calculated the distribution of preference scores with the dependence parameter set to 1, 1.5, and 2.5. As can be seen in Figure 7, varying the dependence parameter has an analogous effect as varying the power function of the GT model. Based on this finding, the AE framework implies that the dependence parameter determines to what degree potential ambivalence translates into felt ambivalence.[5]

Based on the finding that preferences of nodes to remain in their current states are faster decelerating under a high dependence parameter, we expect that factors increasing the dependence parameter also increase felt ambivalence. This hypothesis is indirectly supported by the finding that having to make a decision increases felt ambivalence (e.g., Armitage & Arden, 2007; van Harreveld, Rutjens, Rotteveel, Nordgren, & van der Pligt, 2009; van Harreveld, van der Pligt, et al., 2009). The AE framework holds that basing a decision on an attitude increases the dependence parameter of the attitude network, which results in lower (relative) preference of nodes to remain in their current state if the current configuration is ambivalent.

> **Prediction 10:** Dependence of attitude networks moderates the relation between potential and felt ambivalence. The higher the dependence, the stronger the impact of the first incongruent attitude elements.

> **Prediction 11:** The AE framework assumes that dependence is increased when a decision has to be made. After the decision is made, dependence drops again. This implies that before a decision is contemplated *and* after a decision, correspondence between potential and felt ambivalence is expected to be lower than while contemplating the decision. The same holds for the stability and resistance of the attitude.

Gibbs Entropy as Ambivalence

In our view, it is likely that felt ambivalence, if caused by low preference of nodes to remain in their current state, generally represents a situation-dependent process (e.g., having to make a decision based on an ambivalent attitude). In contrast, felt ambivalence caused by Gibbs entropy reflects a more chronic state of felt ambivalence. Implication II of the AE framework holds that the Gibbs entropy of an attitude network is indirectly evaluated by the stability of the attitude. Based on this implication, we expect that unstable attitude networks in combination with a high dependence parameter cause strong feelings of ambivalence. To investigate under which circumstances low stability and a high

[5]Note that with an increasing dependence parameter, the preference scores of almost all configurations increase. It therefore seems likely that the preference scores are evaluated relative to the dependence parameter.

