Perspectives on Psychological Science
1–39

DOI: 10.1177/17456916231204579
www.psychologicalscience.org/PPS

Toward a General Framework of Biased Reasoning: Coherence-Based Reasoning

Dan Simon[1,2] and Stephen J. Read[2]
[1]Gould School of Law, University of Southern California, and [2]Department of Psychology, University of Southern California

Abstract
A considerable amount of experimental research has been devoted to uncovering biased forms of reasoning. Notwithstanding the richness and overall empirical soundness of the bias research, the field can be described as disjointed, incomplete, and undertheorized. In this article, we seek to address this disconnect by offering "coherence-based reasoning" as a parsimonious theoretical framework that explains a sizable number of important deviations from normative forms of reasoning. Represented in connectionist networks and processed through constraint-satisfaction processing, coherence-based reasoning serves as a ubiquitous, essential, and overwhelmingly adaptive apparatus in people's mental toolbox. This adaptive process, however, can readily be overrun by bias when the network is dominated by nodes or links that are incorrect, overweighted, or otherwise nonnormative. We apply this framework to explain a variety of well-established biased forms of reasoning, including confirmation bias, the halo effect, stereotype spillovers, hindsight bias, motivated reasoning, emotion-driven reasoning, ideological reasoning, and more.

Keywords
bias, cognition, coherence-based reasoning, emotion, affect, judgment, decision-making, motivation, goals, reward, social cognition, thinking, reasoning, unconscious, automatic processing

Reasoning is central to the human experience in that it establishes the basis for comprehending the world, forming behaviors, and making choices. Over the past half century, a considerable amount of experimental research has been devoted to "biased reasoning," a term that we take to capture systematic deviations from normative forms of processing (see Arkes, 1991; Klayman, 1995; Nickerson, 1998). One of the striking features of this body of research—also dubbed the "error paradigm" (Funder, 1995)—is the sheer volume of biases reported in the literature. For illustration, Krueger and Funder (2004) listed 42 types of bias in the domain of social judgment, and Baron (2008, Table 2.1) listed 53 different biases in the field of judgment and decision-making. Probably most comprehensive is the Cognitive Bias Codex, which lists no fewer than 188 discrete biases (Manoogian, 2016).

The richness of the bias observations stands in stark contrast to the field's theoretical footing. Biases have been typically reported as isolated and unique phenomena, with scant attention paid to their commonality with adjacent biases. Little effort has been devoted to grounding them in a comprehensive theoretical framework. However commendable, the few extant theorizations tend to be inconsistent with one another. For example, biases have been viewed as a way of minimizing cognitive effort and simplifying processing (Dawes, 1976; Nisbett & Ross, 1980; Shah & Oppenheimer, 2008; H. A. Simon, 1955), collateral costs of otherwise beneficial cognitive adaptations (Arkes, 1991; Krueger & Funder, 2004; Nisbett & Ross, 1980), a way to avoid costlier errors (Haselton et al., 2015; Johnson et al., 2013; Krizan & Windschitl, 2007; Nickerson, 1998), and a trade-off between the goals of error avoidance and effort minimization (Payne et al., 1992; Weber et al., 1995).

The landscape is richer—but hardly methodic—with respect to the proposed mechanisms that drive biased

Corresponding Author:
Dan Simon, Gould School of Law and Department of Psychology, University of Southern California
Email: dsimon@law.usc.edu

reasoning. For illustration, a number of research programs have listed memory failings as the primary driver while referring to a divergent array of memory-related phenomena, including skewed memory search and retrieval (Dougherty et al., 1999; Edwards & Smith, 1996; Forgas, 2008; Kunda, 1990), noisy deviations in memory-based information processes (Dougherty et al., 1999; Hilbert, 2012; Usher & McClelland, 2001), erroneous assessments of memory strength (Tidwell et al., 2016), and limited capacity in working memory (De Neys & Verschueren, 2006). Other biases have been explained as related to nonsystematic evidence evaluation, such as positive test strategy (Klayman & Ha, 1987; Krizan & Windschitl, 2007; Nickerson, 1998; Wason, 1960), differential scrutiny (Edwards & Smith, 1996; Evans & Curtis-Holmes, 2005; Evans et al., 1983; Pyszczynski & Greenberg, 1987), and selective stopping (Nickerson, 1998; Nisbett & Ross, 1980). A number of biases have been said to be driven by several different concomitant mechanisms (e.g., Krizan & Windschitl, 2007; Nickerson, 1998).

This disjointed, incomplete, and undertheorized array of empirical observations makes for an impoverished scientific landscape. As noted by Lichtenstein and colleagues (1982; see also Tidwell et al., 2016), the bias literature can be deemed a product of "dustbowl empiricism," which borrows its imagery from the drifting dust and tumbleweeds in the drought-stricken American Midwest of the 1930s (see Nesbitt-Larking & Kinnvall, 2012; Paley, 2008). This research tradition has also been likened to a form of hunting and gathering empiricism (De Houwer et al., 2011).

Article Overview

In this article, we seek to address the disconnect between the undeniably valuable empirical observations offered by the biased reasoning research and its lacking theoretical basis. We propose to fill that gap by suggesting that "coherence-based reasoning" offers a parsimonious theoretical framework that explains a sizable number of important and widely studied deviations from normative forms of reasoning. This goal dictates the two objectives of this article.

The first objective is to present the coherence-based reasoning paradigm as a vital and pervasive tool of human cognition. Let us foreshadow the framework's first claim. Dating back to gestalt psychology (Wertheimer, 1923/1938) and Brunswik's (1955) lens model, it has been widely accepted that drawing inferences and making judgments invariably require taking into consideration multitudes of cues. Frequently, those cues are numerous, ambiguous, incongruent, incommensurable, and uncertain. Coherence-based reasoning serves to transform such complexity into comprehensible and sensible conclusions, which extricates people from the taxing state of conflict and enables them to engage successfully with their environment (see Janis & Mann, 1977).

By this account, reasoning tasks are understood to be represented in connectionist networks of interconnected nodes and processed through constraint-satisfaction mechanisms. These processes are driven by structural forces that transform complex representations into states of equilibrium, or coherence. Coherence is defined as the state wherein positively linked elements are similarly activated and negatively linked elements have dissimilar activations, with the winning conclusion being supported by highly activated attributes and the rejected conclusion and its related attributes receiving low activation. Coherence is attained through a coherence-maximizing process that alters the nodes—whether by way of strengthening, weakening, or morphing them—to bring them into line with the emerging conclusion. It must be emphasized that connectionist networks are fundamental to the functioning of the brain and that coherence-based reasoning serves as a ubiquitous, essential, and overwhelmingly adaptive cognitive apparatus.

This brings us to our second objective. Not unlike many other tools in people's mental toolbox, cognitive skills evolved to operate successfully in certain tasks, but they might not stack up to the demands of other contexts (see e.g., Kahneman et al., 1982). Indeed, our second objective is to establish that coherence-based reasoning can provide a generalized framework that helps explain a sizable number of important deviations from normative forms of reasoning. Our framework suggests that this otherwise adaptive cognitive process can readily result in biased conclusions when the network is dominated by nodes or links that are incorrect, overweighted, or otherwise nonnormative. To foreshadow our argument, coherence-based reasoning will naturally drive the process toward the most coherent (i.e., strongly activated) state of the cognitive representation. But normativity is often exogeneous to the process, and when it deviates from the preferred cognitive conclusion, the outcome will be determined by their relative strength. At times, normativity will be overcome by coherence, thus resulting in biased outcomes. It is important to keep in mind that the resulting biases arise not from dysfunctional processing but from the normal operation of the hardware of the brain.

Before we delve into the framework, we note four clarifications and limitations. First, we take as given the literature's findings of biased reasoning without questioning its empirical footing. These biases have largely withstood critical inquiry and have survived the test of time. The objective of the article is to provide this vast and unruly body of literature with a theoretical framework that will help integrate and explain some of the major biases.

Second, we acknowledge that no single theoretical framework could possibly explain each and every one of the vast array of biases that have been recognized by Krueger and Funder (2004), Baron (2008), Manoogian (2016), and others. Our project is humbler than that. We demonstrate our framework by applying it to a select group of biases that are important, familiar, well studied, and highly cited (accounting in total for hundreds of thousands of citations on Google Scholar).

Third, the explanatory power of the coherence framework is particularly germane to reasoning processes that seek to reach a discrete conclusion from multiple features—such as perceptions, beliefs, estimations, preferences, and emotions. Hence, the framework focuses on processes that rely on an integrative task, a feature that accounts for the vast majority of meaningful reasoning tasks. This focus marks an important limitation of our framework, in that it does not examine the sources of those attributes that are brought to the task. In the terms of Brunswik's (1952) lens model, our framework does not dwell on the origin of the cues that represent the distal object but, instead, on how those cues are processed and integrated to produce a representation of that object. The breadth of the framework stems from the fact that many of the most important and ubiquitous forms of human bias fall into this category. Still, to some extent, coherence-based reasoning can influence which cognitions are more likely to be brought to the task, primarily by way of selectively searching and retrieving information based on its propensity to cohere with a desired or emerging conclusion (Bhatia, 2016; Festinger, 1957; Fraser-Mackenzie & Dror, 2009; Jekel et al., 2018). Because of length constraints, this subject is not be explored in this article, although, as discussed below, we encourage that it be addressed in future research.

Fourth, this framework does not seek to contest and undercut all of the literature's extant theories and explanatory mechanisms. Various aspects of our framework are compatible with and partly overlap with work by Arkes (1991), Fischhoff (1975), Morewedge and Kahneman (2010), Wilson and Brekke (1994), and probably others. Our framework resonates most with the recently published work by Oeberst and Imhoff (2023), which we discuss in some detail in the Summary and Implications section. In that section, we will also discuss the relationship between our framework and Bayesian reasoning. In all, we maintain that the coherence account should be deemed the preferred explanation thanks to its parsimony, breadth, and biologically plausible mechanism (see Thagard, 2019).

Norms of Reasoning

In setting the normative standards against which to judge human performance, the field of biased reasoning has largely eschewed the stylized rules of formal logic (cf. Chater & Oaksford, 2012) and the consistency conditions prescribed by rational-choice theory (cf. Luce & Raiffa, 1957). Rather, the field has implicitly adopted informal principles of appropriate reasoning that correspond to the slew of judgments and inferences that people experience in everyday life (Shafir & LeBoeuf, 2002). We discern these principles as clustering roughly into three norms.

First, the norm of justified backing instructs that conclusions should be based only on the task attributes and should be impervious to factors that are task-irrelevant, task-improper, or disproportionately weighted. For example, syllogistic reasoning ought not be influenced by exogenous beliefs (see section on Belief Bias below), the probability of an occurrence should not be affected by its desirability (see section on Motivated Reasoning below), and social judgments should not be affected by the perceiver's preceding emotional state (see section on Emotion-Based Bias below). Second, the norm of invariance instructs that task attributes should remain unchanged, absent an interaction with logically relevant factors. For example, a person's trait should not be affected by other nonplausibly related traits (see section on the Halo Effect below). Finally, the norm of directionality instructs that the reasoning process should proceed exclusively from attributes and premises to conclusions, and not the other way around. For example, evaluation of incoming evidence should not be affected by a preconceived conclusion (see section on Confirmation Bias below), and the causal effect of antecedent events should not be affected by the eventual outcome information (see sections on Hindsight Bias and Outcome Bias below).

These normative principles do not purport to offer analytic exactitude. For one, a norm may vary depending on the nature and context of the task. For example, it would be appropriate to rely on the stereotype "men are taller than women" in concluding that a randomly selected man is taller than a randomly selected woman. However, when a given woman and man are known to have been drawn from samples of equal average heights—not to mention when the sample of women is taller than the sample of men—relying on that stereotype would be deemed biased reasoning (Nelson et al., 1990; Sá et al., 1999). We also note that a given bias may be considered violative of more than one norm. We trust that even absent a formal taxonomy, our readers should encounter little difficulty identifying when reasoning processes violate the proposed norms.

The core function of coherence-based reasoning is to drive the reasoning process toward conclusions that attain the greatest level of coherence given the underlying set of attributes. Thus, coherence-based reasoning is merely a process model, and coherence is a value-neutral property. When all the nodes and links that make up the representation of the task are appropriate, that conclusion will be deemed normative. Undoubtedly, most human-reasoning processes fall into that category. But in some reasoning processes, inappropriate nodes or links win the coherence battle and thus sway the conclusion. Note that we see the thumbprint of coherence-borne biases in grave societal ills, such as racial discrimination (see Eberhardt, 2020), false justifications for going to war (Risen, 2006), heightened political and ideological polarization (Ditto et al., 2019; Stanley et al., 2020), convictions of innocent people (D. Simon, 2012), and more.

Theoretical Framework: Coherence-Based Reasoning

We turn now to the article's first objective of expounding on the coherence-based reasoning framework, which heretofore has not been the subject of a comprehensive review (for limited reviews, see Holyoak & Powell, 2016; D. Simon & Holyoak, 2002). As mentioned, reasoning typically requires taking into consideration multitudes of cues. For example, forming an impression of a person you just met at a dinner party could easily involve cues derived from their affect, age, gender, facial appearance, clothing, accent, wit, linguistic sophistication, apparent socioeconomic class, and much more. That judgement is bound also to involve the reasoner's beliefs and attitudes toward those cues, such as an attraction to people's facial appearance, distaste of their clothing, enjoyment of their wit, and discomfort with their social status. Performing this social-judgment task must somehow integrate this vast array of features into a discrete judgment of the person. We propose that this integration is performed through coherence-based reasoning driven by constraint-satisfaction processes.

Structural dynamics

Coherence-based reasoning originates from a research tradition of structural dynamics (Markus & Zajonc, 1985; Zajonc, 1968), captured most prominently by cognitive-consistency theories (Abelson et al., 1968; Abelson & Rosenberg, 1958; Festinger, 1957; Heider, 1946, 1958; McGuire, 1968). Structural dynamics is premised on the Gestaltian notion that cognition operates holistically rather than elementally (Wertheimer, 1923/1938). As Fritz Heider (1960) explained, cognitive states are to be understood as structural configurations that are determined by their "whole-qualities" rather than as "properties of the[ir] parts" (p. 168). Solomon Asch (1946b) and Abelson and Rosenberg (1958) emphasized that the psychological features of a construct are determined by the organization and interactions that hold its constitutive features together.

Structural properties spur dynamic forces. Positive interrelatedness generates cohesive forces that stabilize structures in states of equilibrium (Rosenberg & Abelson, 1960; Tannenbaum, 1968). Heider (1958) explained that structural configurations seek to settle at states in which their constitutive elements "co-exist without stress" (p. 176), that is, states in which "all parts of a unit have the same dynamic character (i.e., all are positive, or all are negative), and entities with different dynamic character are segregated from each other" (Heider, 1946, p. 107). On the other hand, opposing dynamics destabilize the structure and generate configural forces that seek to restore balance (Heider, 1958; Zajonc, 1983). Consistent with the principle of homeostasis (Rosenberg, 1968), once the system returns to a state of balance, the structure stabilizes, and the configural forces recede (Rosenberg & Abelson, 1960). The concept of equilibrium has been labeled "Prägnanz" (Wertheimer, 1923/1938), "balance" (Heider, 1946, 1958), "congruity" (Osgood & Tannenbaum, 1955), and "consonance" (Festinger, 1957). Following Thagard (1989; Holyoak & Thagard, 1989), we resort to the term "coherence."

Except for the most trivial of instances, the features of a cognitive task do not naturally line up in uniform support of a singular conclusion. Thus, coherence must be constructed, and this construction lies at the heart of this framework. The coherence-maximizing process generates structural forces that reconstruct (Rosenberg & Abelson, 1960) or distort (Asch, 1940) the cognitive elements to bring them to cohere with the network. As Asch (1940) explained, cognitive processing entails changes both in "the judgment of the object" and in "the object of judgment" (p. 455). As the elements give rise to a conclusion, that emerging conclusion reshapes the elements to better fit with it. Hence, the bi-directional nature of the reasoning process (Holyoak & Simon, 1999; Read et al., 1997).

With the advent of modern connectionist architectures (e.g., Hopfield, 1982, 1984; McClelland et al., 1986; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), various researchers, such as Paul Thagard (1989, 2019) and others (Holyoak & Thagard, 1989; Read et al., 1997; Read & Miller, 1994), have noted the parallels between structural dynamics and constraint-satisfaction processes, thus connecting Gestaltian holism to the connectionist architecture of the human brain.

Connectionist representation

The connectionist view of mental representation posits a brain-like neural network in which each of the task elements is represented symbolically as a node (Holyoak, 1991; Hummel & Holyoak, 2003; Thagard, 2019). Elements include any feature that might be involved in a task, including perceptions, preferences, factual propositions, beliefs, evaluations, affective reactions, and so on. An element's level of activation corresponds to the strength with which it is held. The activation is initially set by the element's corresponding background knowledge, preferences, and affective strength. Thus, elements that stand for the person's central attitudes (Judd & Krosnick, 1982) or moral mandates (Skitka, 2010) will be highly activated. Conversely, peripheral beliefs and tenuously held preferences will be weakly activated.

All elements are connected to other elements via relational links through which activation spreads. The existence and type of link will be informed by the person's background knowledge about how features in the world relate to one another. Links are either positive or negative. Positive links connect elements that are deemed to "go together." Elements can go together based on association, causality, similarity, proximity, and the like. In his formulation of balance theory, Heider (1946) proposed that affective elements such as like, love, esteem, and value go together. Note that elements will be considered to go together even absent any inherent relationship as long as they happen to be associated with the same conclusion in the particular task at hand. Conversely, elements will be negatively linked when they are negatively associated or deemed as opposites, such as friend and foe and love and hate. Here, too, elements will be deemed as not going together even absent any inherent relationship as long as they are associated with opposing conclusions in the particular task at hand.

Complex-reasoning tasks are represented in multiple layered networks (Eliasmith, 2013; Smolensky, 1989) in which each level emerges from the level beneath it (Eliasmith, 2013; Thagard, 2019). The uppermost level—labeled the "output layer"—contains the possible conclusions of the reasoning process, such as inferences, judgments, choices, and decisions. Lower levels contain proximal attributes that feed directly into those conclusions. In inference tasks, these will include the premises and evidence items on which the inferences are based, and in judgment tasks, these will include the perceptions and evaluations of the target of judgment. Links can be positive or negative and they can vary in strength. Every element involved in the task will be connected directly to some other elements and through them, to the vying conclusions and ultimately, to the entire representation of the task. Thus, each element constrains the network and is constrained by it in turn.

In all but easy and obvious instances, the task will afford more than one plausible conclusion, which means that the network will entail some form of competition among rivaling conclusions. Competition is often built into the task in that the process requires endorsing one conclusion and rejecting its rivals. Such will be the case, for example, when attempting to judge whether another person is friendly or unfriendly, or when facing a decision which of two capable candidates should be offered the job. Competition implies that the network contains subsets of elements such that the one composed of a_1, $a_2, \ldots a_n$ lends support to Conclusion A, whereas the subset consisting of elements b_1, $b_2, \ldots b_n$ lends support to the rival Conclusion B (assuming just two conclusions). The vying conclusions and their interlinked elements form "attractors," which are local areas of the network that attract a great deal of activation and to which the network will eventually gravitate (Churchland & Sejnowski, 1992; Hopfield, 1982; Read et al., 1997). In reasoning tasks, attractors will typically form around hypotheses or plausible conclusions, but they can also coalesce around any prominently activated feature (or cluster of features), such as a preexisting belief, an emotional state, or a motivated goal.

Constraint-satisfaction processing

Constraint-satisfaction processing occurs through bidirectional activation between interlinked elements. These interactions influence the elements' activation in that positively linked elements (that go together) will excite one another, whereas negatively linked ones will be mutually inhibitory. The degree of activation will be a function of the level of activation of the respective nodes and the weight and sign of the links that bind them. Cross-activation will spread spontaneously and in parallel through all of the connected nodes and thus throughout the entire network. Crucially, the cross-activation alters the nodes to maximize the similarity of activation among positively linked nodes. The process will terminate when the entire network asymptotes or "relaxes" at an attractor or state that reaches the highest level of coherence that can be attained for the given network. Thus, coherence corresponds to the state at which the constraints of the network are maximally satisfied.

The overall effect is that the elements associated with the emergent winning attractor are boosted and those associated with the rejected attractor are suppressed. As a result, the network morphs into a lopsided representation in which the strongly endorsed conclusion is confidently chosen over its rejected rival.