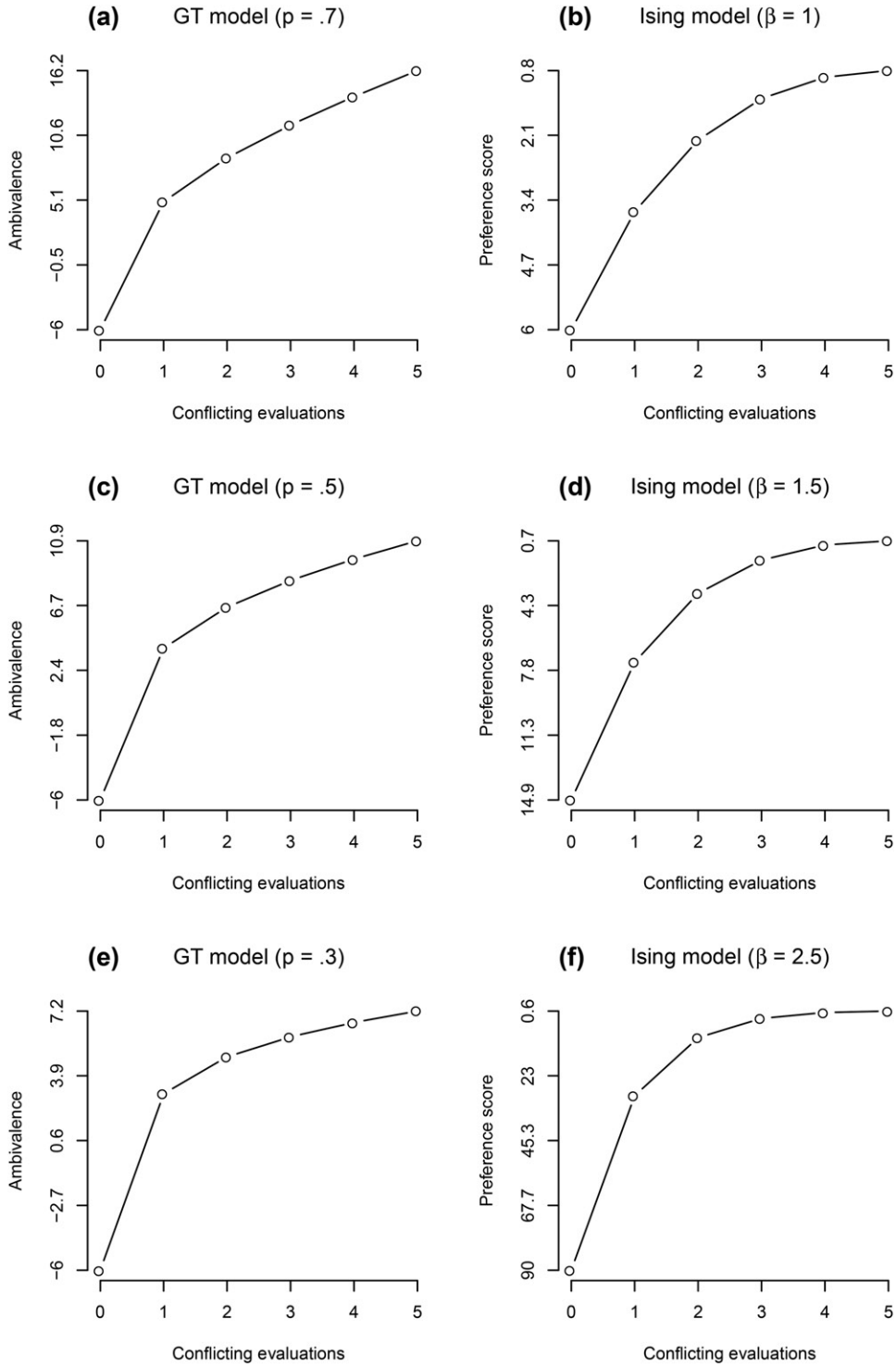


Figure 7. Similarity of ambivalence calculated by the gradual threshold (GT) model and mean preference scores of nodes in Ising networks.

dependence parameter can co-occur, we set up the following simulation.

Simulation 5: Felt ambivalence as Gibbs entropy. For this simulation we again used a fully connected 10-node network with all edge weights set to .1. We varied the thresholds in the following way: (a) all thresholds were set to 0 (representing a situation in which the external information points in no direction), (b) half of the thresholds were set to .1 and the other half were set to −.1 (representing a situation in which an individual receives weak mixed information about an attitude object), (c) half of the thresholds were set to .5 and the other half were set to −.5 (representing a situation in which an individual receives strong mixed information about an attitude object), (d) all thresholds were set to .1 (representing a situation in which the external information points in a weak positive direction), and (e) all thresholds were set to .5 (representing a situation in which the external information points in a strong positive direction). In addition, we varied the dependence parameter between 1, 1.5,

and 2.5 (mirroring Simulation 4). For each combination of thresholds and dependence parameter, we simulated 100 individuals, resulting in the total number of 1,500 simulated individuals. For each individual, we simulated 500 iterations based on Glauber dynamics (the network was again initialized randomly). To evaluate the stability of the attitude network, we calculated the percentage of flipped states for the last 100 iterations.

The results indicated that for $\beta = 1$, only the highly positive thresholds resulted in a relatively stable attitude network (see Figure 8). For $\beta = 1.5$, attitude networks were more stable overall, with the strong positive thresholds resulting in almost perfect stability. The highly mixed thresholds networks remained relatively unstable. For $\beta = 2.5$, only highly mixed thresholds networks did not approach perfect stability. It is our view that such a situation results in the strongest feelings of ambivalence, because stability remains rather low while the dependence parameter is already at a high value.

Based on the results of the simulation, we conclude that high felt ambivalence arises when individuals receive highly mixed information. Felt ambivalence is then amplified by the motivation to reduce attitudinal entropy. Such a situation would arise when individuals hold important attitudes for which they receive mixed information, for instance, when individuals are disposed to a given evaluation (e.g., holding liberal values because you work at a liberal university), whereas significant others endorse a different evaluation (e.g., having parents who hold conservative values). Such a situation has been shown to cause strong feelings of ambivalence (Priester & Petty, 2001). Further support for the relation between stability of an attitude and feelings on ambivalence comes from the finding that ambivalent individuals show physical signs of instability (i.e., moving from one side to the other; Schneider et al., 2013).

Prediction 12: Highly mixed information and high attitude importance result in strong felt ambivalence.

Cognitive Dissonance and Ambivalence Reflect Attitudinal Entropy

Apart from research on ambivalence, the implication that attitudinal entropy causes psychological discomfort is also relevant to research on cognitive dissonance. Similarities between cognitive dissonance and felt ambivalence have been noted by several researchers (e.g., Jonas, Broemer, & Diehl, 2000; McGregor, Newby-Clark, & Zanna, 1999); both concepts describe aversive feelings caused by being aware of incongruence of one's beliefs regarding an attitude object. The main difference between these two concepts concerns the situations by which they are caused (van Harreveld, van der Pligt, et al., 2009). Whereas felt ambivalence arises in situations in which attention is directed at an ambivalent attitude, cognitive dissonance arises in situation in which a univalent attitude is disturbed (e.g., by inducing behavior incongruent with an individual's attitude; Festinger and Carlsmith, 1959). However, the consequences of felt ambivalence and cognitive dissonance are similar. This point is illustrated by the similarities of two experiments focused on the role of arousal in dissonance reduction (Zanna & Cooper, 1974) and on biased information processing serving ambivalence reduction (Study 1; Nordgren, van Harreveld, & van der Pligt, 2006), respectively. In both experiments, participants were first administered a sugar pill but were told that the pill would make them feel either aroused or relaxed. The results in both experiments were similar: When participants were told that the pill would be relaxing, they showed dissonance reduction and biased information processing. In contrast, when they were told that the pill was arousing, participants showed neither dissonance reduction nor biased information processing. Both Zanna and Cooper (1974) and Nordgren et al. (2006) argued that the reason for this pattern of results is that participants attributed their negative feelings caused by cognitive dissonance or ambivalence to the effects of the pill. We take the results of these experiments as indication that negative feelings caused by cognitive dissonance and ambivalence in fact result from attitudinal entropy; the difference is that in cognitive dissonance paradigms entropy is induced and in ambivalence paradigms attention to high entropy attitudes is induced.

Prediction 13: Given that the AE framework assumes that felt ambivalence and cognitive dissonance are caused by aversive configurations of the attitude network in combination with high dependence, felt ambivalence and cognitive dissonance are predicted to have similar consequences.

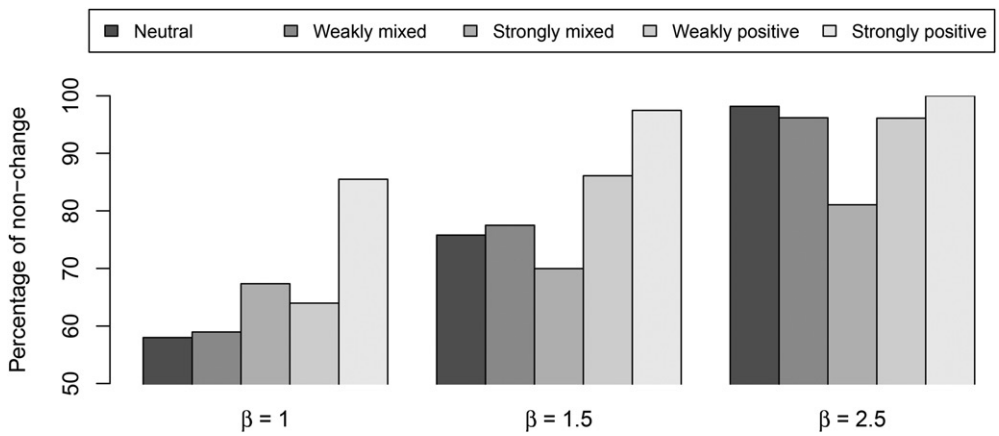


Figure 8. Stability of attitude networks based on different thresholds and dependence parameters.

Future Study of the AE Framework

In the remainder of this article, we address some important opportunities for future study of the AE framework. Apart from the empirical predictions that follow from the AE framework, we highlight the possibility of finding neural substrates of the AE framework's propositions and possibilities for further theoretical integration, and we discuss open questions raised by the AE framework.