The changes that elements undergo as they transform from the original state of conflict and ambiguity to a strong fit with the conclusion has been labeled "coherence shifts" (Holyoak & Simon, 1999; D. Simon, 2004), and the product of these processes amount to "the coherence effect" (Read & Simon, 2012; D. Simon et al., 2015). Constraint-satisfaction processing is deemed an all-purpose mechanism that both integrates information and provides the selection rule for reaching conclusions (see Glöckner & Betsch, 2008). This model has been applied to a wide range of cognitive processing, ranging from low levels of cognition, such as vision (McClelland & Rumelhart, 1981) and perception (McClelland et al., 2014), to high-level processing (Bhatia, 2016; Freeman & Ambady, 2011; Glöckner & Betsch, 2008; Glöckner et al., 2010, 2014; Holyoak & Thagard, 1989; Kunda & Thagard, 1996; MacDonald & Seidenberg, 2006; Suri et al., 2020; Thagard, 2019).

Three aspects of the process are noteworthy. First, although the reasoning process is usually initiated consciously by the reasoner, the process's core mechanism—the spread of activation and construction of coherence—proceeds for the most part automatically (Arkes, 1991). Constraint-satisfaction processing will typically proceed virtually effortlessly, be difficult to control and to stop, and sail mostly beneath the level of conscious awareness (see Bargh, 1989).

Second, the susceptibility of elements to structural forces is not uniform (see Klayman, 1995; Kunda, 1990). Strong and unequivocally held propositions are less likely to be altered or dislodged by structural forces, whereas weakly activated and ambivalently held propositions are more prone to morph to better fit the conclusion. It is also noteworthy that elements are not boundlessly mutable. People tend to avoid reaching conclusions that are blatantly violative of their belief system or of prevailing norms (Boiney et al., 1997; Kunda & Thagard, 1996).

Third, we emphasize that parallel constraint satisfaction is a fundamental aspect of brain-based processing. Drawing on well-understood mathematical models from physics, various researchers (Amit, 1989; Hertz et al., 1991; Hopfield, 1982, 1984; O'Reilly et al., 2020; Rumelhart et al., 1986) have shown that networks of simple elements, such as neurons interconnected through bidirectional relationships, will spontaneously organize over time to minimize the energy of the system. That is, any system with these same characteristics as a brain will spontaneously self-organize to minimize its "energy" by way of maximizing its coherence.

The basic process of constraint satisfaction can be captured by a simple mathematical equation proposed by Hopfield (1982, 1984). If neurons represent cognitions and the weight between them represents whether the two cognitions are consistent or inconsistent with each other, then the energy of the system of cognitions can be captured by the Hopfield energy equation:

$$-\frac{1}{2}\Sigma_j \Sigma_i x_i w_{ij} y_j$$

where x_i and y_i represent the activation of two cognitions and w_{ij} is the weight between them. The total energy of the system is equal to the sum of the energy of each possible pair of nodes or cognitions. Given the minus sign for the equation, energy is minimized to the extent that the activations of the two elements in a pair are consistent with the sign of the weight between them. So, two positively activated nodes with a positive or excitatory link between them (+++) or a positive and a negative node with a negative link between them (+−−) would decrease energy, and two negative nodes with a negative link (−−−) or two positive nodes with a negative link (+−+) would increase energy.

Other authors (e.g., Smolensky, 1989) have shown that one can view this energy minimization as equivalent to coherence or harmony maximization simply by removing the minus sign to get the following equation, in which consistency among pairs of elements increases the value and inconsistency among pairs of elements decreases its value (see Read et al., 1997; Rumelhart et al., 1986; Thagard, 1989):

$$\frac{1}{2}\Sigma_j \Sigma_i x_i w_{ij} y_j$$

This simple function specifies that the coherence of the network can be captured by taking all pairs of nodes that are connected, multiplying the activation of each node in the pair, then multiplying that result by the weight between them, and then summing the result for all pairs of nodes. Because this value is maximized when the weight between two nodes is consistent with the product of the activation of the two connected nodes (e.g., two positive nodes and a positive weight between them or one negative and one positive node with a negative weight between them), this network will adjust the activations until the network has reached the maximum value possible given the constraints. That is, to the extent possible, the network will seek to change any inconsistent relationships (the triple of two nodes and the weight between them) to make them consistent.

As described below, we observe a strong overlap between findings of coherence obtained from experimental participants and computerized constraint-satisfaction simulations (D. Simon et al., 2015, Studies 3, 4).

Word recognition

To exemplify the operation of constraint-satisfaction processing, we turn to a classic application to word recognition (McClelland, 1985; McClelland & Rumelhart, 1981). The recognition of a word can be represented by a network containing iteratively emerging layers, ascending from distal attributes, through proximate attributes, to conclusions. The bottom input layer will contain markings on the page, which can be identified as features of letters in the English language. At the proximate-attribute level, those features will constitute recognizable letters, and at the output level, the letters will be recognized as a word. Under ideal conditions, processing such a network amounts to a straightforward task. When the text symbols are clear (e.g., "H," "A," and "T") and the word they constitute ("hat") is familiar to the reader, skilled readers will perform the task with miniscule effort and in lightning speed. Thanks to the reader's proficiency in the English language, there is no real competition in this network because all elements involved are positively linked both within and between each layer: The letter features combine to form the letters "H," "A," and "T," and those letters readily combine to make the word "hat." Effectively, the network is coherent from the start.

Not all cognitive tasks are so simple or obvious. Take, for example, the task of reading the series of markings in Figure 1 (note, it is a noun).

This task is notable in that each of the markings is ambiguous, making for a network beset with uncertainty and competition. At the letter level, the letters "B" and "D" compete for the interpretation of the first and third markings, and the letters "E" and "F" compete for the middle marking. At the word level, eight alternative conclusions (BFD, DFB, BED, etc.) compete for prominence. Because the viability of any hypothesis is inversely related to the acceptability of its rivals, all horizontal links will be represented as inhibitory activations.

Fig. 1. Obscure word (following McClelland, Rumelhart, & Hinton, 1986).

The network cross-activates bidirectionally between and within the conclusion level, the letter level, and the letter-features level, with inputs from semantic background knowledge. Ultimately, the process yields the word "bed" as the only combination that is consistent with knowledge of the English language. That conclusion serves also to induce a top-down (or "backwards") inference that the three ambiguous markings constitute the letters "B," "E," and "D." Thus, the word "bed" and its constitutive letters "B," "E," and "D" all become accepted as the correct interpretation, which means that they are all highly activated in the network and thus mutually coherent. By the same token, the rival interpretations "D," "F," and "B" and the seven alternative word combinations are suppressed.

The ability to solve such seemingly intractable tasks is immensely useful. But even as the process's prowess is celebrated, it must also be appreciated that the successful performance was enabled through cognitively altering—or, in Solomon Asch's (1940) words, distorting—the perception of the ambiguous markings into the recognizable letters "B," "E," and "D." Two important observations follow. First, the constraint-satisfaction process is an equal-opportunity distorter in that it alters elements that stand in the way of coherence regardless of their normative status. Second, although the normal course of reading proceeds from recognizing letters to recognizing words, the process operates also in the reverse direction, altering elements to support the conclusion. This bidirectionality (see Holyoak & Simon, 1999) renders the process susceptible to backward reasoning. Both observations appear repeatedly in our survey of the biased forms of reasoning.

Higher-level reasoning

To be sure, the power of coherence-based reasoning extends well beyond reading three-letter words. We argue that this mechanism enables people to successfully confront large, complex, and often cacophonous decisions. The process should be credited with affording the largely successful making of considerably more

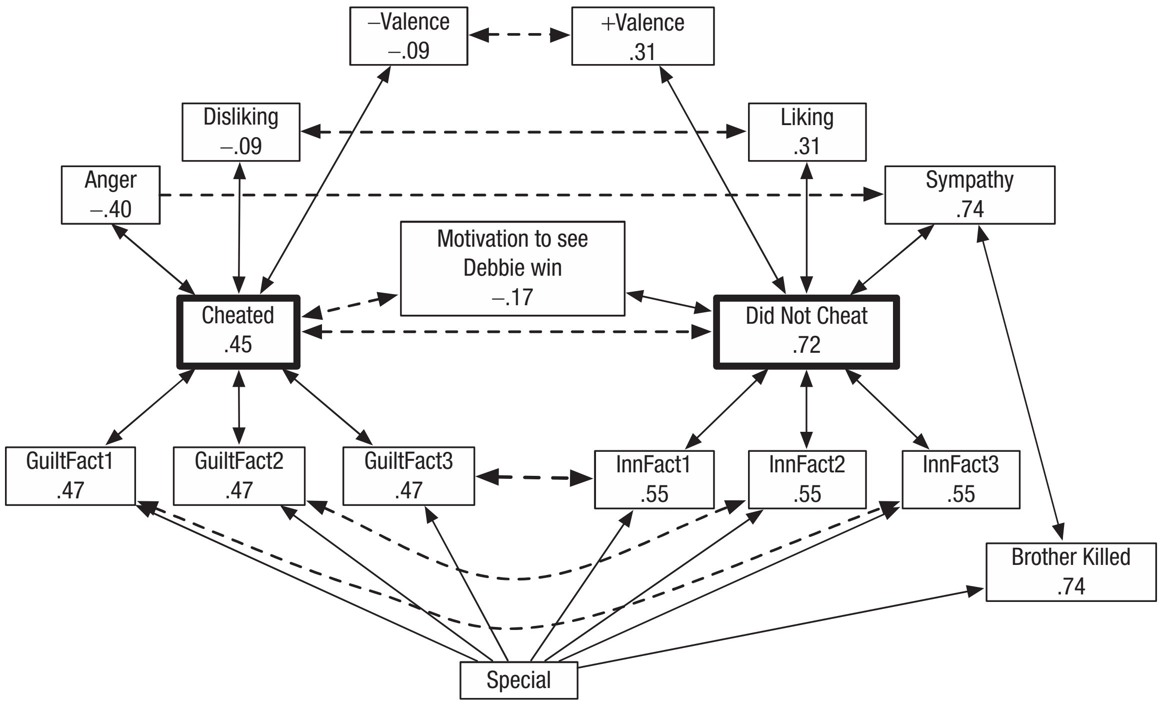

Fig. 2. Network of the coherence simulation in Study 3 of D. Simon et al. (2015) for the sympathy condition (brother-killed manipulation). Numbers in the nodes represent the activations after the network is run (range = −1 to 1).

complex judgments, such as the aforementioned evaluation of the stranger at the dinner party, deciding which career to pursue, whom to choose as a partner, and which home to buy.

To illustrate an application of coherence-based reasoning to higher-level cognition, we constructed and ran a constraint-satisfaction neural network to capture the judging of an allegation of academic misconduct steeped in intricate and ambiguous evidence. The example is based on a case in which a university conducts a misconduct hearing to determine whether a student, Debbie Miller, cheated on an exam (see D. Simon et al., 2015, Studies 3, 4). These studies were designed to test whether constraint-satisfaction networks enmesh "hot" and "cold" cognitions. To run the simulation, we used the "CS" (constraint satisfaction) module in O'Reilly et al.'s (2012) emergent neural-network-modeling system. This program is specifically designed to capture the behavior of a constraint-satisfaction system based on bidirectional connectivity among concepts (represented as nodes). This kind of network could also be specified and run in other bidirectional architectures, such as Thagard's (1989) ECHO or emergent/Leabra by O'Reilly et al. (2020). For a more detailed account of how we constructed these models, see the Supplemental Material available online.

In Study 3 (Fig. 2), we informed half of the participants that the student's brother had previously been killed by a drunken driver. This was expected to manipulate participants' sympathy to side with the student. In Study 4 (Fig. 3), we provided half of the participants with information of rampant cheating that was hurting the university and other students. This was expected to manipulate participants' motivation to side with the university. The bottom layer of each network contains a "Special" node that provides activation to the input elements, that is, the facts and the manipulations (sympathy in Fig. 2 and motivation in Fig. 3). The next layer up contains the facts of the case, followed by the mutually inhibitory conclusions ("cheated" vs. "did not cheat") and then three pairs of hot-cognition measures. The models reveal that activation spread throughout the network and settled at a stable pattern of activation. In Figure 2, increasing sympathy toward the student resulted in higher judgments that she did not cheat, more exculpating factual interpretations, higher valence toward seeing her prevail, less anger and greater liking of her (and lower judgments of all elements related to a guilt finding). In Figure 3, increasing motivation to find the student guilty resulted in opposite judgments throughout. Note that comparing the activation values of the nodes derived from the model with the corresponding mean ratings of the behavioral data provided by participants yielded a correlation of .68.

Experimental evidence

A harbinger of experimental evidence supporting coherence-based reasoning can be found in the remarkably

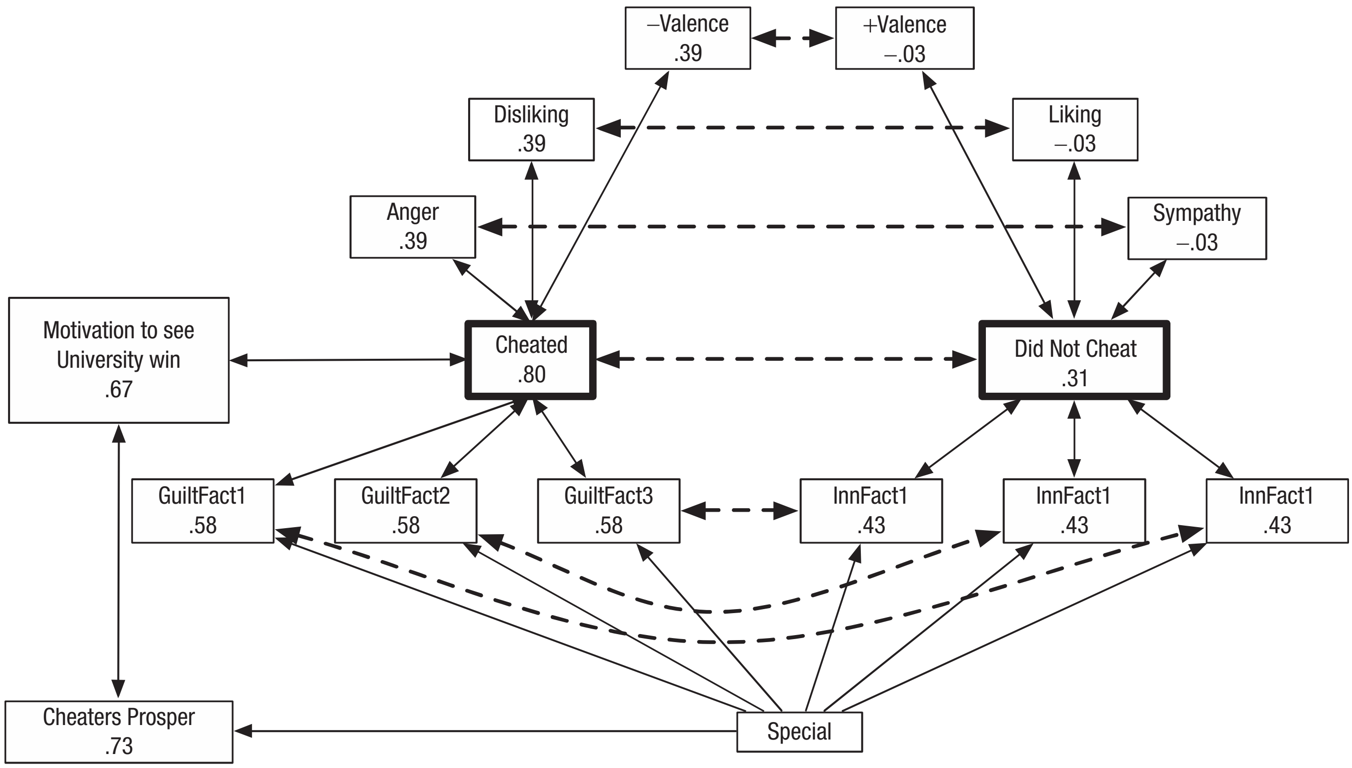

Fig. 3. Network of the coherence simulation in Study 4 of D. Simon et al. (2015) for the "cheaters prosper" manipulation. Numbers in the nodes represent the activation of the node after the network is run.

simple and underappreciated study by Esch (as cited in Heider, 1958, pp. 176–177). Esch presented his participants with the following vignette and asked them to opine how the protagonist (in this case, Bob) would react:

> Bob thinks Jim very stupid, and a first class bore. One day Bob reads some poetry he likes so well that he takes the trouble to track down the author in order to shake his hand. He finds that Jim wrote the poems. (Esch, as cited in Heider, 1958, p. 176)

Bob's discovery of Jim's authorship seems to have generated a state of incoherence, presumably because of the intuition that shallow people cannot create good poetry. Consistent with Heider's (1946, 1958) balance theory, the study illustrates people's aversion to imbalanced states because virtually all participants expected Bob to alter his representation of the situation. The study found that 46% of participants indicated that Bob would improve his view of Jim (e.g., "He grudgingly changes his mind about Jim"), and 29% indicated that he would downgrade his judgment of the poetry (e.g., "He decides the poems are lousy"; Esch, as cited in Heider, 1958, p. 176). Most remaining participants resorted to a variety of other means to impose balance (e.g., "Bob would question Jim's authorship of the poems"). These distortions of the premises demonstrate the power of coherence-borne structural forces on human reasoning. This study serves also to demonstrate the concept of epistemic coherence, that is, the human tendency to impose coherence for the mere purpose of making sense of the world (Festinger, 1957; Heine et al., 2006; Kruglanski, 1990; Thagard & Verbeurgt, 1998).

To gain both a broader and more detailed appreciation of coherence-based reasoning, we turn now to a series of studies that were designed to test the proposition that the coherence framework captures a central aspect of human reasoning and decision-making across a wide range of tasks. We briefly review seven of the core research findings (for more detailed reviews, see Read & Simon, 2012; D. Simon, 2004; D. Simon & Holyoak, 2002).

Coherence shifts. The central finding in our coherence studies is that attributes are altered over the course of the process in that they shift toward greater coherence with the emerging conclusion and with all the other attributes: Attributes supporting the chosen conclusion become more strongly activated, whereas attributes supporting the rejected option attenuate in strength (Chaxel et al., 2016; DeKay, 2015; DeKay et al., 2011; Engel et al., 2020; Engel & Glöckner, 2013; Glöckner et al., 2010; Holyoak & Simon, 1999; Russo et al., 1996, 1998; D. Simon et al., 2015, 2020; D. Simon, Krawczyk, & Holyoak, 2004;

Table 1. Intercorrelations Among the Six Points of Dispute and the Ultimate Verdict in Holyoak and Simon (1999, Study 1) at Pretest and Posttest

	Cause	Motive	Regulation	Speech	Analogy	Verdict
Pretest						
Truth	−.02	.35*	.11	.19	.10	.09
Cause		−.06	.28	−.09	−.30*	.14
Motive			.07	−.14	.03	.05
Regulation				.18	−.05	.11
Speech					.24	.42**
Analogy						−.03
Posttest						
Truth	.34*	.50**	.31*	.40**	.39**	.54**
Cause		.55**	.45**	.34*	.53**	.68**
Motive			.39**	.24	.47**	.52**
Regulation				.43**	.43**	.69**
Speech					.57**	.63**
Analogy						.68**

*$p < .05$. **$p < .01$.

D. Simon, Snow, & Read, 2004; D. Simon & Spiller, 2016; Spellman et al., 1993). Coherence shifts are readily observed by comparing participants' rating of task attributes (e.g., factual inferences, evaluations, beliefs, preferences, probabilities) at two points in time (within-subjects design): once in isolated vignettes (pretest) and once when those attributes are implicated in a mental process, such as reasoning, forming a judgment, or making a decision (posttest). A key feature of this design is that the attributes are designed to be virtually independent of one another so that any interaction that emerges among them is best understood as driven indirectly, that is, via structural forces.

The first measure of coherence effects is derived by simply averaging the evaluations of the attributes. We found consistently that the attributes that go with the eventual winning conclusion increase in strength from pretest to posttest, whereas the attributes associated with the rejected conclusion wane. A second measure of coherence is performed by comparing the intercorrelations between attributes at the pretest and posttest phases. Given the study design, we expect that the intercorrelations will be weak at the pretest phase but will be strong and all-encompassing at the posttest phase. Table 1 depicts the intercorrelations observed in Holyoak and Simon (1999, Study 1). In this study, participants were asked to decide a libel legal case that contained six points of dispute. For each attribute, we measured the correlation with all other attributes and with the eventual verdict. As Table 1 shows, in the pretest phase, the correlations were mostly weak and nonsignificant, with one of the three significant correlations going the wrong way. In contrast, the posttest correlations are consistently strong and highly significant. This discrepancy captures the notion of coherence shifts, by which the loose assortment of unrelated attributes is transformed by coherence-maximizing forces toward a state of strong interrelatedness.