Possible Neural Substrates of the AE Framework

Affective neuroscience has identified several neural substrates of attitude dynamics. Much of this research has focused on finding neural substrates of the reaction to valenced stimuli. This research has identified that the amygdala plays a central role in processing valenced stimuli (e.g., Morris et al., 1996; Phelps, 2006; Zald, 2003). Important to note, the amygdala seems to integrate information from throughout the brain (Cunningham & Zelazo, 2007), which makes it likely that global evaluations are formed in this neural structure. Another neural structure that plays a central role in attitude dynamics seems to be the anterior cingulate cortex (ACC). The ACC plays an important role in the detection of potential conflict (Carter et al., 1998), and it was shown that the ACC is active during the experience of cognitive dissonance (van Veen, Krug, Schooler, & Carter, 2009) and when ambivalent stimuli are processed (Cunningham, Raye, & Johnson, 2004). This makes the ACC a likely candidate for the neural structure involved in translating entropy of attitudes under high dependence into aversive feelings (note that also other neural substrates are likely to be involved in the processing of ambivalent stimuli, such as the insula, the temporal parietal junction, and the posterior cingulate cortex; see Nohlen, van Harreveld, Rotteveel, Lelieveld, & Crone, 2014).

Because the AE framework proposes that directing attention to and thinking about attitude objects serves the function of reducing attitudinal entropy, research on the neural substrates of consciousness is relevant to the AE framework. A recent influential theory of the neural underpinnings of consciousness posits that conscious experience results from neurons engaging in recurrent processing of stimuli, which enables information exchange between several low-level and high-level areas of the brain (Block, 2005, 2007; Lamme, 2003, 2006). It thus seems likely that conscious processing of attitude objects results from integrating different kinds of information regarding the attitude object. This idea is also in line with the information integration theory of consciousness (Tononi, 2004; Tononi & Edelman, 1998), which holds that the level of a system's consciousness depends on the amount of information this system integrates. This again underscores the importance of conscious thought in information integration. Information integration in turn is an important requirement for entropy reduction, thus further supporting the AE framework's assumption that a central function of conscious thought is to reduce attitudinal entropy.

The AE Framework's Relation to Other Models of Attitude

Although it is beyond the scope of our article to discuss the AE framework's relation to all prominent models of attitude, we discuss the framework's relation to three models that are in our view especially relevant: the Iterative Reprocessing (IR) model (Cunningham & Zelazo, 2007), the Attitude as Constraint Satisfaction (ACS) model (Monroe & Read, 2008) as an exemplar of constraint-satisfaction based connectionist models, and the Associative Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006). These models are especially relevant, because they are similar in focus as the AE framework. The basic assumption of the APE model is that evaluations tapped by implicit measures result from associative processes, whereas evaluations tapped by explicit measures result from propositional processes. The APE model further assumes that cognitive consistency is relevant only to propositional processes. Similarly, the AE framework holds that attitudinal entropy reduction, which is mostly pronounced during explicit processing of the attitude object, results in heightened cognitive consistency. However, the models diverge in the assumption that heightened cognitive consistency during explicit processing of the attitude object results from a process that is qualitatively different from implicit processing of the attitude. In this sense, the AE framework is more in line with the IR model and the ACS model, which both assume that implicit and explicit evaluations are based on the same processes.

As we mention in the introduction of the AE framework, the process by which complex attitudinal representations are reduced to a single global evaluation is partly based on the IR model, which assumes that global evaluations are the result of iterative reprocessing of the attitude object, serving the reduction of entropy (Cunningham, Dunfield, & Stillman, 2013). The AE framework has several similarities to the ACS model, as both models assume that the main driving factor in attitude dynamics is the drive for cognitive consistency. The ACS model and the AE framework also share a more technical similarity, because the ACS model is based on Hopfield (1982, 1984) neural networks, which in turn are based on Ising models. In our view, the ACS model and the AE framework are therefore likely to complement each other and have different weaknesses and strengths. A strong feature of the ACS model is that it provides a formalized account of evaluative learning, whereas the AE framework is more parsimonious than the ACS model, which in our view has two advantages: First, parsimony aids the objective of "understanding by building," in the sense that the more parsimonious the model, the more likely it is that we can come to an understanding of the modeled construct. Second, parsimony also aids the development of predictions, because parsimony of a model makes it also less variable. Ultimately, we think that important knowledge can be gained by integrating these different models of attitudes. Based on the similarities between the IR model, the ACS model, and the AE framework, we are optimistic that such integration is possible (for an integration of the IR model

and the ACS model, see Ehret, Monroe, and Read, 2015). As discussed in the introduction of the AE framework, we are currently working on such integration.

Open Questions

The AE framework fosters subsequent research on attitudes in two ways. First, as we discuss throughout this article, several predictions can be straightforwardly derived from the AE framework. Second, the AE framework also identifies several open questions, which we discuss next.