High confidence. Despite the initial balanced strength of the attributes (as indicated by participants' mixed pretest evaluations), by the end of the process, participants consistently reported high levels of confidence in whichever conclusion they reached (Holyoak & Simon, 1999; D. Simon, Krawczyk, & Holyoak, 2004; D. Simon, Snow, & Read, 2004; D. Simon & Spiller, 2016). Strong confidence is consistent with the lopsided representations of the task that follow from the coherence shifts. Note that the confidence levels are found to correlate with the magnitude of the coherence shifts in that the more participants distort the attributes (between pretest and posttest), the more confident their conclusions are (D. Simon, Snow, & Read, 2004; D. Simon & Spiller, 2016). It also has been shown that confidence ratings can be quantitatively predicted very well by neural-network models of coherence-based decision-making (i.e., the parallel constraint-satisfaction model for decision-making; see Glöckner & Betsch, 2012; Glöckner et al., 2014).

Indirect and extraneous influences. The connectionist nature of coherence-based reasoning could give rise to circuitous influences between elements even when they share no logical relation with each other. Recall that in connectionist networks, each node is interconnected with other nodes and through them with the entire network. It follows that manipulating a single task attribute has the potential to trigger a cascade of activation that can sway the conclusion and sweep all other attributes along. The structural forces generated by these cascades will mean that attributes can get activated even by attributes with which they have no logical relation. This intuition was put to the test in a series of studies (Holyoak & Simon, 1999; D. Simon et al., 2015; D. Simon, Krawczyk, & Holyoak, 2004; D. Simon, Snow, & Read, 2004; D. Simon & Spiller, 2016). For illustration, in a study by Holyoak and Simon (1999, Study 3), participants were presented with a task of deciding a legal case in which a company called Quest brought a libel suit against one of its shareholders, Jack Smith. Quest claimed that its collapse was caused by negative rumors that Smith posted about it on an Internet-based forum. The case revolved around six disputed issues, one of which was whether Smith intended to harm the company. To test the aforementioned indirect-influences hypothesis, participants received information about Smith's history—either that he was a conscientious investor or that he was a financial shark who ruined companies to enable him to purchase them on the cheap. As predicted, providing a malevolent (vs. benevolent) history resulted in stronger judgments of Smith's bad motive and also increased the verdicts against him. Note that this manipulation also altered participants' evaluations of the five other points of disputes, which had little or no logical relation to Smith's prior conduct, such as the magnitude of harm caused by his statement, an interpretation of the company's bylaws, and the importance of free speech on the Internet (Holyoak & Simon, 1999). In a criminal case involving an alleged employee theft, adding DNA evidence implicating the suspect not only increased verdicts of guilt but also affected other judgments, such as the reliability of an eyewitness, the strength of an alibi testimony, and the possibility that the suspect harbored a grudge against his employer (D. Simon, Snow, & Read, 2004). In a choice between job offers, adding the fact that one of the offices was located in a fun part of town not only increased choices of that job offer but also influenced participants' preferences for unrelated attributes, such as the size of the office, the salary offered, and the vacation package (D. Simon, Krawczyk, & Holyoak, 2004; D. Simon & Spiller, 2016).

Coherence spreads far. The connectionist nature underlying coherence-based reasoning would also suggest that the spreading activation could reach far afield, much like William McGuire's (1960) imagery of shaking a loose-link fence. To test this intuition, Holyoak and Simon (1999) explored whether coherence could spread from one task to another. One of the six issues in the aforementioned Quest dispute was whether the Internet should be deemed more like a telephone system (which, by law, precludes liability for libel) or like a newspaper (which, by law, is open to libel liability). After deciding the Quest case, participants were given a second case involving a contract dispute at a company called Infoscience, which runs an Internet-based bulletin board. Infoscience's labor contract stipulated that one of the two factors that should determine the bonus the employees deserve would be the level of bonuses paid at similar local information-service firms. One party (counterbalanced) claimed that the Internet is best likened to a telephone system and thus argued that the bonus at Infoscience should match the bonus paid by the local telephone company. The other party claimed that the Internet is best likened to a newspaper, arguing that the Infoscience bonus should match that paid by the local newspaper publisher. Thus, the Analogy factor was a shared point of dispute that served as a bridge between the two cases. Figure 4 presents the results of a structural equation model in which the arrows follow the major predicted flow of influence along the chain of inference. It shows links of influence starting with the experimentally manipulated variable, Smith's history, to Smith's motive, to the Quest verdict, to each of the other five points of dispute, including the Analogy (telephone vs. newspaper). These findings amount to an indirect influence, as discussed in the preceding section. But Figure 4 shows that the activation went further than just swaying the entire Quest network; it also spilled into the lnfoscience case, influencing both the Infoscience verdict and the other factor involved in the Infoscience bonus decision (who deserves the credit for the company's performance). In sum, all indirect effects throughout both cases were found to be significant, including the most remote, six-step indirect influence from Smith's history all the way to the judgments of the Infoscience case.

Coherence runs deep. Coherence not only spreads far but also runs deep. Consider, for example, the application of our framework to the data produced in the canonical study of Lord et al. (1979), which we discuss in detail below. That application will show that coherence effects seep down through four layers of the representation, which is a testament to the sheer force of coherence effects. In our own studies, we used multilayered stimuli and found similar effects. For example, in the aforementioned criminal case involving employee theft, one of the evidence items presented against him was that soon after the incident, he paid back a debt he owed his credit-card company. The man, in turn, claimed that the money was a repayment from his sister of a loan he had given her to fund her flower store. He explained that he could not prove the loan repayment because in the flower industry, transactions are typically done in cash. As predicted, we observed that coherence emerged between the verdict, the judgment of the source of the funds, and the underlying belief of whether transactions in the flower industry are typically done in cash (D. Simon, Snow, & Read, 2004). As predicted, these effects were swayed by the addition of

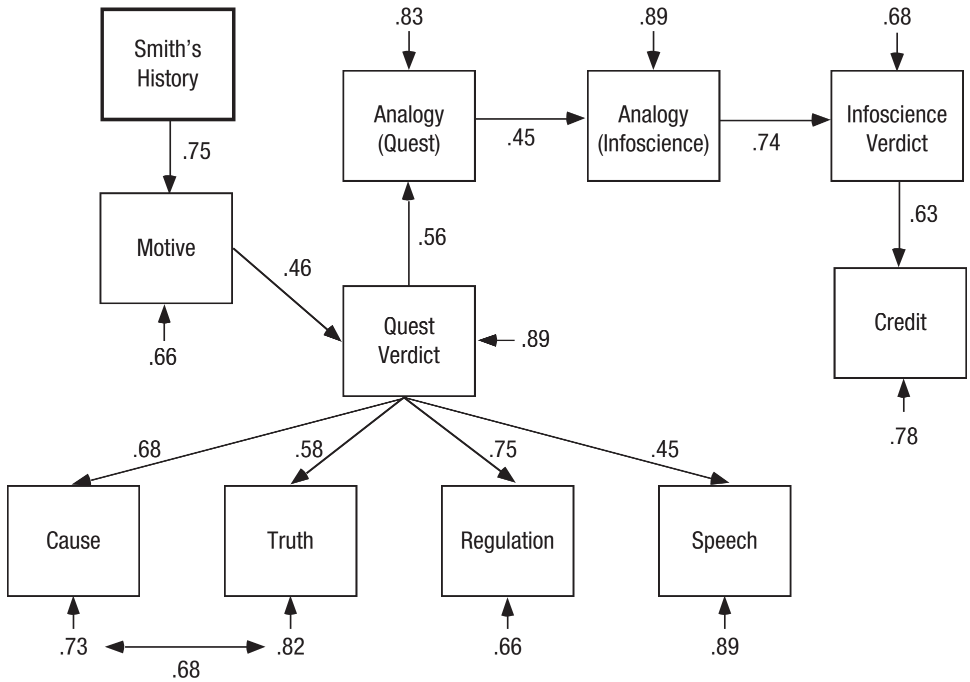

Fig. 4. Graphical summary of the final Equations with Software (EQS) model of coherence shifts in the transfer experiment. Note coherence spreading from the experimental manipulation of Smith's character to the verdicts and all points of dispute in both the Quest and Infoscience cases (Holyoak & Simon, 1999, Experiment 3).

DNA evidence so that the presence of DNA ended up seeping down to affect what people believed on the esoteric (and likely utterly unfamiliar) issue of financial practices in the flower industry.

In another experimental variation, we sought to test whether coherence would spread into the background beliefs that inevitably inform the inferences relating to the task (on the role of background-knowledge structures in reasoning, see Read, 1987; Schank & Abelson, 2013; Wyer & Radvansky, 1999). In this vein, in addition to the usual measures capturing the strength of the attributes, we also included measures of the background knowledge that is bound to inform those judgments. Thus, participants were asked not only to assess the accuracy of a particular eyewitness identification but also how accurate eyewitnesses are in general (D. Simon, Snow, & Read, 2004). Likewise, in a study concerning a case of academic misconduct, participants were asked not only whether a particular Professor Stone was gullible but also whether professors generally tend to be gullible (D. Simon et al., 2015). As predicted, coherence spread from top to bottom, affecting the background beliefs about eyewitnesses and professors in general. The coherence shifts were swayed by the inclusion of extraneous variables, such as that a piece of DNA evidence was added to the mix (in the criminal case) and that the suspected student's brother was killed by a drunk driver some months before the incident (in the academic-misconduct case).

Coherence enmeshes cold and hot cognitions. In the studies discussed thus far, coherence was observed emerging among the chosen conclusion and judgments of various task attributes, all of which would generally be classified as "cold cognitions" (see Abelson, 1963; Lepper, 1994; Sorrentino & Higgins, 1986). Following Thagard (2006) and Lewis (2005), we set out to explore the prospect that coherence would also engulf elements deemed to be "hot cognitions" (see Abelson, 1963; Lepper, 1994; Sorrentino & Higgins, 1986). In that vein, we measured a range of hot cognitions, including emotions (anger and sympathy toward the protagonist), liking of the protagonist, motivation toward the outcome of the case, and valence (feeling good vs. bad) toward either case outcome. These studies (D. Simon et al., 2015) found that coherence spreads across all dependent variables, both cold and hot (see also D. Simon et al., 2020). As described above, we also manipulated hot cognitions to test whether they could trigger coherence shifts and influence cold cognitive judgments. We found that generating sympathy toward Debbie Miller (by informing participants that her teenage brother was killed by a drunk driver) increased judgments of her innocence, alongside a host of concordant hot and cold cognitions. Likewise, the experimental treatment designed to increase participants' motivation to side with the university resulted in equally coherent but roughly opposite evaluations of both hot and cold cognitions. The coherence explanation of enmeshing cold and hot cognitions is compatible with variants of appraisal theory that posit a dynamic and bidirectional interaction between cognitive appraisal and emotional arousal (Frijda, 1986; Keltner et al., 1993; Roseman & Smith, 2001; Scherer, 1984, 2001).

Limited awareness. One of the key features of the coherence effect is that it occurs largely beneath the level of conscious awareness. Our studies (Holyoak & Simon, 1999; D. Simon & Spiller, 2016) found that by the end of a task, participants were incapable of accurately recounting their pretest ratings of the task attributes. Instead, those retrospective ratings were found to approximate the post-coherence-shift ratings. Mirroring previous findings in social psychology (e.g., Bem & McConnell, 1970; Goethals & Reckman, 1973; M. Ross et al., 1981), our participants seemed to believe that their current beliefs were more or less the same ones they held all along, suggesting a lack of awareness of the shifts (Holyoak & Simon, 1999). This finding was obtained even when participants were incentivized monetarily for accurate recall of their original ratings (D. Simon & Spiller, 2016). This metacognitive failure promotes a sense of stability and consistency in one's belief and attitudinal systems, thus helping to maintain naive realism (discussed below) and a rational view of the self (Pronin et al., 2004; Uhlmann & Cohen, 2007).

To summarize the experimental findings, the coherence effect has proven to be a robust and multifaceted empirical phenomenon. It has been demonstrated experimentally in laboratories located in a number of countries and across a wide range of cognitive tasks, including attitude change (Spellman et al., 1993), legal reasoning (Holyoak & Simon, 1999; D. Simon et al., 2001), drawing factual inferences (Engel & Glöckner, 2013; Lundberg, 2007; D. Simon et al., 2015; D. Simon, Snow, & Read, 2004), choices based on probabilistic inferences (Glöckner et al., 2010), analogical reasoning (Holyoak & Thagard, 1989), causal reasoning (Holyoak & Simon, 1999), moral reasoning (Holyoak & Powell, 2016), social judgment (Read & Marcus-Newhall, 1993; D. Simon et al., 2015), financial auditing (Lundberg, 2007; Phillips, 2002), attribution of blame (D. Simon et al., 2015; D. Simon, Snow, & Read, 2004), decision-making (Carpenter et al., 2016; D. Simon, Krawczyk, & Holyoak, 2004; D. Simon & Spiller, 2016), and gambling (Brownstein et al., 2004). Coherence also appears to be incorporated in a commonly held theory of mind in that people impute coherence-based reasoning to other actors (Esch, as cited in Heider, 1958; D. Simon et al., 2020).

Applying Coherence-Based Reasoning to Biased Processing

We now turn to the article's second objective of demonstrating that coherence-based reasoning can serve as an explanatory framework for a wide range of biased reasoning. A key to the following analyses is that normativity is exogenous and orthogonal to coherence. The distinctiveness of these properties is made clear by the aforementioned study by Esch (as cited in Heider, 1958). Recall that after learning that Jim—who was theretofore dismissed as simplistic—turned out to be the author of the beautiful poems, almost half of the participants adopted a more favorable view of him. By doing so, they imposed a state of coherence, that is, that Jim is both smart and writes beautiful poetry. The same goes for participants who responded to that revelation by downgrading their judgment of Jim's poetry. They too reached a state of coherence in that they came to view Jim as both shallow and the writer of lousy poetry. From a strictly cognitive perspective, the two types of reactions are indistinguishable in that they both reach a state of coherence.

These mirroring shifts, however, look very different when seen through the exogenous normative prism. There is nothing wrong with the first group's updating the judgment of a person after being exposed to new and relevant information. Indeed, failing to do so could well be considered unjustified close-mindedness. The same cannot be said for those participants who responded to the new evidence by downgrading the judgment of the poetry. Normatively speaking, most people would agree that the judgment of a poem should not hinge on one's impression of the poet. Thus, Bob had no valid reason for turning sour on Jim's poetry. In other words, based on criteria of sound reasoning, the coherence shift displayed by the first group would be judged appropriate, whereas the second group's symmetrical and cognitively comparable coherence shift should be deemed inappropriate. This difference stems from the fact that normativity is exogenous to the cognitive-processing mechanisms.

The exogeneity of normative reasoning is true for all forms of biased reasoning reviewed here. Constraint-satisfaction processes operate ubiquitously across a vast range of cognitive domains, rarely implicating normative standards. In the domain of reasoning, however, the conclusions that emanate from coherence-based reasoning will be measured against norms of appropriate reasoning and on occasion, will be found to be in violation of them.

Recall that coherence will typically drive the network to cohere around attractors. Attractors are likely to form around plausible intuitions or hypotheses, as well as around strongly activated features such as preexisting beliefs, emotional states, or motivated goals. While the former will often represent normative reasoning outcomes, the latter will frequently run against norms of reasoning. Effectively, the coherence-maximization process often amounts to a competition between normative and nonnormative attractors. When the latter prevails over the former, the reasoning process will produce biased results.

The following sections of the paper are devoted to its second objective, namely, to demonstrate that coherence-based reasoning can help explain specific forms of biased reasoning. The demonstration will unfold in two parts. We first examine reasoning processes that mostly implicate cold cognitive processing and then proceed to examine reasoning that implicates hot cognitions. It warrants emphasizing that both cold and hot cognitions are often concurrently involved in reasoning tasks (Abelson, 1963; Kunda, 1990; D. Simon et al., 2015, 2020; Thagard, 2006), as they are across the human experience (Lepper, 1994; Sorrentino & Higgins, 1986). This separation is intended for presentational convenience alone.

Cold-cognition coherence

Confirmation bias. A central tenet of normative reasoning is that conclusions be based on the attributes and premises of the task. Accordingly, the norm of directionality mandates that the inferences proceed from the attributes and premises toward the conclusion and not the other way around. Yet a large and persistent body of research reveals that incoming evidence and premises are frequently distorted by the reasoner's existing beliefs, expectations, or hypotheses (Arkes, 1991; Klayman, 1995; McKenzie, 2006; Nickerson, 1998). The primary variant of this family of biases is the confirmation bias.

We begin with the classic study by Lord et al. (1979), which sought to establish the phenomena of biased assimilation and attitude polarization. We focus here on the former, which speaks to the fact that incoming evidence can be distorted (assimilated) to support preexisting beliefs. Normatively, one would expect that beliefs would be attenuated by exposure to evidence that is ambiguous, mixed, or contradictory. The researchers hypothesized instead that such an exposure might end up distorting the evidence and even bolstering the preexisting attitudes and beliefs. To test this hypothesis, the researchers recruited participants who held strong positions on the death penalty, both proponents and opponents. Each participant was presented with two fictitious criminological studies that claimed to test the death penalty's deterrent effect. The studies were counterbalanced for both the methodology used and for the results that they purportedly reached (confirming vs. disconfirming the deterrent effect). Each of the studies was also accompanied by a critique and a rebuttal of the critique. Participants were then asked for various judgments of the studies: how convincing, how well done, and their effect on the participants' belief in the practice's deterrent effect. They were asked also for their updated attitudes and beliefs toward the death penalty. The study found that participants' reactions to the studies were highly contingent on whether the studies' putative findings comported with their preexisting beliefs. Supportive studies were deemed to have been well done and convincing, whereas studies that contradicted their beliefs were viewed as poorly done and unconvincing. By the end of the process, participants also reported more extreme beliefs with respect to the death penalty's deterrent effect and expressed stronger attitudes toward the practice. The end result, then, was that exposure to mixed evidence strengthened, rather than weakened, preexisting attitudes and beliefs. It is notable that participants in the Lord et al. study displayed strong coherence effects in that their attitudes, beliefs, and judgments of the studies all coalesced to form a coherent concordant set of propositions, whether in support or in opposition to the practice.

To further explore the applicability of coherence-based reasoning to this study, we simulated the behavioral results using Thagard's (1989) ECHO program, which is a computational implementation of constraint-satisfaction processes. Figure 5 presents the results of the simulation for participants who held positive attitudes toward the death penalty. A simulation of the anti-death-penalty participants would yield the same pattern with inverse values and thus be redundant.

Each of the nodes in the network represents a key proposition in the participants' network, organized in four layers. The top two layers represent participants' attitudes toward the death penalty and their belief in its deterrent effect, respectively. The bottom two layers represent participants' evaluations of the studies' support of the deterrent effect and their judgments about the studies themselves (convincingness, well conducted), respectively. Lines represent the relationships between the nodes. Solid double-headed arrows represent supportive or excitatory bidirectional connections (weight of .04), and the dashed double-headed arrows represent contradictory or inhibitory bidirectional connections (weight of –.06) between the nodes.

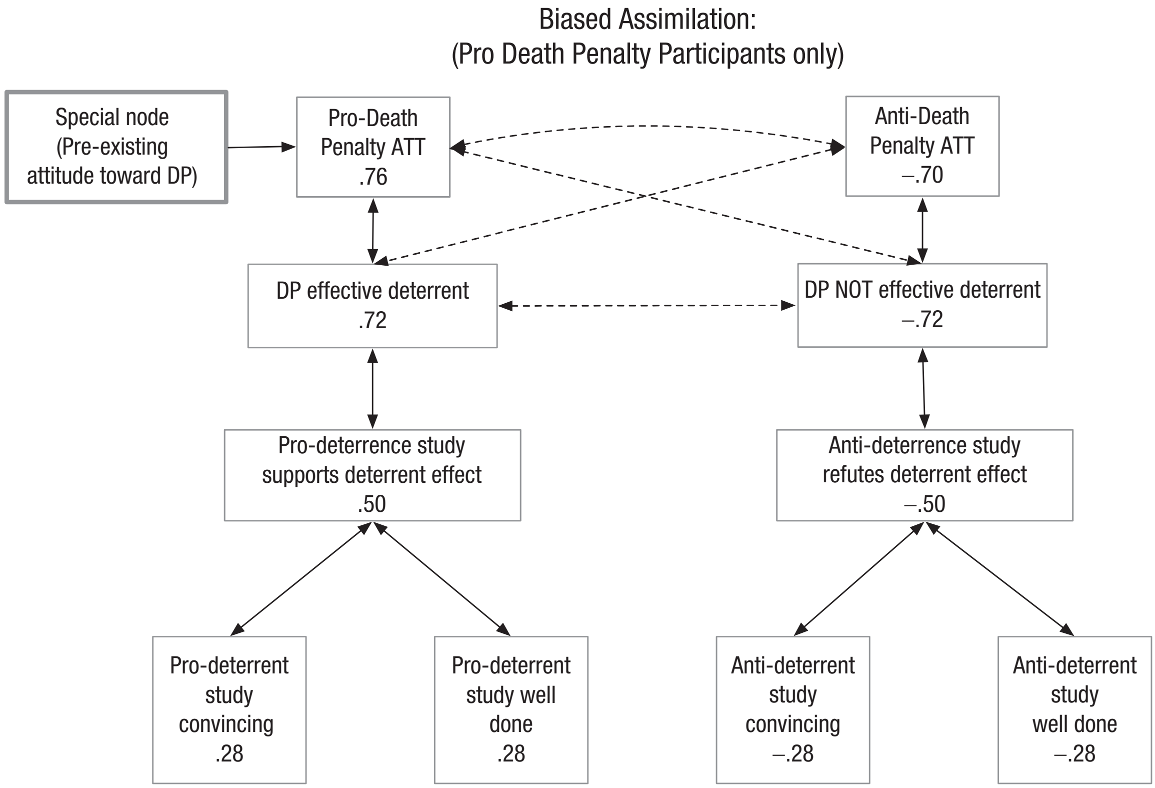

Fig. 5. ECHO (Thagard, 1989) simulation of Lord et al.'s (1979) simulation of the impact of death-penalty attitudes on biased assimilation of evidence. Numbers in the nodes represent the strength of activation of that concept after the network has settled. Note that in the original data, the bottom layer is not perfectly symmetrical in that participants with anti-death-penalty attitudes do not find antideterrence studies "well done" (or "convincing") but that they were indistinguishable from indifference and distinguishable from the judgments of the supporters of the death penalty.