Open Question 1: The exact nature of attitude elements needs to be further investigated. In our earlier work on attitude networks (Dalege et al., 2016; Dalege, Borsboom, van Harreveld, van der Maas, 2017, 2018; Dalege et al., 2017) we treated rather general beliefs (e.g., judging a presidential candidate as honest) and feelings (feeling anger toward a presidential candidate), as well as concrete behaviors (voting for a presidential candidate) as attitude elements. However, it might also be possible that more low-level beliefs (e.g., episodic memories of a person acting in a specific way) and feelings (e.g., recalling situations in which a person made one feel in a given way) are alternative operationalizations of attitude elements.

Open Question 2: Although we have focused on determinants of entropy reduction, it is also relevant to investigate determinants that make individuals more tolerant to attitudinal entropy. A possible such determinant might be that individuals are highly motivated to be accurate.

Open Question 3: Can one level of attitudinal entropy reduction substitute for the other (e.g., is commitment to a given evaluation always necessary to reach higher levels of attitudinal entropy reduction or would something like relevance of the attitude to a decision be sufficient)?

Open Question 4: The AE framework assumes that attitudinal entropy is evaluated through two processes—the energy of a given attitudinal configuration and the instability of an attitude. However, the extent to which these processes are linked is a matter for future research.

Open Question 5: Although we discussed attitudinal entropy reduction mostly as an intrapersonal process, it is certainly also possible that there are interpersonal effects on attitudinal entropy reduction. A question needs to be addressed: How often individuals spontaneously reduce attitudinal entropy compared to how often this is socially instigated?

Open Question 6a: How pronounced are individual differences in attitudinal entropy reduction? Indirect evidence points to the existence of substantial differences, as individuals differ in their preference for consistency (Cialdini, Trost, & Newsom, 1995).

Open Question 6b: How pronounced are cultural differences in attitudinal entropy reduction? Similar to Open Question 6a, indirect evidence supports the hypothesis that cultural differences exist in how routinely individuals engage in attitudinal entropy reduction, as individuals from collectivistic cultures are less likely to experience cognitive dissonance than individuals from independent cultures (Heine & Lehman, 1997; Hoshino-Browne et al., 2005; Kitayama, Snibbe, Markus, & Suzuki, 2004).

Open Question 6c: Combining Open Questions 6a and 6b leads to the question of whether individuals might even differ qualitatively in attitudinal entropy reduction: Are there individuals who do not engage in attitudinal entropy reduction?

Open Question 7: In the current article we focused on single attitudes. Attitudes, however, do not exist in independence from one another, and future study of the AE framework should explore whether its principles also extend to interattitudinal processes.

Conclusion

In this article, we introduced the AE framework, which holds that (a) attitude inconsistency is entropy, (b) energy of attitude configurations serves as a local processing strategy to reduce the global entropy of attitude networks, and (c) directing attention to and thinking about attitude objects reduces attitudinal entropy by increasing the dependence parameter of attitude networks. The level of attitudinal entropy reduction depends on several factors, with merely directing attention to and thinking shortly about the attitude object representing the initial levels. Thinking more elaborately about an attitude object and commitment to an evaluation and relevance to decisions of the attitude represent the intermediate levels and high attitude importance represents the final level in attitudinal entropy reduction. We discussed the AE framework's relevance to research on ambivalence, the mere thought effect on attitude polarization, attitude strength, heuristic versus systematic persuasion, and implicit versus explicit measurements of attitude, thereby underscoring the integrative power of the AE framework. We also discussed several predictions that follow from the AE framework and several open questions identified by the AE framework. It is our view that because of its abilities in integration and spurring novel research questions, the AE framework represents a significant advancement in the theoretical understanding of attitudes. Furthermore, the AE framework places attitude dynamics into a broader dynamical systems context, further underscoring that reduction of entropy is the defining feature of living systems—both in a biological and a psychological sense. Ultimately, this might help to answer the question *why* it is that we think: to reduce the entropy of our mental representations.

References

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review*, *78*(3), 171–206.

Armitage, C. J., & Arden, M. A. (2007). Felt and potential ambivalence across the stages of change. *Journal of Health Psychology*, *12*(1), 149–158.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*(3), 668–688.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*(1), 53–58.