Each of the nodes in Figure 5 has a starting activation of .01, except for the special-evidence node, which has a starting activation of 1 and a weight of .04. The special-evidence node indicates that the linked proposition is strongly supported or believed. Activation is initiated primarily at the special-evidence node, and a minor amount spreads also from each individual node. Activation spreads across the network iteratively, until the activations reach stability and "settle" at a state in which the constraints are maximally settled.

The decimal numbers in the nodes in Figure 5 represent the final activation of each node after the network settles. Note that all the nodes (except the special-evidence node) start with a small positive activation (.01), but as the network reaches coherence, the nodes tend to move toward either strong positive or negative values based on the pattern of excitatory and inhibitory connections among them. Positive values indicate that the individual believes the respective proposition, whereas negative values indicate that the individual does not believe it.

Note that the simulation produces a coherent network. The left side of Figure 5 shows that the model predicts pro-death-penalty individuals will view the death penalty as an effective deterrent, strongly endorse the study that purports to back that proposition, and deem the study to be convincing and well conducted. By the same token, the right side of Figure 5 shows disagreement with the proposition that the death penalty is ineffective, disbelief that the antideterrence study casts doubt over the practice's effectiveness, and evaluations of the study as unconvincing and poorly done. The similarity between the model and the original behavioral findings is compelling. We note, in particular, that coherence spread vertically, traversing all four layers of the network, which is a testament to the depth of the effect.[1]

In sum, the behavioral data reported by Lord et al. (1979) are generally both consistent with the coherence effect and successfully predicted by the constraint-satisfaction model. But recall that coherence and normativity are orthogonal. From a normative point of view, incoming evidence should be evaluated objectively for its inferential value and thus have the potential to alter the conclusion. Yet as this study demonstrated, incoming evidence is assimilated into the preexisting conclusion. This constitutes a blatant violation of the norm of directionality, which instructs that the reasoning process should proceed exclusively from attributes and premises to conclusions and not the other way around.

For another example of the confirmation bias and the backward reasoning that lies at its core, we turn to a study by Mahoney (1977). Mahoney used the naturalistic peer-review process of a social-psychological journal by soliciting reviews of an article from referees. The reviewers were asked to offer a global recommendation regarding the article's publication worthiness, and they were asked also to evaluate it on each of the criteria on which those recommendations were supposed to be based: the soundness of the methodology, the quality of data presentation, and the article's scientific contribution. A normative-reasoning process requires, first, that the reviewers' judgments should not normally be influenced by the referees' preconception of the study's findings. Second, the overall recommendation should be based on the judgments of the underlying decision criteria. Unbeknownst to the reviewers, this article was constructed and circulated only for the purposes of the study, and it was sent to reviewers whose positions on the article's topic were known in advance. As hypothesized, the study revealed that the referees' prior beliefs affected their judgments: The modal recommendation by referees whose preexisting beliefs were buttressed by the results was to accept the article with moderate revisions, whereas most of the referees who disagreed with the results recommended to reject it or require a major revision. Second, confirmatory reasoning affected all the judgments of the decision criteria, aligning them with the overall recommendation, which was aligned, in turn, with the referees' preexisting beliefs. In effect, the study demonstrates a double normative violation: The reviewers' preconceived beliefs swayed their overall recommendation and also seeped backward to contaminate their judgments of the underlying criteria. In a classic exemplar of circular reasoning, the entire task was distorted by the very beliefs whose empirical basis was being put to the test. The fact that this normative violation pervaded judgments made by scholarly experts in the pursuit of a scientific (psychological) endeavor is as disconcerting as it is illustrative of the strength of the bias.

Confirmation bias has been observed also in the domains of scientific research (Greenhoot et al., 2004), medical decision-making (Graber et al., 2005; Kohn et al., 2000; Ludolph & Schulz, 2018; O'Sullivan & Schofield, 2018; Wallsten, 1981), psychiatric diagnoses (Langer & Abelson, 1974), clinical psychology (Ben-Shakhar et al., 1998; Zapf et al., 2018), and other contexts. Researchers have also identified the reciprocal "disconfirmation bias," by which evidence that is incompatible with one's beliefs is dismissed or perceived as too weak to affect the conclusion (Edwards & Smith, 1996; Taber et al., 2009).

We propose that the confirmation bias is best understood as driven by coherence-based reasoning. Incoming evidence that comports with the preconceived conclusion generates a state of coherence, thus smoothly resulting in a correspondent conclusion. In contrast, a poor fit between the two triggers incoherence, which means that a coherent conclusion will need to be constructed. The ensuing instability could be resolved in a normative manner by way of revising the preconceived belief in light of the incoming evidence. Alternatively, coherence can be attained by altering the evidence to comport with the preconceived beliefs, as observed in the studies by Lord et al. (1979) and Mahoney (1977). For illustration, in the latter study, sound methods that lead to a preferred conclusion create a state of coherence that should easily bring a referee to recommend the article for publication. By the same token, unsound methods and a disfavored conclusion are coherent, and together, they enable a solid recommendation to reject. But the combination of solid scientific work that leads to a disfavored conclusion (and unsound scientific work that leads to a favored conclusion) generates a state of incoherence, which will throw the network into instability. The coherence-maximizing process takes the form of a competitive activation between the two attractors: the preconceived favored conclusion and the conclusion that emerges from the evidence. The network will cross-activate until one attractor overpowers its rival. When the node representing the normative judgment—that is, the scientific merit of the research—prevails, the referee will have performed the review task appropriately. However, when the referees' scientific judgment is dominated by their favored conclusion, they will have engaged in biased reasoning.

Halo effect. Introduced by Edward Thorndike (1920), the "halo effect" stands for the "tendency to think of the person in general as rather good or rather inferior and to color the judgments of the qualities by this general feeling" (p. 25). Thorndike examined how U.S. military officers rated the air cadets under their supervision during the First World War. The officers were asked to provide ratings on a number of attributes and were admonished to judge each trait independently of the other traits. Thorndike's key finding was that the correlations among the ratings were too high and too similar in magnitude to be true (Thorndike, 1920). This observation was developed further in work by Solomon Asch (1946a), who showed that "The trait develops its full content and weight only when it finds its place within the whole impression" (p. 284), which shapes "into a single, consistent view" (p. 261). Adding the trait "warm" to a set of traits not only contributes an additional facet but also alters all the other traits, resulting in the target person seeming more generous, wise, and imaginative. This assimilative effect also works in the opposite direction. For example, the trait "calm" is deemed synonymous with "serene" when it follows favorable traits, such as kind, wise, and honest, but is seen as cold, icy, calculating, and scheming when it is follows negative ones, such as cruel, shrewd, and unscrupulous (Asch, 1946a). In a study by Nisbett and Wilson (1977), a teacher addressed his students either in a cold and unfriendly manner or in a warm and friendly manner. Unsurprisingly, the students were less amenable to the former than to the latter. The notable finding was that the students also judged the former to be less physically attractive, more unpleasant, and his (French) accent to be more irritating. Note that the participants denied being influenced by the teacher's likeableness, and displaying a lack of introspection, they insisted that the likeableness judgments were brought about by his accent, mannerisms, and physical unattractiveness (all of which were kept constant; Nisbett & Wilson, 1977; Wilson & Brekke, 1994). In other words, participants engaged in backward reasoning and thus violated the norm of directionality while claiming to have adhered to it.

The overlap between the halo effect and coherence-based reasoning is inescapable. The halo effect is a classic instantiation of the coherence effect in that the perceived traits are altered by one another as part of the construction of a coherent impression of the person. The overlap is unsurprising given that both phenomena emanate from the same Gestaltian sources (see also Heider, 1958; Ichheiser, 1949; Stagner, 1951).

Stereotype spillovers. Essential to adaptive social functioning is the ability to judge other people correctly. Given people's heavy reliance on categorization as a fundamental cognitive tool (Medin & Heit, 1999; Rosch, 1978), they habitually categorize other people into groups and extract information from their group membership (S. T. Fiske, 1998; Judd & Park, 1993; Secord, 1959). This is a useful strategy because first, group traits often shed light on the individual member. To a certain extent, bookish people are prone to be introverts, accountants tend to be pedantic, and engineers are generally precise. Second, in many instances, the individuating information about the target person is limited. It follows that when both group-based and individuating information is available, the social judgment will likely be based on an integration of those two. Kunda and Thagard (1996) offered a series of computational simulations of this integration task.

A number of studies have presented participants with the task of estimating heights of men and women using photos of the targets in a seated position. Generally speaking, participants were found to follow the gender cue correctly by assigning taller heights to men than to women. Yet they readily got tripped up by the gender cue and relied on it when it was unhelpful and even misleading. Participants persisted in estimating greater male height after being told that the men and women targets were drawn from equal height samples and were admonished not to rely on it (Nelson et al., 1990; Sá et al., 1999), even when they were incentivized monetarily to avoid doing so (Dorrough et al., 2017). In other words, people tend to rely on stereotypes even when they have no diagnostic value, hence the term "spillover."

The misuse of stereotypes becomes particularly problematic when they are fueled by prejudice or animus toward marginalized groups (Devine, 1989; Dovidio et al., 1986; Eberhardt, 2020). A classic study by Duncan (1976) showed that a stereotypical view of an African American man can result in judgments of aggression when interpreting an ambiguous physical interaction between a White man and a Black man: White participants interpreted the same behavior as a jovial shove when the actor was White but as a violent act when the actor was Black. In a study by Darley and Gross (1983), judgments of a schoolgirl were found to be influenced by information about her social-economic background. After viewing a video depicting mixed performance on a school test, participants judged the girl's academic performance less favorably when told that she comes from a low socioeconomic background compared with a privileged background. Consistent with connectionist activation, that impression spread throughout the task and swept with it judgments of her traits and character, inferring worse work habits, weaker cognitive skills, lower motivation, and lesser sociability and maturity.

These findings are consistent with the automatic activation of constraint satisfaction in that stereotypes create powerful and habitual attractors unconsciously, and those attractors could sway the network even when the information they provide is irrelevant, wrong, or non-normative. The activation of negative stereotypes such as "Blackness implies aggression" and "poverty implies incompetence" does not readily comport with individuating behavior that is neutral or ambiguous, not to mention favorable. To attain coherence, the reasoner needs to either suppress the stereotype or alter the interpretation of the target's behavior. Given that stereotypes tend to be deeply entrenched (Allport et al., 1954; Banaji et al., 2001), the more likely route to attaining coherence will be via morphing the individuating information toward a stereotype-conforming interpretation. These skewed judgments, in turn, serve to reinforce the stereotypes that drove them in the first place. Herein lies the pernicious circularity of stereotypical reasoning afforded by coherence-based reasoning.

Moreover, derogatory beliefs borne by prejudicial stereotypes are often coupled with negative affect felt toward the target group. The aforementioned enmeshing of cognitive and emotional elements makes stereotypes especially powerful and resilient.

Hindsight bias. In many walks of life, the prediction of future events can be a matter of importance, such as when deciding to start a new business or when a country contemplates going to war (Hastie & Dawes, 2009; Vlek, 1984). Often, people seek to evaluate the quality of those predictions. Such inquiries will frequently be done ex post, after the predicted event has already materialized, or failed to materialize. Normatively speaking, the evaluation of a prediction should be done based on the conditions and information that existed at the time the prediction was made, and all subsequent events should be disregarded. However, research on hindsight bias has shown that retrospective judgments of (putatively) ex ante predictions are readily susceptible to intrusion by ex post information. It follows that predictions are judged with the benefit of information that was not known to the predictor (for reviews, see Guilbault et al., 2004; Hawkins & Hastie, 1990; Roese, 2004).

In a classic study by Baruch Fischhoff (1975), participants were given accounts of historical events that included the outcomes of those events. They were then asked to provide ex ante estimates of the probability of those outcomes. For example, participants were informed of the 19th-century war between the British and the Gurkas (in today's Afghanistan) and were given outcome information (i.e., the British won or the Gurkas won). Participants were then asked to estimate the ex ante foreseeability of that outcome. The study found that the estimates were affected by the knowledge of the outcome. Participants who were told of a British victory were more likely to endorse that outcome as foreseeable and vice versa for participants who were told of a Gurka victory. Consistent with the Mahoney (1977) study, the biasing impact of the outcome information spilled over into judgments of all other aspects of the events. For example, learning of a British victory led participants to assign greater relevance to factors such as "The Gurkas were only some 12,000 strong," whereas participants who were told of a British defeat assigned greater relevance to factors such as the Gurkas "were brave fighters, fighting in territory well-suited to their raiding tactics" (Fischhoff, 1975). The retrospective knowledge also affected participants' memory of the event, creating a sense of "I knew it all along" (Fischhoff, 1975, 1977; Powell, 1988). The effects of the hindsight bias were found to be resilient to educating participants of the bias and to explicitly instructing them to ignore the outcome information (Fischhoff, 1977).

We propose that the hindsight bias can be understood as an instantiation of coherence-based reasoning. Consistent with schematic knowledge, people tend to view outcomes as flowing from their potentiating environment. Mismatches between actual outcomes and predicted outcomes will be incoherent and thus unstable. Given that the outcome is dictated by the experimental materials and thus taken by the participants as given, the only route to attaining coherence is by altering its predictability. Consistent with coherence-based reasoning, the biasing effect spreads to all corners of the representation of the task, affecting all judgments small and large, and even seeps into the memory system (for evidence of coherence affecting background knowledge, see D. Simon, Snow, & Read, 2004).

Note that Fischhoff (1975) offered a theoretical explanation for the hindsight bias that is a kindred spirit of coherence-based reasoning. He posited a process of "creeping determinism" by which the perceived inevitability "is imposed upon, rather than legitimately inferred from, the available evidence" (Fischhoff, 1975, p. 293). Participants assimilate the hindsight information with the rest of the knowledge as a way to "make sense, or a coherent whole, out of all that he knows about the event" (Fischhoff, 1977, p. 356; see also Roese & Vohs, 2012). Again, consistent with coherence-based reasoning, Fischhoff (1977) noted that the hindsight bias is difficult to undo or resist because the intrusion of ex post information occurs naturally, immediately, and absent conscious awareness.

Outcome bias. The outcome bias is similar to the hindsight bias in that it, too, concerns the susceptibility of judgment to knowledge that becomes known only after the fact. In a classic demonstration, Mark Alicke and colleagues (1994) found that participants gave favorable ratings to a psychiatrist's decision to release a mental patient when the patient was said to turn his life around, but the same decision was rated negatively—with overtones of blame—when the patient eventually suffered a breakdown and killed his landlord (see also Alicke & Davis, 1989). Likewise, Baron and Hershey (1988) found that decisions are unduly judged by the way they turned out even when the outcome was unpredictable to the decision maker. This finding was obtained even when the favorable outcome followed an inferior decision and the unfavorable outcome followed a superior decision (Baron & Hershey, 1988; see also Stanovich & West, 2008).

We propose that the outcome bias is yet another instantiation of coherence-based reasoning. Again, based on schematic knowledge, people tend to perceive congruence between the quality of decisions and the favorability of their outcomes. Thus, a network representing a mismatch between the two will be incoherent and thus unstable. Given the relative immutability of the given outcome, the easiest route to attain coherence is to skew the evaluation of the decision and thus align it with the outcome.

Belief bias. Unlike most other forms of biased reasoning, one strand of research—dubbed "belief bias" (Evans & Curtis-Holmes, 2005; Markovits & Nantel, 1989; Sá et al., 1999)—has focused on deductive reasoning, specifically, syllogistic logic. The rule of modus ponens mandates that the pair of premises "If P, then Q" and "P is true" should lead one to conclude, "Therefore, Q." A central feature of syllogistic reasoning is that it should be driven by the internal logic of the task regardless of the content or context of the premises (Chater & Oaksford, 2012; Evans et al., 1983; Piaget, 1972).

Nonetheless, a century's worth of research has revealed that syllogistic reasoning is frequently skewed by beliefs that are exogeneous to the task, typically, the believability of the conclusion (Evans et al., 1983; Morgan & Morton, 1944; Woodworth & Sells, 1935). Studies have shown that deductive conclusions that are both valid and believable are readily accepted, just as conclusions that are both invalid and unbelievable are summarily rejected. Both scenarios are internally coherent, and they have resulted in accuracy rates that exceed 90% (Evans et al., 1983). This level of performance decreases considerably when the task precipitates a state of incoherence, that is, a case in which a syllogistic conclusion is internally valid but not believable (e.g., Premise 1: All mammals walk. Premise 2: Whales are mammals. Conclusion: Whales walk.) or the conclusion is believable but logically invalid (e.g., Premise 1: All living things need water. Premise 2: Roses need water. Conclusion: Roses are living things.). Typically, performance levels on such tasks drop to 50%–70% (Evans et al., 1983; see also Markovits & Nantel, 1989; Sá et al., 1999).

States of incoherence will be resolved only when one of the vying attractors—validity and believability—manages to overpower the other. When the former prevails, the reasoning process will have complied with the norms of reasoning. But when the believability attractor carries the day, the reasoner will have reached a biased result. These findings have obtained even when participants were explicitly admonished to accept the premises as true, confine their responses to the logical conclusion that follows from those premises, and refrain from using information that is extraneous to them (Evans et al., 1983; Markovits & Nantel, 1989; Sá et al., 1999).

Naive realism. To make sense of the world, people need first to interpret it (Asch, 1952; Ichheiser, 1943; Krech & Crutchfield, 1948). This truism, however, is generally unnoticed by our metacognitive system. "Naive realism" (L. Ross & Ward, 1996) stands for the observation that people tend to equate their perceptions and judgments with objective reality, failing to realize that those perceptions are unconsciously framed and interpreted through their idiosyncratic and situationally specific lens. Given this perceived access to objective truth, people tend to believe that others will share the same impression and that divergent perceptions stem from being uninformed, unreasonable, or biased (Pronin et al., 2004; Pronin & Kugler, 2007; L. Ross & Ward, 1996).

Naive realism and the ensuing "bias blind spot" (Pronin et al., 2002) have traditionally been explained as deriving from the discrepancy between the richness of self-introspection and the paucity of introspection into other people's minds (Molouki & Pronin, 2015; Pronin, 2008; Pronin & Kugler, 2007). We suggest that this systemic bias can be deemed also as a product of coherence-based reasoning in that the sense of objective reality stems from the distortion of the entire task environment to cohere with the conclusion. This state of global coherence generates a sense of inevitable correctness and unmediated realism, which makes any deviation seem wrong.

Hot-cognition coherence

The biases discussed heretofore have involved reaching conclusions from the features of the task, which are effectively determined by nature and thus external to the reasoner. In contrast, biases triggered by hot cognitions—such as motivation and emotion—are typically internal to the reasoner and thus exogenous to the attributes of the task (Baumeister & Newman, 1994; Kunda, 1990). Thus, any impact of internally generated hot cognitions on the reasoning task is bound to be illogical: The odds of winning a lottery should be unaffected by a gambler's desire to win, and a judgment of a person should not hinge on the observer's emotional state (cf. Bodenhausen, 1993; Loewenstein & Lerner, 2003). Given this logical irrelevance, judgments influenced by hot cognitions will frequently constitute violations of normative reasoning.

We reiterate that often, cold and hot cognitions are both implicated in reasoning tasks (Abelson, 1963; Kunda, 1990; D. Simon et al., 2015, 2020; Thagard, 2006), as they are across the human experience (Lepper, 1994; Sorrentino & Higgins, 1986). The ordering of topics around this distinction is intended for presentational convenience alone.

Motivated reasoning. Perhaps the strongest and most ubiquitous contribution of hot cognitions to biased reasoning stems from human motivation. Motivation is a foundational construct in that it guides the attainment of goals and thus enables the all-important satisfaction of needs (Baumeister, 2016; Dweck, 2017; Maslow, 1943). Invariably, the meeting of goals and satisfaction of needs are accompanied by positive valence, such as the pleasant affect one experiences when winning a competition, receiving a gift, or feeling loved (Roese & Olson, 2007; Suri & Gross, 2012; Weiner, 1985; Zajonc, 1980). As proposed by Kunda (1990), reasoning tasks can be driven by the goal of reaching accurate conclusions, which typically promotes normative-reasoning processes. Frequently, however, reasoning pertains to a motivational goal, that is, a "wish, desire, or preference that concerns the outcome of a given reasoning task" (Kunda, 1990, p. 480). A large body of research on motivated reasoning shows that directional goals tend to hijack the reasoning task, resulting in distortions of the relevant evidence (Dunning, 2015; Erisen et al., 2014; Kunda, 1990; D. Simon et al., 2015). Motivated reasoning comes in the form of both unwarranted credulity of favorable information and unwarranted skepticism toward unfavorable information (Clark & Winegard, 2020; Ditto & Lopez, 1992; Taber & Lodge, 2006). The next four sections examine familiar reasoning processes that are understood to be skewed by motivation.

We argue that these types of biased reasoning can all be explained by coherence-based reasoning in that the motivations are represented in the networks as attractors, which generate global forces that skew the network to cohere with them. The ensuing conclusions then form around those motivations, thereby enhancing positive affect and averting negative states, often in violation of the norms of reasoning.

Desirability bias. The desirability bias—also known as unrealistic optimism (Weinstein, 1980) and the Pollyanna principle (Matlin & Stang, 1978)—captures the notion that people tend to overpredict positive events and downplay the likelihood of negative ones. For example, studies have found that participants tend to inflate the odds that a card drawn from a deck will confer monetary gains rather than monetary losses (Irwin, 1953; Irwin & Metzger, 1966), that their chosen horse will win the race (Brownstein et al., 2004), that their sports team will win an upcoming match (Massey et al., 2011), that they will succeed in a task that earns monetary payoffs (Windschitl et al., 2013), and that their businesses will succeed (Boiney et al., 1997; A. C. Cooper et al., 1988; Larwood & Whittaker, 1977). The desirability bias can also alter the way in which people judge other people and situations. For example, participants were found to exaggerate the skills of their prospective partners in an upcoming history-trivia game (Klein & Kunda, 1989).

Motivated reasoning can even alter the desirability of the outcome itself (McGuire, 1960). For example, participants rated a prospective dating partner as more desirable when the probability of the date materializing increased (Wilson et al., 2004). Altering the desirability of an outcome is particularly useful when the desired outcome becomes unattainable. Consistent with the "sour grapes" phenomenon (Elster, 1983), studies have shown that desirable outcomes, such as a victory by one's preferred political candidate or a tuition decrease, become less desirable as they appear less likely to materialize (Kay et al., 2002), and a prize becomes less attractive as it seems less attainable (Pyszczynski, 1982). Altering the desirability of an outcome can also prove useful in the opposite situation of "sweet lemons," such as an electoral victory by an opposing candidate becomes more palatable as its likelihood increases (Kay et al., 2002; Wilson et al., 2004).

Self-concept maintenance. As stated by Gordon Allport (1937), maintaining one's self-concept is "nature's eldest law," which makes the self-concept fertile ground for the experience of both positive and negative valence. The need for a positive sense of self is manifested in people's needs for achievement (Atkinson & Raynor, 1974), status (Anderson et al., 2015), social approval (Marlow & Crowne, 1961), social acceptance (Baumeister & Leary, 2017), and control (Heine et al., 2006; Pittman & D'Agostino, 1989). Maintaining the self-concept is manifested by the complementary constructs of self-enhancement and self-protection (Alicke & Sedikides, 2009).

Self-enhancement refers to people's tendency to portray themselves in a favorable light. People tend to describe themselves favorably across a slew of traits, such as fairness (McPherson Frantz, 2006), friendliness (Fetchenhauer & Dunning, 2006), honesty (Ehrlinger et al., 2005), morality (Epley & Dunning, 2000; Miller & Ratner, 1998), and ironically, objectivity and freedom from bias (Ehrlinger et al., 2005; Pronin et al., 2002, 2004; D. Simon et al., 2020; Uhlmann & Cohen, 2007). People tend to rate themselves as being above average across a wide range of domains (Cross, 1977; Dunning et al., 2004; Svenson et al., 1985) and inflate outside evaluations of their performance and rate their own performance even higher (Alicke, 1985). Typically, expressions of self-enhancement are accompanied by boosts in positive affect (McFarland & Ross, 1982; Roese & Olson, 2007; Wheeler & Miyake, 1992).

Probably the most ubiquitous and fiercest motivations are evoked in response to threats to the self-concept. Echoing Freud's concept of defense mechanisms, experimental studies have shown that people react to self-threats by adopting a suite of strategies, including reaction formation, isolation, and denial (Baumeister et al., 1998). People tend to dismiss putative personality tests that are said to detect a negative trait by deeming them irrelevant and poorly conducted but to herald the same tests when they are said to detect positive traits (Pyszczynski et al., 1985; Tesser & Paulhus, 1983). When faced with a medical test that indicates an unfavorable diagnosis, people tend to downplay the seriousness of the condition (Ditto et al., 1988) and question the validity of the test (Ditto & Lopez, 1992). College students believe that they are less likely than their peers to develop a drinking problem or suffer a heart attack (Weinstein, 1980). Motorcyclists believe they are less likely than other bikers to get into accidents (Rutter et al., 1998), and bungee jumpers believe they are less likely to get injured than fellow jumpers (Middleton et al., 1996). As demonstrated by Festinger and Carlsmith (1959), rather than reckon with their dishonesty, people will distort their own judgments to maintain a positive self-image.

In-group bias. Intergroup conflict is another domain that is rife with biased reasoning. Studies have revealed a host of both in- and out-group differences across a range of judgments pertaining to competence (Hinkle & Schopler, 1986; Sachdev & Bourhis, 1987; Sherif et al., 1961), honesty and cooperation (Gaertner et al., 1990), attributions (Stewart et al., 1985), and traits (Rosenbaum, 1986). These findings have been observed across a variety of affiliations, including minimal groups (Tajfel et al., 1971), political membership (Munro et al., 2010), and national conflicts (Halperin, 2008; Halperin & Bar-Tal, 2011). The classic study by Hastorf and Cantril (1954) demonstrated that belonging to a college community resulted in polarized distortions of a contested football game: which activities on the field amounted to rule breaking, how severe were those violations, who was to blame for the fracas, and how genuine were the ensuing complaints. Findings of in-group bias are consistent with the view of group membership as an extension of one's selfhood (Sherman & Kim, 2005; Tajfel & Turner, 1979).

Role-induced bias. In a seminal study, Janis and King (1954) assigned participants (random) to argue for a particular position in a debate format. The study found that participants came to endorse and internalize that (randomly assigned) position. This finding has been labeled the role-induced bias, and it has replicated in both experimental settings (Egli Anthonioz et al., 2019; Engel & Glöckner, 2013; Melnikoff & Strohminger, 2020; D. Simon et al., 2020) and naturalistic settings among debaters (Eigen & Listokin, 2012; Schwardmann et al., 2022) and legal expert witnesses (Murrie & Boccaccini, 2015; Murrie et al., 2013). The role-induced bias persists even in the face of monetary incentives to provide unbiased judgments (Schwardmann et al., 2022), even with incentives as high as €100 (Engel & Glöckner, 2013). The bias has been observed even with minimalistic experimental treatment, that is, by merely asking participants to imagine their role (D. Simon et al., 2020). Studies show that the role assignment not only affects the overall preference for the assigned position, but it also sways the perception of the entire task, including all the corresponding evidence and propositions (Engel & Glöckner, 2013; Melnikoff & Strohminger, 2020), and it triggers a motivation to see the respective side prevail (D. Simon et al., 2020).

To recap, we argue that the four types of motivation exert their biasing effects in a similar manner. These motivations—whether a desired state of affairs, a positive self-image, a positive sense of one's group, or identification with one's role—all generate strong attractors that can dominate the coherence-maximization process and result in conclusions that affirm those motivations.

Emotion-driven reasoning. It is widely accepted that emotions serve important intrapersonal and social functions and that they are intimately linked to a person's cognitive appraisal of the environment (see e.g., Ellsworth & Scherer, 2003; Frijda, 1986; Lazarus, 1966; Smith & Ellsworth, 1985). Yet emotions also have the capacity to distort human judgment. Note that anger has been shown to affect judgments of other people's conduct, resulting in inferring more personal rather than situational causal attributions (Keltner et al., 1993; Quigley & Tedeschi, 1996), more severe judgments of responsibility (Bodenhausen et al., 1994; Goldberg et al., 1999; Lerner et al., 1998), higher imputed intentionality (Ask & Pina, 2011), greater discounting of alternative explanations and mitigating circumstances (Georges et al., 2013), and resorting to lower thresholds of evidence for apportioning blame (Tetlock, 2002). In all, these tendencies spur a mindset of blame validation (Alicke, 2000) and punitiveness (Goldberg et al., 1999; Nuñez et al., 2015). These effects have been observed even when the arousal of anger was incidental to the target judgment, that is, when the target person played no role in the arousal of the anger (Bodenhausen, 1993; Loewenstein & Lerner, 2003).

These reactions to anger can be explained by coherence-based reasoning. One of the distinct characteristics of anger is that it triggers an approach mindset (Carver, 2006) that induces people to confront and rectify the situation (Carver & Harmon-Jones, 2009; Fischer & Roseman, 2007; Frijda, 1986), including the infliction of pain and harm on the offending other (Shaver et al., 1987). But to maintain coherence with one's self-image as a moral and reasonable person, this vigilance must be accompanied by a belief that the target person does indeed deserve the punishment. Hence, the cognitive process distorts the situation to reach heightened attributions of blameworthiness by assigning personal responsibility, imputing intentionality, discounting alternative explanations, and the like.

Similar effects have been found with feelings of sympathy. As mentioned above, D. Simon et al. (2015) found that judgments of an allegation of student misconduct were influenced by the knowledge that the suspected student's brother had been killed by a drunk driver some months prior. This knowledge lowered findings of guilt and swayed the entire representation of the task toward more lenient interpretations of the evidence. From a normative point of view, feelings of sympathy should not influence factual judgments of an unrelated incident that occurred months later. These distortions can be viewed as driven by coherence-based reasoning in that feeling sympathy toward the student impels one to rule in her favor, but doing so requires sustaining a belief that she engaged in no wrongdoing. Coherence is thus restored by altering the evaluation of the evidence to make her seem more innocent.

Attitude-driven reasoning. A large body of research has shown that people's reasoning can be skewed by their attitudes. Given the close relationship between attitudes and beliefs—which in itself is an instantiation of a coherence effect—there is bound to be an overlap between attitude-driven reasoning and confirmation bias. As shown above, in the study by Lord et al. (1979), participants' attitudes about the death penalty's deterrent effect were coupled with concordant attitudes toward the practice. Recall that the study found strong distortions of incoming evidence, resulting in the strengthening of both the attitudes toward the death penalty and the beliefs in its deterrent effect (Lord et al., 1979). Attitudes have likewise been shown to distort judgments in the context of media bias (Matheson & Dursun, 2001; Vallone et al., 1985), affirmative action (Taber & Lodge, 2006), gun control (Taber & Lodge, 2006), legalization of marijuana (Taber et al., 2009), historical analysis (Hulsizer et al., 2004), climate science (Kahan et al., 2012), and other scientific information (Hornsey et al., 2020; Munro, 2010). Note that the presentation of balanced evidence supporting both sides of an issue fails to temper preexisting attitudes and under certain conditions, can even result in heightened polarization (Lord et al., 1979; Taber et al., 2009).

Attitude-borne biases can be explained as driven by the need to interpret evidence to cohere with those attitudes. Indeed, the multifaceted nature of attitudes can be said to compound the biasing effect of coherence observed in the domains of confirmation bias, motivated reasoning, and emotional reasoning (Hornsey et al., 2020; Kunda, 1990; Osgood et al., 1957; Taber & Lodge, 2006).

Political reasoning and partisan worldviews. A distinctive feature of contemporary politics (especially in the United States) is the power of partisanship in shaping people's views on matters of public policy. A series of studies has shown that the mere mention of a party's endorsement of a policy leads people to adopt positions that they might otherwise reject (Cohen, 2003; Guilbeault et al., 2018; Smith et al., 2012). Studies show that party affiliation leads people to judge newspaper accounts as more flattering to their preferred candidate (Meffert et al., 2006), believe that their candidate outperformed their rival in a debate (Munro et al., 2002; Richardson et al., 2008), and perceive the media as being hostile to their candidate (Vallone et al., 1985). Partisanship explained the polarized reactions to Bill Clinton's adulterous conduct in the Lewinsky scandal (Fischle, 2000). The results of the 2020 U.S. presidential election were trusted by 95% of Democrats but a mere 24% of Republicans (Montaro, 2020), whereas George W. Bush's victory in 2004 was distrusted by Democrats and praised by Republicans (Stolberg & Dao, 2005). We propose that these behaviors are facilitated by coherence-based reasoning (see Thagard, 2019). Party identification serves as an attractor that alters people's values, identity, group membership, and the like.

More ominously, coherence serves as an enabler of the fermentation of ardent ideology that is increasingly engrossing people's belief systems (Brandt & Crawford, 2020). Worldviews of this kind tend to be grounded in a powerful ideological epistemology (Clark & Winegard, 2020), are rife with sacred values (A. P. Fiske & Tetlock, 1997), and are largely impervious to external reality checks (Zmigrod et al., 2019). These worldviews are undergirded by concordant interpretations of the facts, a core feature of coherence-based reasoning. These factual foundations can be sufficiently powerful to defy rational and scientific findings, resulting in positions such as distrust of vaccinations and climate-change denial (Hornsey et al., 2020; Kahan et al., 2012; Lewandowsky et al., 2013). Consistent with the coherence effect, evidence to the contrary is readily dismissed as fake news (Allcott & Gentzkow, 2017). Such an epistemology has fueled a rejection of scientific expertise with respect to the COVID-19 pandemic, thus leading to distorted risk perceptions and the flouting of sound public-health advice (Allcott et al., 2020).

The coherence framework also suggests that in the absence of supportive facts, ideological believers might turn to concocting facts, often in the form of conspiracy theories. Thus, people are told that the U.S. government was complicit in the 9/11 attack on the World Trade Center (Swami et al., 2010), Princess Diana was murdered (Douglas & Sutton, 2008), COVID-19 was a plot to enable Bill Gates to plant microchips in people's bodies, and the 2020 presidential election was actually won by President Trump but stolen by Joe Biden (Haberman & Schmidt, 2020).

Moreover, the coherence effect can cascade and lead people to far-flung and possibly grave places. For various reasons (notably, the desire to ride a strong economy into the 2020 presidential election), former President Trump sought to belittle the impact of the pandemic on daily life. This motivation led to a series of behaviors, including downplaying the pandemic's threat to public health, predicting its prompt disappearance, repudiating the calls to wear masks, holding demonstrably mask-less campaign events and social gatherings, promoting untested cures, criticizing local governments for imposing closures, championing renegade scientists, and undercutting public-health experts serving on his COVID-19 task force (Abutaleb et al., 2020). We suggest that a conceivable explanation for this confluence of behaviors is that they all cohere with the underlying motivation. Seeking to maintain their support of the president, his followers defied closures, propagated conspiracy theories, launched antimask protests, vandalized state legislature houses, plotted to kidnap the governor of Michigan, and issued death threats to public-health scientists and officials (Ames, 2020; Bella, 2020; Cohen, 2003; Withycombe & Barreda, 2020). The cascading beliefs can readily generate antagonism toward people who challenge the group's gospel. Although the beliefs and motivations underlying such actions are undoubtedly multidetermined, we maintain that a cascade of coherence effects plays a central role in guiding these cavalcades of otherwise inexplicable behaviors.

Summary and Implications

In the first half of the article, we built up the case to support the first objective of the article, that is, that coherence-based reasoning serves as a vital and pervasive tool of human cognition. Reasoning is understood to be represented in connectionist networks of interconnected nodes and processed through constraint-satisfaction mechanisms. These processes are driven by structural forces that transform complex representations into states of equilibrium, or coherence, in which the winning conclusion is supported by highly activated attributes and the rejected conclusion and its supporting attributes end up with low activation. Coherence is attained through a coherence-maximizing process that alters the nodes—whether by way of strengthening, weakening, or morphing them—to bring them into line with the emerging conclusion. It must be emphasized that connectionist networks are fundamental to the functioning of the brain and that coherence-based reasoning serves as a ubiquitous, essential, and overwhelmingly adaptive apparatus in people's mental toolbox.

The evidence supporting this conception is, in our view, compelling. As mentioned above, a number of research groups, ours included, have found a slew of findings that match the predictions from a constraint-satisfaction process. These include systematic changes in the levels of activation of the task attributes; high intercorrelations among those attributes; cross-activation between attributes that are not logically related; bidirectional activations among and within entire sets of attributes; high levels of confidence (despite the comparable strength of the vying conclusions); a relationship between the magnitude of coherence shifts and the ensuing confidence; impact of extraneous variables on all other attributes; spreading of coherence from one task to an unrelated task; multidirectional impacts among motivation, emotion, and attribute strength; seeping of coherence down to lower layers of processing; and limited awareness of the coherence shifts. To reiterate, these findings and others have been replicated across a wide range of stimuli testing different tasks and performed in multiple languages by multiple labs in the United States and abroad.

In the second half of the article, we addressed the article's second objective, namely, that coherence-based reasoning can provide a framework that integrates and illuminates a number of important deviations from normative forms of reasoning. Our framework suggests that the otherwise adaptive cognitive process can readily result in biased conclusions when the network is dominated by nodes or links that are incorrect, overweighted, or otherwise nonnormative. Given the exogeneity of normativity to the cognitive process, the most coherent state of the cognitive representation might or might not be consistent with normative principles. The process will be prone to bias when it contains a powerful attractor of the sort covered throughout this article: strong beliefs, stereotypes, hindsight knowledge, outcome knowledge, motivations, aroused emotions, and so on. When these attractors are stronger than the reasoner's commitment to the norms of biased reasoning, they will likely win the activation battle and result in a biased outcome. To reiterate, the resulting bias arises not from dysfunctional processing but from the normal operation of the hardware of the brain.

The remainder of this section is devoted to a variety of implications that flow from the framework. But before we proceed to discuss those implications, it is incumbent upon us to briefly touch on the framework's relationship to a recently published article by Oeberst and Imhoff (2023).

Oeberst and Imhoff's (2023) belief-consistent information processing

We feel much affinity toward Oeberst and Imhoff's (2023) framework. They, too, start with lamenting the fact that the numerous strands of bias research are both scattered and ignorant of one another, denying the field the benefits of integrated common principles. As with our own project, Oeberst and Imhoff's primary mission was to introduce a parsimonious framework to integrate and better understand biases. They propose that many biases share the same "recipe," by which incoming propositions are processed to be consistent with preexisting beliefs. The driving principle, then, is that "prior beliefs plus belief-consistent information processing" lie at the heart of biased reasoning. It is not hard to see the similarity between this operative principle and the principle of coherence that we advocate. Both frameworks also rely on similar distortions of the propositions in the service of attaining the desired states.

The two approaches, however, have some distinct differences. We first note the difference in scope. Recall that our framework sets out to encompass biases borne by cold cognitions as well as biases borne by motivations and emotions. In contrast, Oeberst and Imhoff's (2023) framework is explicitly limited only to belief-based biases, as denoted in the framework's label. That is inevitable because of the grounding of their framework in the concept of belief-consistent information processing, specifically, deeming biases as "variants of 'confirmation bias'" as propounded by Nickerson (1998). Indeed, in their article, they state repeatedly that motivation is not a necessary ingredient for bias. That proposition is, of course, correct. Indeed, some biases emanate exclusively from cold cognitions, as expounded in our article. By the same token, our article shows that numerous and important biases emanate also from motivations and emotions. Yet, these important and pervasive biases are left unaddressed by Oeberst and Imhoff. Curtailing the scope of the belief-consistent information processing framework in this manner brings it closer to the disjointed and isolated kind of literature that both approaches avowedly eschew. We also note a sense of puzzlement by Oeberst and Imhoff's insistent repudiation of the role of motivation given that motivations underlie some of the core beliefs on which their framework rests. Indeed, three of the six exemplars presented in their Table 1 emanate straight from propositions that are indisputably motivational in nature ("I am good," "My group is good," and "I make correct assessments of the world"). The scope of Oeberst and Imhoff's framework is limited also in that it applies only to dyads (the underlying beliefs and the incoming propositions), and only to propositions located in the topmost layer of reasoning.