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, *9*(2), 46–52.
Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*, 481–499.
Boltzmann, L. (1877). Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht. [On the relationship between the second law of the mechanical theory of heat and the probability calculus]. *Wiener Berichte*, *76*, 373–435.
Boninger, D. S., Krosnick, J. A., & Berent, M. K. (1995). Origins of attitude importance: Self-interest, social identification, and value relevance. *Journal of Personality and Social Psychology*, *68*(1), 61–80.
Cacioppo, J. T., Petty, R. E., & Green, T. R. (1989). From the tripartite model to the homeostasis model of attitudes. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 275–310). Hillsdale, NJ: Lawrence Erlbaum.
Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–749.
Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing with and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.
Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology*, *69*(2), 318.
Clark, J. K., Wegener, D. T., & Fabrigar, L. R. (2008). Attitudinal ambivalence and message-based persuasion: Motivated processing of proattitudinal information and avoidance of counterattitudinal information. *Personality and Social Psychology Bulletin*, *34*(4), 565–577.
Cramer, A. O., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PloS One*, *11*(12), e0167490
Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2–3), 137–150; discussion 150-193. doi:10.1017/S0140525X09991567
Cunningham, W. A., Dunfield, K. A., & Stillman, P. E. (2013). Emotional states from affective dynamics. *Emotion Review*, *5*(4), 344–355.
Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience*, *16*(10), 1717–1729.
Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104.
Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude (CAN) model. *Psychological Review*, *123*(1), 2–22.
Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. (2017). Network analysis on attitudes: A brief tutorial. *Social Psychological and Personality Science*, *8*(5), 528–537.
Dalege J., Borsboom, van Harreveld F., & van der Maas H. L. J. (2018). A network perspective on attitude strength: Testing the connectivity hypothesis. *Social Psychological and Personality Science.* Advance online publication. doi:10.1177/1948550618781062
Dalege, J., Borsboom, D., van Harreveld, F., Waldorp, L. J., & van der Maas, H. L. J. (2017). Network structure explains the impact of attitudes on voting decisions. *Scientific Reports*, *7*, 4909.
Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, *25*(5), 582–602.
Ehret, P. J., Monroe, B. M., & Read, S. J. (2015). Modeling the dynamics of evaluation: A multilevel neural network implementation of the iterative reprocessing model. *Personality and Social Psychology Review*, *19*(2), 148–176.
Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Hillsdale, NJ: Lawrence Erlbaum.
Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition. research: Their meaning and use. *Annual Review of Psychology*, *54*(1), 297–327.
Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *J Abnorm Psychol*, *58*(2), 203–210.
Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading: Addison-Wesley.
Flay, B. R. (1978). Catastrophe theory in social psychology: Some applications to attitudes and social behavior. *Behavioral Science*, *23*(4), 335–350.
Fraiman, D., Balenzuela, P., Foss, J., & Chialvo, D. R. (2009). Ising-like dynamics in large-scale functional brain networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *79*(6 Pt 1), 061922
Fredrickson, G. H., & Andersen, H. C. (1984). Kinetic ising model of the glass transition. *Physical Review Letters*, *53*(13), 1244–1247.
Galam, S., Gefen, Y., & Shapir, Y. (1982). Sociophysics: A new approach of sociological collective behaviour. I. mean-behaviour description of a strike. *Journal of Mathematical Sociology*, *9*(1), 1–13.
Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731.
Gawronski, B., LeBel, E. P., & Peters, K. P. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*(2), 181–193.
Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, *43*(3), 300–312.
Gawronski, B., & Strack, F. (2012). Cognitive consistency as a basic principle of social information processing. In B. Gawronski & F. Strack (eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 1–16). New York, NY: Guilford Press.
Gilmore, R. (1981). *Catastrophe theory for scientists and engineers*. New York, NY: John Wiley.
Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, *132*(5), 778–822.
Glauber, R. J. (1963). Time-dependent statistics of the Ising model. *Journal of Mathematical Physics*, *4*(2), 294–307.
Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.
Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, *10*(4), 371
Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*, 9(4), 393–401. doi: 10.1177/1948550617711229
Heider, F. (1958). *The psychology of cognitive organization*. Hoboken, NJ: Wiley.
Heine, S. J., & Lehman, D. R. (1997). Culture, dissonance, and self-affirmation. *Personality and Social Psychology Bulletin*, *23*(4), 389–400.
Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*(2), 304–320.
Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385.
Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, *128*(1), 3–31.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8), 2554–2558.

Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(10), 3088–3092.