The two frameworks diverge also with respect to their grounding in a theoretical foundation. Whereas we situate the coherence framework in a mechanistic, computational account that fits within a fully fledged model of cognition, Oeberst and Imhoff's (2023) framework offers little in the way of a cognitive theory or an explanatory cognitive mechanism. They base their framework on a review of a body of research findings that support belief-consistent processing, leading them to conclude that "belief-consistent information processing seems to be a fundamental principle in human information processing." In the Oeberst and Imhoff framework, confirmation bias serves as the explanatory template for all other biases (an explanans), whereas in our framework, it is just one of the phenomena that are products of coherence-based reasoning (an explanandum). The lack of an underlying cognitive theory begs a host of questions: Given that the paradigm speaks only of pairs of related variables, how does one explain spreading of bias across large sets of variables, remote effects by extraneous variables, or spreading between variables that have no logical connection? What factors moderate the process? Given that belief-consistent information processing cannot plausibly be expected to overcome every contrary interpretation, what are the countervailing factors, and what determines the outcome of dialectical conflicts?

In sum, we deem both frameworks to be instructive and revealing, and we feel that they coexist comfortably with each other. At the same time, we resist the notion that the insights provided by coherence-based reasoning have been exhausted by the belief-consistent information-processing framework.

Coherence and Bayesianism

Numerous researchers have argued that human cognition is best characterized in terms of Bayesian mechanisms (e.g., Griffiths et al., 2008, 2010; Sanborn & Chater, 2016), and some have argued that such mechanisms are inconsistent with the connectionist mechanisms that underlie coherence-based reasoning and coherence effects (Griffiths et al., 2010). However, we believe the two approaches are not incompatible.

First, many Bayesian modelers have acknowledged that the human brain is only approximately Bayesian. Because of limitations in time and resources, the brain cannot calculate the complete joint probability or prior probability distributions required for many kinds of reasonable problems. Sanborn and Chater (2016) argued that the brain solves that problem similarly to how Bayesian modelers do: They create approximations by using various sampling techniques, such as Markov chain Monte Carlo sampling, to iteratively build a probability distribution. However, time and resource limitations limit the sampling process and therefore the accuracy of the estimated probability distributions. In addition, characteristics of the sampling process as implemented in a biological brain might result in further limitations or biases in the sampling process.

Related to this argument, the resource-rational perspective (e.g., Bhui et al., 2021; Lieder & Griffiths, 2020) argues that rather than examine cognitive processes from an optimally rational perspective, one should focus on the idea that what the brain does with a specific task is rational given all the relevant constraints, such as time, cognitive resources, environment, evolutionary pressures, biological constraints, and so on. This resource-rational perspective has led to models whose core processes strongly resemble coherence-based reasoning (Bhatia, personal communication, December 8, 2021). For example, some resource-rational models allow for preferences and beliefs to bias search, which can be resource rational when search is expensive. But the result is an algorithm that is very close to key aspects of coherence-based reasoning, such as the importance of bidirectional influence among nodes in the network. Furthermore, it has been shown that even simple constraint-satisfaction networks generate predictions that align closely with rational solutions (e.g., Jekel et al., 2012).

Our second reason is based on Marr's (1982) classic distinction between levels of analysis: computational, algorithmic, and implementational. Bayesian models focus on the computational and algorithmic levels, the formal mechanisms by which Bayesian updating is done. In contrast, coherence-based reasoning focuses more on the implementational level, that is, the hardware (e.g., neurons and networks of neurons) in which reasoning is implemented, and on the implications of that hardware for processing. Because Bayesian models largely ignore the implementation processes done by specific hardware, they overlook the possibility that the brain's implementation of Bayesian processing may have important implications for its accuracy or lack of bias. However, when one looks at how neurons and networks of neurons operate, this specific hardware implementation has implications for how reasoning proceeds (note that according to the resource-rational perspective, these implementational issues may be an important constraint on the resource rationality of a process or mechanism). As we argued in other sections of this article, the brain behaves like an energy-minimizing system that operates as a parallel constraint-satisfaction process (Hopfield, 1982, 1984; Rumelhart et al., 1986; Smolensky, 1986). Inference can be viewed as an energy-minimizing process that seeks to maximize the coherence of a network of beliefs and inferences (Thagard, 1989, 2019).

It should also be noted that in this article, we have have applied coherence-based reasoning to explain a wide variety of biases and failures of reasoning. We know of no Bayesian analysis of reasoning biases that has the coverage of our coherence-based analysis.

Coherence and dual-process models

According to dual-process models of cognition, biased reasoning is largely explainable by the distinction between Type 1 reasoning, which is characterized as fast, intuitive, automatic, nonconscious, associative, and requiring little or no working memory, and Type 2 reasoning, which is said to capture slow, deliberative, conscious, rule-based, and working-memory-dependent processing (e.g., Evans & Stanovich, 2013; Kahneman & Frederick, 2002; Stanovich, 1999). This approach posits that cognitive errors are due mostly to the failure of Type 2 processing to override the quick and error-prone intuitions of Type 1 (Evans & Stanovich, 2013; Frederick, 2005). We question whether biased reasoning is best understood as an override failure. In fact, many of the abovementioned studies that displayed biased reasoning in the coherence framework entailed deliberate processing (Engel & Glöckner, 2013; Holyoak & Simon, 1999; D. Simon et al., 2015, 2020; D. Simon, Krawczyk, & Holyoak, 2004; D. Simon, Snow, & Read, 2004; D. Simon & Spiller, 2016). In other words, these biases seem to be a direct failure of the deliberate Type 2 process, not merely a failure to correct a misguided Type 1 process. Furthermore, the fact that constraint-satisfaction processing relies on a single system calls into question the bifurcated nature of dual-process models (see Arkes, 2016; Keren & Schul, 2009; Melnikoff & Bargh, 2018; Osman, 2004, 2013). Indeed, Glöckner and Betsch (2008) proposed a single constraint-satisfaction model that simulates both types of reasoning.

Coherence and cognitive-dissonance theory

There are obvious and important overlaps between cognitive-dissonance theory (Festinger, 1957, 1964) and coherence-based reasoning. Both approaches emanate from the structural-dynamics approach (Zajonc, 1968) and thus share the tenet that cognitive states are affected by the interrelationship among the elements involved in the task (see Shultz & Lepper, 1996). Both approaches also start with a broad conception of consonance/coherence as a generalized epistemic motivation that helps people see the world as comprehensible, predictable, and somewhat controllable (Festinger, 1957; see also Heine et al., 2006; Kruglanski, 1990; Thagard &

Verbeurgt, 1998). However, after presenting a general epistemic framework, Festinger (1957) divided up the theory into five paradigms, only two of which have garnered more than scant research attention.

The paradigm of decision-making, or choice between alternatives (Festinger, 1957, Chapters 2 and 3; see also Brehm, 1956), closely overlaps with studies of coherence-based decision-making (Holyoak & Simon, 1999; Russo et al., 2008; D. Simon, Krawczyk, & Holyoak, 2004; D. Simon & Spiller, 2016). The two approaches differ, however, with respect to dissonance's insistence that the spreading of alternatives is strictly a postdecisional phenomenon (Brehm, 1956; Festinger, 1957) that is triggered only after the making of a commitment to a decision (Brehm & Cohen, 1962; Festinger, 1964). Rather, the research on the coherence effect has shown that the spreading apart occurs for the most part predecisionally (Holyoak & Simon, 1999; D. Simon et al., 2001; D. Simon, Krawczyk, & Holyoak, 2004), a notion that is compatible with a host of findings of predecisional distortion (DeKay et al., 2012, 2014; Janis & Mann, 1977; Montgomery & Willen, 1999; Russo et al., 1998, 2008; Svenson, 1999. For reviews, see Brownstein, 2003; Russo, 2014). Coherence shifts are observed even when the processing task entails no decision-making at all (D. Simon et al., 2001, 2015, Study 1). We agree with Bruner (1957) and Abelson (1983) that to gain meaningful insight into human decision-making, one ought to focus on the processes that guide the making of decisions rather than on the retroactive rationalization of imperfect decisions (see also McGuire, 1968).

The overwhelming interest in dissonance theory has been directed at the "forced compliance" paradigm (Festinger, 1957, Chapters 4 and 5), which captures the change in people's attitudes after being lured into behaving in ways that violate their value systems (J. Cooper & Fazio, 1984; Festinger & Carlsmith, 1959). Notwithstanding the paradigm's important impact on the history of psychological thought (see J. Cooper, 2007; Harmon-Jones, 2019), we maintain that cognitive dissonance theory is best viewed as just a particular application of cognitive-consistency theories (cf. Abelson et al., 1968; Heider, 1946, 1958; McGuire, 1968; Rosenberg & Abelson, 1960). We share the views of fellow consistency theorists who lament dissonance theory's failure to meet the broader goals of cognitive-consistency theories (Abelson, 1983; Berkowitz & Devine, 1989; McGuire, 1968). Moreover, we agree with Aronson (1969, 1992) and Kunda (1990) that this research program is effectively confined to people's reactions to a particular type of threat to their self-concept. As such, we maintain, forced compliance is best understood as a straightforward example of the self-concept maintenance prong of motivated reasoning, and thus as a mere instantiation of coherence-based reasoning.

Lack of awareness and resiliency

One of the signature features of biased reasoning is that people are largely unaware of its biased nature, and when probed, they tend to misidentify the true drivers of their conclusions (Bargh, 1999; Fischhoff, 1975, 1977; Greenwald & Banaji, 1995; Wilson, 2002; Wilson & Brekke, 1994). Biased reasoning is generally unaffected by mitigation interventions, such as calling attention to the bias, explaining its adverse effects, and admonishing against engaging in it (Alicke et al., 1994; Dorrough et al., 2017; Evans et al., 1983; Fischhoff, 1975, 1977; Macrae et al., 1994; Markovits & Nantel, 1989; Monteith et al., 1998; Sá et al., 1999; Sassenberg & Moskowitz, 2005; Smith et al., 2012; Söllner et al., 2014). Biased reasoning persists in the face of feedback (Massey et al., 2011) and even in the presence of monetary incentives to avoid it (Dorrough et al., 2017; Engel & Glöckner, 2013; Irwin & Metzger, 1966; Lerner et al., 2004; Massey et al., 2011; Schwardmann et al., 2022; Simmons & Massey, 2012). These phenomena are consistent with the findings from the coherence studies that show that people are generally unaware that their judgments are being distorted by coherence shifts (Engel & Glöckner, 2013; Holyoak & Simon, 1999; D. Simon & Spiller, 2016), as described above in the section on Experimental Evidence. The beauty of coherence-based reasoning is that it enables people to solve seemingly intractable problems and navigate through complex worlds, and it does so with relatively little effort and largely beneath the level of conscious awareness. That very lack of awareness is what shields biased reasoning from introspection and correction and, as described above, enables naive realism.

Future directions and predictions

In addition to the experimental support summarized above in the section on Experimental Evidence, our framework provides fertile ground for additional lines of research.

In this article, we focused on reasoning tasks that entail an integration of multiple factors in the pursuit of a single conclusion. As mentioned above, this meant that we excluded examining biases that might affect the sources of the attributes that are brought to the task. Future work could go beyond that limitation to explore the role of coherence-based reasoning in forming and perhaps also biasing the attributes before they are subjected to the coherence-maximizing structural forces. Initial leads in this direction suggest that coherence may be playing a role in determining which cognitions are more likely to be brought to the task, primarily by way of selectively searching and retrieving information based on its propensity to cohere with a desired or emerging conclusion (Bhatia, 2016; Festinger, 1957; Fraser-Mackenzie & Dror, 2009; Jekel et al., 2018). This possibility calls for more research.

Another question worth exploring is the relationship between the emergence of coherence and the amount of thought devoted to the task. The prediction is that coherence will strengthen with the amount and intensity of thought (Chaiken & Yates, 1985; Tesser, 1978). Consistent with Tesser's (1978) work on mere thought leading to attitude polarization, Monroe and Read (2008) used a constraint-satisfaction model of attitudes to show that more time spent on processing a network could lead to more polarized attitudes. This also suggests that coherence will likely be mediated by individual differences that typically drive greater thought, such as the reasoner's need for consistency or closure (Festinger, 1957; Kruglanski & Freund, 1983; Russo et al., 2008; Swann & Brooks, 2012). At the same time, it is also possible that a lack of thought could lead to the formation of local attractors representing superficial propositions and a failure to explore and cross-activate countervailing factors.

Coherence may also be moderated by the demands of the task, such as an explicit decision is expected or a deadline is imposed. As discussed in the section Coherence and Cognitive Dissonance Theory, a series of studies has found that coherence emerges early on in the process, well before the decision is made, and even absent a decision altogether. Still, different processing tasks seem to result in different strengths of coherence effects. For example, although making a tentative decision or just memorizing the task yields significant coherence shifts, those effects tend to intensify once a commitment to a final decision has been made (Holyoak & Simon, 1999; D. Simon et al., 2001), which presumably requires greater processing and a greater need for certainty. Thus, we would expect that coherence effects would be increased by provision of an explicit deadline or decision.

It is also possible that some tasks do not lend themselves to reaching coherence because strong attributes lie on either side of the dilemma, such as a clash between central attitudes (see Howe & Krosnick, 2017; Judd & Krosnick, 1982), strongly held principles, or moral mandates (see Skitka, 2010). In such situations, the network can be said to be pulled in opposite directions by comparably strong attractors. Resolving such conflicts is bound to require extra mental effort and will likely result in less coherent and more ambivalent conclusions. At the extreme, the distortions required to attain coherence might prove untenable, and the ensuing state of conflict could lead to the postponement or abandonment of the task (see Janis & Mann, 1977).

Other research could focus on the fact that constraint satisfaction is an inherently dynamic process. Networks do not immediately jump to coherent states. Rather, they evolve over time, and these temporal dynamics might have important implications. For example, the gradual progression could predict that the impact of information will depend on when it is received. Information that is received early in the process could have more of an impact than information received later. As coherence evolves, the state of the system moves closer to and deeper into an attractor. Before the network has evolved toward a strong attractor, incoming information could be expected to be evaluated at face value. Although, once the network has evolved toward a strong attractor, it should become harder to reverse that progression. It should also be noted that difficult reasoning tasks could entail consecutive alternations between two vying models, in which case the dynamic nature of the process could favor the later simulated model. Temporal dynamics suggest that it should be instructive to use methods such as mouse tracking, eye tracking, or reaction times. For example, in a decision task with two alternatives, one could use mouse tracking to follow the evolution of the system as it moves from the initial state to the final decision. One could examine the mouse trajectory for signs of ambivalence (movement back and forth between the alternatives) or signs that the system steadily moves in one direction. One could also study the relative strengths of different sets of beliefs by examining the extent to which early movement is biased toward one alternative or the other. Consistent with these suggestions, Glöckner and colleagues (Fiedler & Glöckner, 2012; Glöckner et al., 2012, 2014) have used eye-tracking and reaction-time methods and have shown that it successfully predicts many aspects of the temporal dynamics of decision-making.

Our framework could be used to make predictions about the outcome of different configurations of facts, beliefs, and interventions in specific contexts. The computational model that we have provided should allow researchers to make explicit predictions concerning beliefs and relationships that have been appropriately measured. Possible predictions include the degree and direction of polarization, the existence and extent of bias, and the dynamics of the reasoning process.

Another testable proposition is that a biased conclusion is more likely to emerge when the nonnormative conclusion is couched in an organizing conceptual structure, such as a story, analogy, or other compelling conceptual framework. These structures provide tighter and more grounded connections among their building blocks, more pathways for the spread of activation, and stronger constraints. Future research could examine the proposition that networks are swayed by factors that increase the salience of certain pieces of information, such as when the biasing attributes are particularly fluent (Newman et al., 2020; Novemsky et al., 2007).

In this article, we fit our model to the data provided by Lord et al. (1979) and to data collected in our own studies (D. Simon et al., 2015). Because of length constraints, we could not have done the same with respect to every one of the biases discussed here. We invite other researchers to extend this fitting task to other data sets pertaining to biased reasoning to further bolster the framework and develop it into an overarching theory. That should also better position the framework to be tested against competing models.

Conclusion

One of the bedrocks of experimental psychology is that human reasoning is essential for interpreting and deciphering the stimuli through which the world presents itself (see Asch, 1952; Ichheiser, 1943; Krech & Crutchfield, 1948). Although these stimuli are frequently ambiguous, complex, incomplete, and numerous, the tasks people face often require reaching discrete and implementable conclusions, hence the indispensability of the integrative process, which we identify here as coherence-based reasoning. This form of reasoning, we propose, is driven by parallel constraint-satisfaction processes, which are premised on neuron-like brain structures and are fundamental to brain processing. The imposition of coherence on representations is adaptive in that it enables people to navigate their complicated worlds with confidence and resolve. But it must be acknowledged that this process is driven by global forces that might be dominated by nodes or links that are incorrect, overweighted, or otherwise nonnormative. At times, these global forces will be sufficiently strong to overwhelm the norms of principled reasoning, thus constituting a bias. In this article, we posited that coherence-based reasoning elucidates the mechanism by which some important biases occur and thus provides a comprehensive theoretical explanation for their occurrence. We hope that this parsimonious account will provide some depth and consolidation to the currently disjointed, incomplete, and undertheorized field of biased reasoning.

Transparency

Action Editor: John Dovidio
Editor: Interim Editorial Panel
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

ORCID iD

Dan Simon https://orcid.org/0000-0002-9821-3607

Acknowledgments

For helpful comments on an early draft, we thank Hal Arkes and Paul Thagard.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/17456916231204579

Note

1. We note that in the original data, the bottom layer is not as perfectly symmetrical as the model predicts for one of the four conditions (the anti-death-penalty participants judging antideterrence studies). Still, those judgments were distinguishable from judgments of pro-deterrence studies and were also distinguishable from the judgments of those same studies by supporters of the death penalty. Lacking access to the original data, we are reliant on the means reported in the article, which could mean that this discrepancy is due to differences in the structure of the underlying beliefs between pro- and anti-death-penalty adherents. Either way, the overall fit is, in our view, remarkable.

References

Abelson, R. P. (1963). Computer simulation of “hot cognition.” In S. Tomkins & S. Messick (Eds.), *Computer simulation of personality* (pp. 277–298). Wiley.

Abelson, R. P. (1983). Whatever became of consistency theory? *Personality and Social Psychology Bulletin*, *9*(1), 37–54.

Abelson, R. P., Aronson, E. E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, P. H. (1968). *Theories of cognitive consistency: A sourcebook*. Rand McNally.

Abelson, R. P., & Rosenberg, M. (1958). Symbolic psychologic: A model of attitudinal cognition. *Behavioral Science*, *3*(1), 1–8.

Abutaleb, Y., Parker, A., Dawsey, J., & Rucker, P. (2020, December 19). The inside story of how Trump’s denial, mismanagement and magical thinking led to the pandemic’s dark winter. *The Washington Post*. https://www.washingtonpost.com/graphics/2020/politics/trump-covid-pandemic-dark-winter/

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*(6), 1621–1630.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556–574.

Alicke, M. D., & Davis, T. L. (1989). The role of a posteriori victim information in judgments of blame and sanction. *Journal of Experimental Social Psychology*, *25*(4), 362–377.

Alicke, M. D., Davis, T. L., & Pezzo, M. V. (1994). A posteriori adjustment of a priori decision criteria. *Social Cognition*, *12*(4), 281–308.

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*(1), 1–48.

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, *191*, Article 104254. https://doi.org/10.1016/j.jpubeco.2020.104254

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236.

Allport, G. W. (1937). *Personality: A psychological interpretation*. Holt.

Allport, G. W., Clark, K., & Pettigrew, T. (1954). *The nature of prejudice*. Addison-Wesley.

Ames, M. (2020, December 21). How Ammon Bundy helped foment an anti-masker rebellion in Idaho. *The New Yorker*. https://www.newyorker.com/news/us-journal/how-ammon-bundy-helped-foment-an-anti-masker-rebellion-in-idaho

Amit, D. J. (1989). *Modeling brain function: The world of attractor neural networks*. Cambridge University Press.

Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological Bulletin*, *141*(3), 574–601.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486–498.

Arkes, H. R. (2016). A levels of processing interpretation of dual-system theories of judgment and decision making. *Theory & Psychology*, *26*(4), 459–475.

Aronson, E. (1969). The theory of cognitive dissonance: A current perspective. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4, pp. 1–34). Elsevier.

Aronson, E. (1992). The return of the repressed: Dissonance theory makes a comeback. *Psychological Inquiry*, *3*(4), 303–311.

Asch, S. E. (1940). Studies in the principles of judgments and attitudes: II. Determination of judgments by group and by ego standards. *The Journal of Social Psychology*, *12*(2), 433–465.

Asch, S. E. (1946a). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*, 258–290.

Asch, S. E. (1946b). Max Wertheimer's contribution to modern psychology. *Social Research*, *13*, 81–102.

Asch, S. E. (1952). *Social psychology*. Prentice-Hall.

Ask, K., & Pina, A. (2011). On being angry and punitive: How anger alters perception of criminal intent. *Social Psychological and Personality Science*, *2*(5), 494–499.

Atkinson, J. W., & Raynor, J. O. (1974). *Motivation and achievement*. Winston.

Banaji, M. R., Lemm, K. M., & Carpenter, S. J. (2001). The social unconscious. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 134–158). Blackwell Publishers.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–21). The Guilford Press.

Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). The Guilford Press.

Baron, J. (2008). *Thinking and deciding* (4th ed.). Cambridge University Press.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569–579.

Baumeister, R. F. (2016). Toward a general theory of motivation: Problems, challenges, opportunities, and the big picture. *Motivation and Emotion*, *40*(1), 1–10.

Baumeister, R. F., Dale, K., & Sommer, K. L. (1998). Freudian defense mechanisms and empirical findings in modern social psychology: Reaction formation, projection, displacement, undoing, isolation, sublimation, and denial. *Journal of Personality*, *66*(6), 1081–1124.

Baumeister, R. F., & Leary, M. R. (2017). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Interpersonal Development*, *117*(3), 57–89.

Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes. *Personality and Social Psychology Bulletin*, *20*(1), 3–19.

Bella, T. (2020, December 22). Maskless California men charged with shooting AK-47 at strip club after mask dispute, authorities say. *The Washington Post*. https://www.washingtonpost.com/nation/2020/12/22/california-shooting-mask-stripclub-ak47/

Bem, D. J., & McConnell, H. K. (1970). Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulation attitudes. *Journal of Personality and Social Psychology*, *14*(1), 23–31.

Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y., & Shefler, G. (1998). Seek and ye shall find: Test results are what you hypothesize they are. *Journal of Behavioral Decision Making*, *11*(4), 235–249.

Berkowitz, L., & Devine, P. G. (1989). Research traditions, analysis, and synthesis in social psychological theories: The case of dissonance theory. *Personality and Social Psychology Bulletin*, *15*(4), 493–507.

Bhatia, S. (2016). The dynamics of bidirectional thought. *Thinking & Reasoning*, *22*(4), 397–442.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21.

Bodenhausen, G. V. (1993). Emotions, arousal, and stereotypic judgments: A heuristic model of affect and stereotyping. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition and stereotyping* (pp. 13–37). Academic Press.

Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgment: The differential impact of anger and sadness. *European Journal of Social Psychology*, *24*(1), 45–62.

Boiney, L. G., Kennedy, J., & Nye, P. (1997). Instrumental bias in motivated reasoning: More when more is needed. *Organizational Behavior and Human Decision Processes*, *72*(1), 1–24.

Brandt, M. J., & Crawford, J. T. (2020). Worldview conflict and prejudice. In B. Gawronski (Ed.), *Advances in experimental social psychology* (Vol. 61, pp. 1–66). Academic Press.

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, *52*, 384–389. http://dx.doi.org/10.1037/h0041006

Brehm, J. W., & Cohen, A. R. (1962). *Explorations in cognitive dissonance*. Wiley. http://dx.doi.org/10.1037/11622-000

Brownstein, A. L. (2003). Biased predecision processing. *Psychological Bulletin*, *129*, 545–568. http://dx.doi.org/10.1037/0033-2909.129.4.545

Brownstein, A. L., Read, S. J., & Simon, D. (2004). Bias at the racetrack: Effects of individual expertise and task importance on predecision reevaluation of alternatives. *Personality and Social Psychology Bulletin*, *30*(7), 891–904.

Bruner, J. S. (1957). Discussion. In J. S. Bruner, E. Brunswik, L. Festinger, F. Heider, K. F. Muenzinger, C. E. Osgood, & D. Rapaport (Eds.), *Contemporary approaches to cognition* (pp. 151–156). Harvard University Press.

Brunswik, E. (1952). *The conceptual framework of psychology*. University of Chicago Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217.

Carpenter, S. M., Yates, J. F., Preston, S. D., & Chen, L. (2016). Regulating emotions during difficult multiattribute decision making: The role of pre-decisional coherence shifting. *PLOS ONE*, *11*(3), Article e0150873. https://doi.org/10.1371/journal.pone.0150873

Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and Emotion*, *30*(2), 105–110.

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin*, *135*(2), 183–204.

Chaiken, S., & Yates, S. (1985). Affective-cognitive consistency and thought-induced attitude polarization. *Journal of Personality and Social Psychology*, *49*(6), 1470–1481.

Chater, N., & Oaksford, M. (2012). Normative systems: Logic, probability, and rational choice. In K. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 11–21). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0002

Chaxel, A. S., Russo, J. E., & Wiggins, C. (2016). A goal-priming approach to cognitive consistency: Applications to judgment. *Journal of Behavioral Decision Making*, *29*(1), 37–51.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. MIT Press.

Clark, C. J., & Winegard, B. M. (2020). Tribalism in war and peace: The nature and evolution of ideological epistemology and its significance for modern social science. *Psychological Inquiry*, *31*(1), 1–22.

Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology*, *85*(5), 808–822.

Cooper, A. C., Woo, C. Y., & Dunkelberg, W. C. (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing*, *3*(2), 97–108.

Cooper, J. (2007). *Cognitive dissonance: 50 years of a classic theory*. Sage.

Cooper, J., & Fazio, R. H. (1984). A new look at dissonance theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 229–266). Academic Press.

Cross, K. P. (1977). Not can, but will college teaching be improved? *New Directions for Higher Education*, *1977*(17), 1–15.

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*(1), 20–33. https://doi.org/10.1037/0022-3514.44.1.20

Dawes, R. M. (1976). Shallow psychology. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 3–12). Lawrence Erlbaum.

De Houwer, J., Fiedler, K., & Moors, A. (2011). Strengths and limitations of theoretical explanations in psychology: Introduction to the special section. *Perspectives on Psychological Science*, *6*(2), 161–162.

DeKay, M. L. (2015). Predecisional information distortion and the self-fulfilling prophecy of early preferences in choice. *Current Directions in Psychological Science*, *24*(5), 405–411.

DeKay, M. L., Miller, S. A., Schley, D. R., & Erford, B. M. (2014). Proleader and antileader information distortion and their effects on choice and postchoice memory. *Organizational Behavior and Human Decision Processes*, *125*, 134–150. http://dx.doi.org/10.1016j.obhdp .2014.07.003

DeKay, M. L., Stone, E. R., & Sorenson, C. M. (2012). Sizing up information distortion: Quantifying its effect on the subjective values of choice options. *Psychonomic Bulletin & Review*, *19*, 349–356. http://dx.doi.org/10.3758/s13423-011-0184-8

DeKay, M. L., Stone, E. R., & Miller, S. A. (2011). Leader-driven distortion of probability and payoff information affects choices between risky prospects. *Journal of Behavioral Decision Making*, *24*(4), 394–411.

De Neys, W., & Verschueren, N. (2006). Working memory capacity and a notorious brain teaser: The case of the Monty Hall Dilemma. *Experimental Psychology*, *53*(2), 123–131.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18.

Ditto, P. H., Jemmott, J. B., III, & Darley, J. M. (1988). Appraising the threat of illness: A mental representational approach. *Health Psychology*, *7*(2), 183–201.

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., Celniker, J. B., & Zinger, J. F. (2019). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science*, *14*(2), 273–291.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and non-preferred conclusions. *Journal of Personality and Social Psychology*, *63*(4), 568–584.

Dorrough, A. R., Glöckner, A., Betsch, T., & Wille, A. (2017). When knowledge activated from memory intrudes on probabilistic inferences from description-the case of stereotypes. *Acta Psychologica*, *180*, 64–78.

Dougherty, M. R., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*(1), 180–209.

Douglas, K. M., & Sutton, R. M. (2008). The hidden impact of conspiracy theories: Perceived and actual influence of theories surrounding the death of Princess Diana. *The Journal of Social Psychology*, *148*(2), 210–222.

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*(1), 22–37.

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, *34*(4), 590–598.

Dunning, D. (2015). Motivated cognition in self and social thought. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Vol. 1. Attitudes and social cognition* (pp. 777–803). American Psychological Association.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*(3), 69–106.

Dweck, C. S. (2017). From needs to goals and representations: Foundations for a unified theory of motivation, personality, and development. *Psychological Review*, *124*(6), 689–719.

Eberhardt, J. L. (2020). *Biased: Uncovering the hidden prejudice that shapes what we see, think, and do*. Penguin Books.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, *71*(1), 5–24.

Egli Anthonioz, N., Schweizer, M., Vuille, J., & Kuhn, A. (2019). Role-induced bias in criminal prosecutions. *European Journal of Criminology*, *16*(4), 452–465.

Ehrlinger, J., Gilovich, T., & Ross, L. (2005). Peering into the bias blind spot: People's assessments of bias in themselves and others. *Personality and Social Psychology Bulletin*, *31*(5), 680–692.

Eigen, Z. J., & Listokin, Y. (2012). Do lawyers really believe their own hype, and should they? A natural experiment. *The Journal of Legal Studies*, *41*(2), 239–267.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). Oxford University Press.

Elster, J. (1983). *Sour grapes*. Cambridge University Press.

Engel, C., & Glöckner, A. (2013). Role-induced bias in court: An experimental analysis. *Journal of Behavioral Decision Making*, *26*(3), 272–284.

Engel, C., Timme, S., & Glöckner, A. (2020). Coherence-based reasoning and order effects in legal judgments. *Psychology, Public Policy, and Law*, *26*(3), 333–352.

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology*, *79*(6), 861–875.

Erisen, C., Lodge, M., & Taber, C. S. (2014). Affective contagion in effortful political thinking. *Political Psychology*, *35*(2), 187–206.

Evans, J. S. B., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306.

Evans, J. S. B., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382–389.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

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

Festinger, L. (1964). *Conflict, decision, and dissonance*. Stanford University Press.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, *58*(2), 203–210.

Fetchenhauer, D., & Dunning, D. (2006). Perceptions of prosociality and solidarity in self and others. In D. Fetchenhauer, A. Flache, A. P. Buunk, & S. M. Lindenberg (Eds.), *Solidarity and prosocial behavior* (pp. 61–74). Springer.

Fiedler, S., & Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis. *Frontiers in Psychology*, *3*, Article 335. https://doi.org/10.3389/fpsyg.2012.00335

Fischer, A. H., & Roseman, I. J. (2007). Beat them or ban them: The characteristics and social functions of anger and contempt. *Journal of Personality and Social Psychology*, *93*(1), 103–115.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288–299.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(2), 349–358.

Fischle, M. (2000). Mass response to the Lewinsky Scandal: Motivated reasoning or Bayesian updating? *Political Psychology*, *21*(1), 135–159.

Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology*, *18*(2), 255–297.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 357–411). McGraw-Hill.

Forgas, J. P. (2008). Affect and cognition. *Perspectives on Psychological Science*, *3*(2), 94–101.

Fraser-Mackenzie, P. A., & Dror, I. E. (2009). Selective information sampling: Cognitive coherence in evaluation of a novel item. *Judgment and Decision Making*, *4*(4), 307–316.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*(2), 247–279.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*(4), 652–670.

Gaertner, S. L., Mann, J. A., Dovidio, J. F., Murrell, A. J., & Pomare, M. (1990). How does cooperation reduce intergroup bias? *Journal of Personality and Social Psychology*, *59*(4), 692–704.

Georges, L. C., Wiener, R. L., & Keller, S. R. (2013). The angry juror: Sentencing decisions in first-degree murder. *Applied Cognitive Psychology*, *27*(2), 156–166.

Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1055–1075.

Glöckner, A., & Betsch, T. (2012). Decisions beyond boundaries: When more information is processed faster than less. *Acta Psychologica*, *139*, 532–542.

Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making*, *23*(5), 439–462.

Glöckner, A., Heinen, T., Johnson, J. G., & Raab, M. (2012). Network approaches for expert decisions in sports. *Human Movement Science*, *31*(2), 318–333. https://doi.org/10.1016/j.humov.2010.11.002

Glöckner, A., Hilbig, B. E., & Jekel, M. (2014). What is adaptive about adaptive decision making? A parallel constraint satisfaction account. *Cognition*, *133*(3), 641–666.

Goethals, G. R., & Reckman, R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology*, *9*(6), 491–501.

Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology*, *29*(5–6), 781–795.

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine*, *165*(13), 1493–1499.

Greenhoot, A. F., Semb, G., Colombo, J., & Schreiber, T. (2004). Prior beliefs and methodological concepts in scientific reasoning. *Applied Cognitive Psychology*, *18*(2), 203–221.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 59–100). Cambridge University Press.

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, *26*(2–3), 103–117.

Guilbeault, D., Becker, J., & Centola, D. (2018). Social learning and partisan bias in the interpretation of climate trends. *Proceedings of the National Academy of Sciences, USA*, *115*(39), 9714–9719.

Haberman, M., & Schmidt, M. S. (2020, December 23). President Trump: Unhappy, unleashed and unpredictable. *The New York Times*. https://www.nytimes.com/2020/12/23/us/politics/trump-end-of-presidency.html

Halperin, E. (2008). Group-based hatred in intractable conflict in Israel. *Journal of Conflict Resolution*, *52*(5), 713–736.

Halperin, E., & Bar-Tal, D. (2011). Socio-psychological barriers to peace making: An empirical examination within the Israeli Jewish society. *Journal of Peace Research*, *48*(5), 637–651.

Harmon-Jones, E. E. (2019). *Cognitive dissonance: Reexamining a pivotal theory in psychology*. American Psychological Association.

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 968–987). John Wiley & Sons.

Hastie, R., & Dawes, R. M. (2009). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Sage.

Hastorf, A. H., & Cantril, H. (1954). They saw a game; a case study. *The Journal of Abnormal and Social Psychology*, *49*(1), 129–134.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311–327.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, *21*, 107–112.

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

Heider, F. (1960). The gestalt theory of motivation. In *Nebraska symposium on motivation* (Vol. 8, pp. 145–172). University of Nebraska Press.

Heine, S. J., Proulx, T., & Vohs, K. D. (2006). The meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review*, *10*(2), 88–110.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Addison-Wesley.

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, *138*(2), 211–237.

Hinkle, S., & Schopler, J. (1986). Bias in the evaluation of in-group and out-group performance. *Psychology of Intergroup Relations*, *2*, 196–212.

Holyoak, K. J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 301–335). Cambridge University Press.

Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin*, *142*(11), 1179–1203.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology-General*, *128*(1), 3–31. https://doi.org/10.1037/0096-3445.128.1.3

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*(3), 295–355.

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

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

Hornsey, M. J., Finlayson, M., Chatwood, G., & Begeny, C. T. (2020). Donald Trump and vaccination: The effect of political identity, conspiracist ideation and presidential tweets on vaccine hesitancy. *Journal of Experimental Social Psychology*, *88*, Article 103947. https://doi.org/10.1016/j.jesp.2019.103947

Howe, L. C., & Krosnick, J. A. (2017). Attitude strength. *Annual Review of Psychology*, *68*(1), 327–351.

Hulsizer, M. R., Munro, G. D., Fagerlin, A., & Taylor, S. P. (2004). Molding the past: Biased assimilation of historical information. *Journal of Applied Social Psychology*, *34*(5), 1048–1074.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*(2), 220–264.

Ichheiser, G. (1943). Misinterpretations of personality in everyday life and the psychologist's frame of reference. *Character & Personality; A Quarterly for Psychodiagnostic & Allied Studies*, *12*, 145–160. https://doi.org/10.1111/j.1467-6494.1943.tb01953.x

Ichheiser, G. (1949). Misunderstandings in human relations: A study in false social perception. *American Journal of Sociology*, *55*(Pt. 2), 1–72.

Irwin, F. W. (1953). Stated expectations as functions of probability and desirability of outcomes. *Journal of Personality*, *21*, 329–335. https://doi.org/10.1111/j.1467-6494.1953.tb01775.x

Irwin, F. W., & Metzger, M. J. (1966). Effects of probabilistic independent outcomes upon predictions. *Psychonomic Science*, *5*(2), 79–80.

Janis, I. L., & King, B. T. (1954). The influence of role playing on opinion change. *The Journal of Abnormal and Social Psychology*, *49*(2), 211–218.

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. Free Press.

Jekel, M., Glöckner, A., & Bröder, A. (2018). A new and unique prediction for cue-search in a parallel-constraint satisfaction network model: The attraction search effect. *Psychological Review*, *125*(5), 744–768.

Jekel, M., Glöckner, A., Fiedler, S., & Bröder, A. (2012). The rationality of different kinds of intuitive decision processes. *Synthese*, *189*(1), 147–160.

Johnson, D. D., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution*, *28*(8), 474–481.

Judd, C. M., & Krosnick, J. A. (1982). Attitude centrality, organization, and measurement. *Journal of Personality and Social Psychology*, *42*(3), 436–447.

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, *100*(1), 109–128. https://doi.org/10.1037/0033-295X.100.1.109

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732–735.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Kay, A. C., Jimenez, M. C., & Jost, J. T. (2002). Sour grapes, sweet lemons, and the anticipatory rationalization of the status quo. *Personality and Social Psychology Bulletin*, *28*(9), 1300–1312.

Keltner, D., Ellsworth, P. C., & Edwards, K. (1993). Beyond simple pessimism: Effects of sadness and anger on social perception. *Journal of Personality and Social Psychology*, *64*(5), 740–752.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*(6), 533–550.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, *32*, 385–418.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211–228.

Klein, W., & Kunda, Z. (1989, March). *Motivated person perception: Justifying desired conclusions* [Paper presentation]. Meeting of the Eastern Psychological Association, Boston, MA, United States.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *Institute of medicine. To err is human: Building a safer health system*. National Academy Press.

Krech, D., & Crutchfield, R. S. (1948). Perceiving the world. In D. Krech & R. S. Crutchfield, *Theory and problems of social psychology* (pp. 76–109). McGraw-Hill. https://doi.org/10.1037/10024-003

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, *133*(1), 95–121.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, *27*(3), 313–327.

Kruglanski, A. W. (1990). Lay epistemic theory in social-cognitive psychology. *Psychological Inquiry*, *1*(3), 181–197.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, *19*(5), 448–468.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

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

Langer, E. J., & Abelson, R. P. (1974). A patient by any other name. . .: Clinician group difference in labeling bias. *Journal of Consulting and Clinical Psychology*, *42*(1), 4–9.

Larwood, L., & Whittaker, W. (1977). Managerial myopia: Self-serving biases in organizational planning. *Journal of Applied Psychology*, *62*(2), 194–198.

Lazarus, R. S. (1966). *Psychological stress and the coping process*. McGraw-Hill.

Lepper, M. R. (1994). "Hot" versus "cold" cognition: An Abelsonian voyage. In R. C. Schank & E. Langer (Eds.), *Beliefs, reasoning, and decision making: Psycho-logic in honor of Bob Abelson* (pp. 237–275). Erlbaum.

Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin*, *24*(6), 563–574.

Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science*, *15*(5), 337–341.

Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). The role of conspiracist ideation and worldviews in predicting rejection of science. *PLOS ONE*, *8*(10), Article e75637. https://doi.org/10.1371/journal.pone.0075637

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and Brain Sciences*, *28*, 169–245.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge University Press.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, Article e1. https://doi.org/10.1017/S0140525X1900061X

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 619–642). Oxford University Press.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. John Wiley & Sons.

Ludolph, R., & Schulz, P. J. (2018). Debiasing health-related judgments and decision making: A systematic review. *Medical Decision Making*, *38*(1), 3–13.

Lundberg, C. G. (2007). Models of emerging contexts in risky and complex decision settings. *European Journal of Operational Research*, *177*, 1363–1374.

MacDonald, M. C., & Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In M. A. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 581–611). Elsevier.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, *67*(5), 808–817.

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, *1*(2), 161–175.

Manoogian, J. (2016). *Cognitive bias codex*. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Cognitive_bias_codex_en.svg

Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, *17*, 11–17.

Markus, H., & Zajonc, R. B. (1985). The cognitive perspective in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 1, pp. 137–230). Random House.

Marlow, D., & Crowne, D. P. (1961). Social desirability and response to perceived situational demands. *Journal of Consulting Psychology*, *25*(2), 109–115.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370–396. https://doi.org/10.1037/h0054346

Massey, C., Simmons, J. P., & Armor, D. A. (2011). Hope over experience: Desirability and the persistence of optimism. *Psychological Science*, *22*(2), 274–281.