Hoshino-Browne, E., Zanna, A. S., Spencer, S. J., Zanna, M. P., Kitayama, S., & Lackenbauer, S. (2005). On the cultural guises of cognitive dissonance: The case of easterners and westerners. *Journal of Personality and Social Psychology*, *89*(3), 294–310.

Howe, L. C., & Krosnick, J. A. (2017). Attitude strength. *Annual Review of Psychology*, *68*, 327–351.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus [Contribution to the theory of ferromagnetism.]. *Zeitschrift Für Physik*, *31*(1), 253–258.

Jaynes, E. T. (1965). Gibbs vs Boltzmann entropies. *American Journal of Physics*, *33*(5), 391–398.

Jonas, K., Broemer, P., & Diehl, M. (2000). Attitudinal ambivalence. *European Review of Social Psychology*, *11*(1), 35–74.

Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly, 24*, 163–204.

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. New York: Oxford University Press.

Kim, B. J., Liu, J., Um, J., & Lee, S. I. (2005). Instability of defensive alliances in the predator-prey model on complex networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *72*(4 Pt 1), 041906

Kindermann, R., & Snell, J. L. (1980). *Markov random fields and their applications*. Providence, RI: American Mathematics Society.

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.

Krosnick, J. A., Boninger, D. S., Chuang, Y. C., Berent, M. K., & Carnot, C. G. (1993). Attitude strength: One construct or many related constructs? *Journal of Personality and Social Psychology*, *65*, 1132–1151.

Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Hillsdale, NJ: Lawrence Erlbaum.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284–308.

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18.

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501.

Latané, B., & Nowak, A. (1994). Attitudes as catastrophes: From dimensions to categories with increasing involvement. In R. R. Vallacher & A. Nowak (Eds.), *Dynamical systems in social psychology* (pp. 219–249). San Diego, CA: Academic Press.

LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, *37*(4), 570–583.

Liu, J. H., & Latané, B. (1998). The catastrophic linke between the importance and extremity of political attitudes. *Political Behavior*, *20*, 105–126.

McGregor, I., Newby-Clark, I. R., & Zanna, M. P. (1999). "Remembering" dissonance simultaneous accessibility of inconsistent cognitive elements moderates epistemic discomfort. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 325–352). Washington: American Psychological Association.

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), 733–759.

Newby-Clark, I. R., McGregor, I., & Zanna, M. P. (2002). Thinking and caring about cognitive inconsistency: When and for whom does attitudinal ambivalence feel uncomfortable?. *Journal of Personality and Social Psychology*, *82*(2), 157–166.

Nohlen, H. U., van Harreveld, F., Rotteveel, M., Lelieveld, G. J., & Crone, E. A. (2014). Evaluating ambivalence: Social-cognitive and affective brain regions associated with ambivalent decision-making. *Social Cognitive and Affective Neuroscience*, *9*(7), 924–931.

Nordgren, L. F., Van Harreveld, F., & Van Der Pligt, J. (2006). Ambivalence, discomfort, and motivated information processing. *Journal of Experimental Social Psychology*, *42*(2), 252–258.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition*, *19*(6), 625–666.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248.

Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*(1), 69–81.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*(5), 847–855.

Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research*, *10*(2), 135–146.

Priester, J. R., & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality and Social Psychology*, *71*(3), 431–449.

Priester, J. R., & Petty, R. E. (2001). Extending the bases of subjective attitudinal ambivalence: interpersonal and intrapersonal antecedents of evaluative tension. *Journal of Personality and Social Psychology*, *80*(1), 19–34.

Refling, E. J., Calnan, C. M., Fabrigar, L. R., MacDonald, T. K., Johnson, V. C., & Smith, S. M. (2013). To partition or not to partition evaluative judgments: Comparing measures of structural ambivalence. *Social Psychological and Personality Science*, *4*(4), 387–394.

Rosenberg, M. J., Hovland, C. I., McGuire, W. J., Abelson, R. P., & Brehm, J. W. (1960). *Attitude organization and change: An analysis of consistency among attitude components*. New Haven, CT: Yale University Press.

Schneider, I. K., Eerland, A., van Harreveld, F., Rotteveel, M., van der Pligt, J., van der Stoep, N., & Zwaan, R. A. (2013). One way and the other: The bidirectional relationship between ambivalence and body movement. *Psychological Science*, *24*(3), 319–325.

SchröDinger, E. (1944). *What is life*? Cambridge: Cambridge University Press.

Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review*, *103*(2), 219–240.