Matheson, K., & Dursun, S. (2001). Social identity precursors to the hostile media phenomenon: Partisan perceptions of coverage of the Bosnian conflict. *Group Processes & Intergroup Relations*, *4*(2), 116–125.

Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna principle: Selectivity in language, memory, and thought*. Schenkman Publishing Company.

McClelland, J. L. (1985). Putting knowledge in its place: A scheme for programming parallel processing structures on the fly. *Cognitive Science*, *9*(1), 113–146.

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, *38*(6), 1139–1189.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I.

An account of basic findings. *Psychological Review*, *88*(5), 375–407.

McClelland, J. L., & Rumelhart, D. E., & PDP Research Group. (1986). *Parallel distributed processing* (Vol. 2). MIT Press.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & P. R. Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 3–44). MIT Press.

McFarland, C., & Ross, M. (1982). Impact of causal attributions on affective reactions to success and failure. *Journal of Personality and Social Psychology*, *43*(5), 937–946.

McGuire, W. J. (1960). Cognitive consistency and attitude change. *The Journal of Abnormal and Social Psychology*, *60*(3), 345–353.

McGuire, W. J. (1968). Theory of the structure of human thought. In R. P. Abelson, E. E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 140–162). Rand McNally.

McKenzie, C. R. (2006). Increased sensitivity to differentially diagnostic answers using familiar materials: Implications for confirmation bias. *Memory & Cognition*, *34*(3), 577–588.

McPherson Frantz, C. (2006). I AM being fair: The bias blind spot as a stumbling block to seeing both sides. *Basic and Applied Social Psychology*, *28*(2), 157–167.

Medin, D. L., & Heit, E. (1999). Categorization. In B. M. Bly & D. E. Rumelhart (Eds.), *Cognitive science* (pp. 99–143). Elsevier.

Meffert, M. F., Chung, S., Joiner, A. J., Waks, L., & Garst, J. (2006). The effects of negativity and motivated information processing during a political campaign. *Journal of Communication*, *56*(1), 27–51.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293.

Melnikoff, D. E., & Strohminger, N. (2020). The automatic influence of advocacy on lawyers and novices. *Nature Human Behaviour*, *4*(12), 1258–1264.

Middleton, W., Harris, P., & Surman, M. (1996). Give em enough rope: Perception of health and safety risks in bungee jumpers. *Journal of Social and Clinical Psychology*, *15*(1), 68–79.

Miller, D. T., & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology*, *74*(1), 53–62.

Molouki, S., & Pronin, E. (2015). Self and other. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Vol. 1. Attitudes and social cognition* (pp. 387–414). American Psychological Association. https://doi.org/10.1037/14341-013

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.

Montaro, D. (2020, December 9). Poll: Only quarter of Republicans accept 2020 election results. *npr*. https://www.npr.org/2020/12/09/944385798/poll-just-a-quarter-of-republicans-accept-election-outcome

Monteith, M. J., Sherman, J. W., & Devine, P. G. (1998). Suppression as a stereotype control strategy. *Personality and Social Psychology Review*, *2*(1), 63–82.

Montgomery, H., & Willén, H. (1999). Decision making and action: The search for a good structure. In P. Juslin & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (pp. 147–173). Erlbaum.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, *14*(10), 435–440.

Morgan, J. J., & Morton, J. T. (1944). The distortion of syllogistic reasoning produced by personal convictions. *The Journal of Social Psychology*, *20*(1), 39–59.

Munro, G. D. (2010). The scientific impotence excuse: Discounting belief-threatening scientific abstracts. *Journal of Applied Social Psychology*, *40*(3), 579–600.

Munro, G. D., Ditto, P. H., Lockhart, L. K., Fagerlin, A., Gready, M., & Peterson, E. (2002). Biased assimilation of sociopolitical arguments: Evaluating the 1996 US presidential debate. *Basic and Applied Social Psychology*, *24*(1), 15–26.

Munro, G. D., Lasane, T. P., & Leary, S. P. (2010). Political partisan prejudice: Selective distortion and weighting of evaluative categories in college admissions applications. *Journal of Applied Social Psychology*, *40*(9), 2434–2462.

Murrie, D. C., & Boccaccini, M. T. (2015). Adversarial allegiance among expert witnesses. *Annual Review of Law and Social Science*, *11*, 37–55.

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science*, *24*(10), 1889–1897.

Nelson, T. E., Biernat, M. R., & Manis, M. (1990). Everyday base rates (sex stereotypes): Potent and resilient. *Journal of Personality and Social Psychology*, *59*(4), 664–675.

Nesbitt-Larking, P., & Kinnvall, C. (2012). The discursive frames of political psychology. *Political Psychology*, *33*(1), 45–59.

Newman, E. J., Jalbert, M. C., Schwarz, N., & Ly, D. P. (2020). Truthiness, the illusory truth effect, and the role of need for cognition. *Consciousness and Cognition*, *78*, Article 102866. https://doi.org/10.1016/j.concog.2019.102866

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.

Nisbett, R. E., & Ross, L. D. (1980). *Human inference: Strategies and shortcomings of social judgment*. Prentice-Hall.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. https://doi.org/10.1037/0022-3514.35.4.250

Novemsky, N., Dhar, R., Schwarz, N., & Simonson, I. (2007). Preference fluency in choice. *Journal of Marketing Research*, *44*(3), 347–356.

Nuñez, N., Schweitzer, K., Chai, C. A., & Myers, B. (2015). Negative emotions felt during trial: The effect of fear, anger, and sadness on juror decision making. *Applied Cognitive Psychology*, *29*(2), 200–209.

Oeberst, A., & Imhoff, R. (2023). Toward parsimony in bias research: A proposed common framework of belief-consistent information processing for a set of biases. *Perspectives on Psychological Science*. Advance online publication. https://doi.org/10.1177/17456916221148147

O'Reilly, R. C., Munakata, Y., Frank, M. J., & Hazy, T. E. (2012). *Computational cognitive neuroscience*. Wiki Book. https://github.com/CompCogNeuro/ed4

O'Reilly, R. C., Munakata, Y., Frank, M. J., & Hazy, T. E., & Contributors. (2020). *Computational cognitive neuroscience* (4th ed.). WikiBook. https://CompCogNeuro.org

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review*, *62*(1), 42–55.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, *11*(6), 988–1010.

Osman, M. (2013). A case study: Dual-process theories of higher cognition—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, *8*(3), 248–252.

O'Sullivan, E. D., & Schofield, S. (2018). Cognitive bias in clinical medicine. *Journal of the Royal College of Physicians of Edinburgh*, *48*(3), 225–231.

Paley, J. (2008). Empiricism. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research: Methods* (Vol. 2). Sage. https://doi.org/10.4135/9781412963909

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*(1), 87–131.

Phillips, F. (2002). The distortion of criteria after decision-making. *Organizational Behavior and Human Decision Processes*, *88*(2), 769–784.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development*, *15*(1), 1–12.

Pittman, T. S., & D'Agostino, P. R. (1989). Motivation and cognition: Control deprivation and the nature of subsequent information processing. *Journal of Experimental Social Psychology*, *25*(6), 465–480.

Powell, J. L. (1988). A test of the knew-it-all-along effect in the 1984 presidential and statewide elections. *Journal of Applied Social Psychology*, *18*(9), 760–773.

Pronin, E. (2008). How we see ourselves and how we see others. *Science*, *320*(5880), 1177–1180.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, *111*(3), 781–799.

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, *43*(4), 565–578.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177/0146167202286008

Pyszczynski, T. (1982). Cognitive strategies for coping with uncertain outcomes. *Journal of Research in Personality*, *16*(3), 386–399.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. *Advances in Experimental Social Psychology*, *20*, 297–340.

Pyszczynski, T., Greenberg, J., & Holt, K. (1985). Maintaining consistency between self-serving beliefs and available data: A bias in information evaluation. *Personality and Social Psychology Bulletin*, *11*(2), 179–190.

Quigley, B. M., & Tedeschi, J. T. (1996). Mediating effects of blame attributions on feelings of anger. *Personality and Social Psychology Bulletin*, *22*(12), 1280–1288.

Read, S. J. (1987). Constructing causal scenarios: A knowledge structure approach to causal reasoning. *Journal of Personality and Social Psychology*, *52*(2), 288–302.

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, *65*(3), 429–447.

Read, S. J., & Miller, L. C. (1994). Dissonance and balance in belief systems: The promise of parallel constraint satisfaction processes and connectionist modeling approaches. In R. C. Schank & E. J. Langer (Eds.), *Beliefs, reasoning, and decision making: Psycho-logic in honor of Bob Abelson* (pp. 209–235). Erlbaum.

Read, S. J., & Simon, D. (2012). Parallel constraint satisfaction as a mechanism for cognitive consistency. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 66–86). The Guilford Press.

Read, S. J., Vanman, E. J., & Miller, L. C. (1997). Connectionism, parallel constraint satisfaction processes, and gestalt principles: (Re)introducing cognitive dynamics to social psychology. *Personality and Social Psychology Review*, *1*(1), 26–53.

Richardson, J. D., Huddy, W. P., & Morgan, S. M. (2008). The hostile media effect, biased assimilation, and perceptions of a presidential debate. *Journal of Applied Social Psychology*, *38*(5), 1255–1270. https://doi.org/10.1111/j.1559-1816.2008.00347.x

Risen, J. (2006). *State of war: The secret history of the CIA and the Bush administration*. Simon & Schuster.

Roese, N. J. (2004). Twisted pair: Counterfactual thinking and the hindsight bias. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 258–273). Blackwell.

Roese, N. J., & Olson, J. M. (2007). Better, stronger, faster: Self-serving judgment, affect regulation, and the optimal vigilance hypothesis. *Perspectives on Psychological Science*, *2*(2), 124–141.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Erlbaum.

Roseman, I. J., & Smith, C. A. (2001). Appraisal theory. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 3–19). Oxford University Press.

Rosenbaum, M. E. (1986). The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality*

and Social Psychology, *51*(6), 1156–1166. https://doi.org/10.1037/0022-3514.51.6.1156
Rosenberg, M. J. (1968). Hedonism, inauthenticity, and other goads toward expansion of a consistency theory. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 73–111). Rand McNally.
Rosenberg, M. J., & Abelson, R. P. (1960). An analysis of cognitive balancing. In M. J. Rosenberg, C. I. Hovland, R. P. Abelson, W. J. McGuire, & J. W. Brehm (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (Vol. 3, pp. 112–163). Yale University Press.
Ross, L., & Ward, A. (1996). Naive realism: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *Values and knowledge* (pp. 103–135). Erlbaum.
Ross, M., McFarland, C., & Fletcher, G. J. (1981). The effect of attitude on the recall of personal histories. *Journal of Personality and Social Psychology*, *40*(4), 627–634.
Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60–94.
Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 7–57). MIT Press.
Russo, J. E. (2014). The predecisional distortion of information. In E. A. Wilhelms & V. F. Reyna (Eds.), *Neuroeconomics, judgment, and decision making* (pp. 91–110). Psychology Press.
Russo, J. E., Carlson, K. A., Meloy, M. G., & Yong, K. (2008). The goal of consistency as a cause of information distortion. *Journal of Experimental Psychology: General*, *137*(3), 456–470.
Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, *66*(1), 102–110.
Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional distortion of product information. *Journal of Marketing Research*, *35*(4), 438–452.
Rutter, D. R., Quine, L., & Albery, I. P. (1998). Perceptions of risk in motorcyclists: Unrealistic optimism, relative realism and predictions of behaviour. *British Journal of Psychology*, *89*(4), 681–696.
Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *91*(3), 497–510.
Sachdev, I., & Bourhis, R. Y. (1987). Status differentials and intergroup behaviour. *European Journal of Social Psychology*, *17*(3), 277–293.
Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893.
Sassenberg, K., & Moskowitz, G. B. (2005). Don't stereotype, think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology*, *41*(5), 506–514.
Schank, R. C., & Abelson, R. P. (2013). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press.
Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. *Review of Personality & Social Psychology*, *5*, 57–63.
Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (Vol. 92, pp. 92–120). Oxford University Press.
Schwardmann, P., Tripodi, E., & Van der Weele, J. J. (2022). Self-persuasion: Evidence from field experiments at international debating competitions. *American Economic Review*, *112*(4), 1118–1146.
Secord, P. F. (1959). Stereotyping and favorableness in the perception of Negro faces. *The Journal of Abnormal and Social Psychology*, *59*(3), 309–314.
Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, *53*(1), 491–517.
Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207–222.
Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061–1086.
Sherif, M., Sherif, M., Harvey, O. J., Hood, W. R., Sherif, C. W., & White, J. (1961). *The robbers cave experiment: Intergroup conflict and cooperation* (Vol. 10). University of Oklahoma.
Sherman, D. K., & Kim, H. S. (2005). Is there an "I" in "team"? The role of the self in group-serving judgments. *Journal of Personality and Social Psychology*, *88*(1), 108–120.
Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review*, *103*(2), 219–240. https://doi.org/10.1037/0033-295X.103.2.219
Simmons, J. P., & Massey, C. (2012). Is optimism real? *Journal of Experimental Psychology: General*, *141*(4), 630–634.
Simon, D. (2004). A third view of the black box: Cognitive coherence in legal decision making. *University of Chicago Law Review*, *71*(2), 511–586.
Simon, D. (2012). *In doubt: The psychology of the criminal justice process*. Harvard University Press.
Simon, D., Ahn, M., Stenstrom, D. M., & Read, S. J. (2020). The adversarial mindset. *Psychology, Public Policy, and Law*, *26*(3), 353–377.
Simon, D., & Holyoak, K. (2002). Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review*, *6*(4), 283–294. https://doi.org/10.1207/S15327957PSPR0604_03
Simon, D., Krawczyk, D., & Holyoak, K. (2004). Construction of preferences by constraint satisfaction. *Psychological Science*, *15*(5), 331–336. https://doi.org/10.1111/j.0956-7976.2004.00678.x

Simon, D., Pham, L., Le, Q., & Holyoak, K. (2001). The emergence of coherence over the course of decision making. *Journal of Experimental Psychology-Learning Memory and Cognition*, *27*(5), 1250–1260. https://doi.org/10.1037//0278-7393.27.5.1250

Simon, D., Snow, C., & Read, S. (2004). The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*, *86*(6), 814–837. https://doi.org/10.1037/0022-3514.86.6.814

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.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, *4*(4), 267–281.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*, 813–838.

Smith, C. T., Ratliff, K. A., & Nosek, B. A. (2012). Rapid assimilation: Automatically integrating new information with existing beliefs. *Social Cognition*, *30*(2), 199–219.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1., pp. 194–281). MIT Press.

Smolensky, P. (1989). Connectionist modeling: Neural computation/mental connections. In L. Nadel, L. A. Cooper, P. Culicover, & R. M. Harnish (Eds.), *Computational models of cognition and perception. Neural connections, mental computation* (pp. 49–67). MIT Press.

Söllner, A., Bröder, A., Glöckner, A., & Betsch, T. (2014). Single-process versus multiple-strategy models of decision making: Evidence from an information intrusion paradigm. *Acta Psychologica*, *146*, 84–96.

Sorrentino, R. M., & Higgins, E. T. (1986). Motivation and cognition: Warming up to synergism. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 1, pp. 3–19). The Guilford Press.

Spellman, B. A., Ullman, J. B., & Holyoak, K. J. (1993). A coherence model of cognitive consistency: Dynamics of attitude change during the Persian Gulf War. *Journal of Social Issues*, *49*(4), 147–165.

Stagner, R. (1951). Homeostasis as a unifying concept in personality theory. *Psychological Review*, *58*(1), 5–17. https://doi.org/10.1037/h0063598

Stanley, M. L., Henne, P., Yang, B. W., & De Brigard, F. (2020). Resistance to position change, motivated reasoning, and polarization. *Political Behavior*, *42*(3), 891–913.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Psychology Press.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695.

Stewart, M. A., Ryan, E. B., & Giles, H. (1985). Accent and social class effects on status and solidarity evaluations. *Personality and Social Psychology Bulletin*, *11*(1), 98–105.

Stolberg, S. G., & Dao, J. (2005, January 7). Congress ratifies Bush victory after challenge. *The New York Times*. https://www.nytimes.com/2005/01/07/politics/congress-ratifies-bush-victory-after-challenge.html

Suri, G., & Gross, J. J. (2012). Emotion regulation and successful aging. *Trends in Cognitive Sciences*, *16*(8), 409–410.

Suri, G., Gross, J. J., & McClelland, J. L. (2020). Value-based decision making: An interactive activation perspective. *Psychological Review*, *127*(2), 153–185.

Svenson, O. (1999). Differentiation and consolidation theory: Decision making processes before and after a choice. In P. Juslin & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (pp. 175–197). Erlbaum.

Svenson, O., Fischhoff, B., & MacGregor, D. (1985). Perceived driving safety and seatbelt usage. *Accident Analysis & Prevention*, *17*(2), 119–133.

Swami, V., Chamorro-Premuzic, T., & Furnham, A. (2010). Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs. *Applied Cognitive Psychology*, *24*(6), 749–761.

Swann, W. B., Jr., & Brooks, M. (2012). Why threats trigger compensatory reactions: The need for coherence and quest for self-verification. *Social Cognition*, *30*(6), 758–777.

Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, *31*(2), 137–155.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178.

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.

Tannenbaum, P. H. (1968). The congruity principle: Retrospective reflections and recent research. In R. P. Abelson, E. E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 52–72). Rand McNally.

Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 289–338). Elsevier.

Tesser, A., & Paulhus, D. (1983). The definition of self: Private and public self-evaluation management strategies. *Journal of Personality and Social Psychology*, *44*(4), 672–682. https://doi.org/10.1037/0022-3514.44.4.672

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, *109*(3), 451–471.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(3), 435–502.

Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. MIT Press.

Thagard, P. (2019). *Brain-mind: From neurons to consciousness and creativity (treatise on mind and society)*. Oxford University Press.

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, *22*(1), 1–24.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*(1), 25–29.

Tidwell, J. W., Buttaccio, D., Chrabaszcz, J. S., Dougherty, M. R., & Thomas, R. P. (2016). Sources of bias in judgment and decision making. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 109–125). Oxford University Press.

Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes*, *104*(2), 207–223.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, *49*(3), 577–585. https://doi.org/10.1037/0022-3514.49.3.577

Vlek, C. (1984). What constitutes "a good decision"? A panel discussion among Ward Edwards, István Kiss, Giandomenico Majone and Masanao Toda. *Acta Psychologica*, *56*(1–3), 5–27.

Wallsten, T. S. (1981). Physician and medical student bias in evaluating diagnostic information. *Medical Decision Making*, *1*(2), 145–164.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140.

Weber, E. U., Goldstein, W. M., & Barlas, S. (1995). And let us not forget memory: The role of memory processes and techniques in the study of judgment and choice. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 32, pp. 33–81). Elsevier.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548–573.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*(5), 806–820.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–94). Hartcourt, Brace and Co. (Original work published 1923)

Wheeler, L., & Miyake, K. (1992). Social comparison in everyday life. *Journal of Personality and Social Psychology*, *62*(5), 760–773. https://doi.org/10.1037/0022-3514.62.5.760

Wilson, T. D. (2002). *Strangers to ourselves*. Harvard University Press.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117–142.

Wilson, T. D., Wheatley, T. P., Kurtz, J. L., Dunn, E. W., & Gilbert, D. T. (2004). When to fire: Anticipatory versus postevent reconstrual of uncontrollable events. *Personality and Social Psychology Bulletin*, *30*(3), 340–351.

Windschitl, P. D., Scherer, A. M., Smith, A. R., & Rose, J. P. (2013). Why so confident? The influence of outcome desirability on selective exposure and likelihood judgment. *Organizational Behavior and Human Decision Processes*, *120*(1), 73–86.

Withycombe, C., & Barreda, V. (2020, December 21). Protesters clash with police, smash glass doors, disrupt lawmakers on "unnerving" day at Oregon State Capitol in Salem. *USA Today*. https://amp.usatoday.com/amp/4003055001

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in syllogistic reasoning. *Journal of Experimental Psychology-General*, *18*, 451–460.

Wyer, R. S., Jr., & Radvansky, G. A. (1999). The comprehension and validation of social information. *Psychological Review*, *106*(1), 89–118.

Zajonc, R. B. (1968). Cognitive theories in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 1, pp. 320–411). Addison-Wesley.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175.

Zajonc, R. B. (1983). Validating the confluence model. *Psychological Bulletin*, *93*(3), 457–480.

Zapf, P. A., Kukucka, J., Kassin, S. M., & Dror, I. E. (2018). Cognitive bias in forensic mental health assessment: Evaluator beliefs about its nature and scope. *Psychology, Public Policy, and Law*, *24*(1), 1–10. https://doi.org/10.1037/law0000153

Zmigrod, L., Rentfrow, P. J., Zmigrod, S., & Robbins, T. W. (2019). Cognitive flexibility and religious disbelief. *Psychological Research*, *83*(8), 1749–1759.