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of preferences by constraint satisfaction. *Psychological Science*, *15*(5), 331–336.

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The emergence of coherence over the course of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1250–1260.

Simon, D., Snow, C. J., & Read, S. J. (2004). The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*(6), 814–837.

Simon, D., & Spiller, S. A. (2016). The elasticity of preferences. *Psychological Science*, *27*(12), 1588–1599.
Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending cold and hot cognitions. *Journal of Personality and Social Psychology*, *109*(3), 369–394.
Sitnov, M. I., Sharma, A. S., Papadopoulos, K., & Vassiliadis, D. (2001). Modeling substorm dynamics of the magnetosphere: From self-organization and self-organized criticality to nonequilibrium phase transitions. *Physical Review E*, *65*, 016116.
Smith, M. B., Bruner, J. S., & White, R. W. (1956). *Opinions and personality*. Oxford, UK: John Wiley & Sons.
Tesser, A. (1978). Self-generated attitude change. *Advances in Experimental Social Psychology*, *11*, 289–338.
Tesser, A., & Conlee, M. C. (1975). Some effects of time and thought on attitude polarization. *Journal of Personality and Social Psychology*, *31*(2), 262–270.
Tesser, A., & Leone, C. (1977). Cognitive schemas and thought as determinants of attitude change. *Journal of Experimental Social Psychology*, *13*(4), 340–356.
Tesser, A., Martin, L., & Mendola, M. (1995). In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 73–92). Hillsdale, NJ: Lawrence Erlbaum.
Millar, M. G., & Tesser, A. (1986). Thought-induced attitude change: The effects of schema structure and commitment. *Journal of Personality and Social Psychology*, *51*(2), 259–269.
Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, *383*(6603), 812–815.
Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, *57*, 27–53.
Thom, R. (1972). *Stabilitè structurelle et morphogenesè [Structural Stability and Morphogenesis]*. New York, NY: Benjamin.
Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5*, 42.
Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science (New York, N.Y.)*, *282*(5395), 1846–1851.
van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. doi: 10.1037/0033-295X.113.4.842
van der Maas, H. L. J., Kolstein, R., & van der Pligt, J. (2003). Sudden transitions in attitudes. *Sociological Methods & Research*, *32*(2), 125–152.
Van Harreveld, F., Rutjens, B. T., Rotteveel, M., Nordgren, L. F., & Van Der Pligt, J. (2009). Ambivalence and decisional conflict as a cause of psychological discomfort: Feeling tense before jumping off the fence. *Journal of Experimental Social Psychology*, *45*(1), 167–173.
van Harreveld, F., van der Pligt, J., & de Liver, Y. N. (2009). The agony of ambivalence and ways to resolve it: introducing the MAID model. *Personality and Social Psychology Review : an Official Journal of the Society for Personality and Social Psychology, Inc*, *13*(1), 45–61.
van Veen, V., Krug, M. K., Schooler, J. W., & Carter, C. S. (2009). Neural activity predicts attitude change in cognitive dissonance. *Nature Neuroscience*, *12*(11), 1469–1474.
Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews*, *41*(1), 88–123.
Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology*, *29*(5), 703–709.
Zanna, M. P., & Rempel, J. K. (1988). Attitudes: A new look at an old concept. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The psychology of knowledge* (pp. 315–334). Cambridge: Cambridge University Press.
Zeeman, E. C. (1976). Catastrophe theory. *Scientific American*, *234*(4), 65–83.

Appendix

Supplementary Information on Simulation 1b

Table A1. Probability of assignment of subject with given threshold to a given group in Simulation 1b.

	Thresholds									
Group	−0.2	−0.1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
1	.19	.17	.15	.13	.11	.09	.07	.05	.03	.01
2	.17	.15	.14	.12	.11	.09	.08	.06	.05	.03
3	.15	.14	.13	.12	.11	.09	.08	.07	.06	.05
4	.13	.12	.12	.11	.10	.10	.09	.08	.08	.07
5	.11	.11	.11	.10	.10	.10	.10	.09	.09	.09
6	.09	.09	.10	.10	.10	.10	.10	.11	.11	.11
7	.07	.08	.08	.09	.10	.10	11	.12.	.12	.13
8	.05	.06	.07	.08	.10	.11	.12	.13	.14	.15
9	.03	.05	.06	.08	.09	.11	.12	.14	.15	.17
10	.01	.03	.05	.07	.09	.11	.13	.15	.17	.19