Toward A Cognitive Science Of Beliefs

Joseph Sommer, Julien Musolino, and Pernille Hemmer

Rutgers University, New Brunswick

Abstract

Beliefs are, or at least appear to be, integral to cognition and action. Though there are scarcely features of human psychology more intuitive to their bearers, beliefs are surprisingly elusive targets of study. In this chapter, we consider some perennial questions about beliefs and suggest that some clarity might be achieved by viewing beliefs through the lens of cognitive psychology. We discuss psychological findings and evolutionary considerations which seem to imply that the mind is not designed to form true beliefs, but beliefs that are instrumentally useful. This issue is redolent of debates over whether people are rational or irrational and whether beliefs aim at truth or serve other psychological functions. We survey a series of practical tradeoffs and computational constraints that limit the attainment of true beliefs, and which may be responsible for apparent irrationality. Additionally, the origin of false or irrational-seeming beliefs may be inadequately specified by behavioral data, which implies that a deeper understanding of processes and prior knowledge inside the head is essential for a science of beliefs. We conclude by noting that a view of irrational beliefs as the result of prior knowledge, rather than irrational processes may have optimistic implications for improving people's beliefs.

Keywords: belief; irrationality; instrumental belief; computational constraints; bounded rationality; prior knowledge

1. Introduction

Beliefs play a central role in our lives: They lie at the heart of what makes us human, they shape the organization and functioning of our minds, they define the boundaries of our cultures, and they guide our motivation and behavior. It is no surprise then that belief has been studied in a broad range of disciplines, including anthropology, sociology, political science, economics, philosophy, and psychology. However, trying to determine what beliefs are, or what they ought to be, is far from obvious. For example, the philosophical literature on belief offers a dizzying array of possibilities, including one form or another of representationalism, dispositionalism, interpretationism, functionalism, eliminativism, instrumentalism, atomism, holism, internalism and externalism. The psychology of belief has long history from introspectionist studies (Okabe, 1910) and theorizing (James, 1889), to early AI models (Abelson, 1973; Colby, 1964; 1973), and belief has been studied across numerous psychological domains (see Porot & Mandelbaum, Chapter 2, this volume). Today, it seems that interest is coalescing around belief as a field of its own (Connors & Halligan, 2015). Still, psychology has often struggled to assemble a comprehensive treatment of belief (Egan, 1986). Part of the difficulty may be that the range of phenomena, patterns, and distinctions that seem to fall within the scope of a theory of belief is arresting.

To see this, consider the following observations: (a) Beliefs can have different origins. They can be formed through direct sensory experience, through interactions with others, or via written or audiovisual information channels; (b) Beliefs can be held at different levels of awareness. While some of our beliefs are accessible to conscious inspection, others may not be; (c) Beliefs can be held with different levels of conviction. People can be very confident about some of the things they believe and much less so about others; (d) Beliefs vary in their susceptibility to change. Some beliefs appear to be very stubborn and difficult to change while others seem much easier to modify or even abandon; (e) Beliefs vary in their generality and scope. Some beliefs apply to single objects or individuals while others apply to entire classes of entities; (f) Beliefs vary in the extent to which they guide our behavior. Someone might be afraid of flying and avoid doing so and yet hold the explicit belief that air travel is perfectly safe; (g) Beliefs can produce a range of emotional effects. While some beliefs are benign or even helpful, others can have dire emotional consequences; (h) Beliefs vary in the degree to which they are shared by other people. While some beliefs are idiosyncratic, others can be quite widespread and perhaps even universal.

This diversity has led to disparate conceptions across fields and to literatures that often do not make contact with each other, especially in the context of hyper-specialized academic disciplines. Since complex problems can seldom be comprehensively studied within a single discipline, different aspects of the question of beliefs have been studied by different investigators, using different approaches, and yielding different and sometimes seemingly contradictory results. This fractured picture calls for a systematic effort to integrate these disconnected lines of research and start a broader dialogue on the nature, role, and consequences of beliefs. This is the goal that we set out to achieve in the present volume. Because beliefs are anchored in minds, it is fitting that the integration we are calling for here be conducted within Cognitive Science, the interdisciplinary study of mind.

In this opening chapter, we offer a brief introduction to some of the perennial questions posed by the study of beliefs. In Section 2, we begin by discussing the nature of beliefs and their place in the study of mind. Section 3 asks why biological organisms – human beings at least – should have

beliefs in the first place. Section 4 briefly discusses research on belief formation and updating which seems indicative of irrationality. These findings, as well as evolutionary considerations, have led to proposals about functional roles for beliefs other than aiming at truth. Central to these issues is the question of whether, or the extent to which, human beings can be regarded as rational. In Section 5, we review considerations about epistemic tradeoffs, computational constraints, and the effects of prior knowledge, which suggest that ascriptions of irrationality may be tempered by the complexity of the problems that belief systems face. We close the section with a note on irrationality. Finally, in Section 6, we offer a brief overview of the content of the book.

2. What are beliefs?

The nature of beliefs has been debated for centuries and philosophers have yet to reach a consensus definition (Quilty-Dunn & Mandelbaum, 2018). As mentioned above, some of the difficulty stems from the variety of phenomena that we intuitively label as belief. Beliefs are often unequivocal assertions of flat-out belief (Frankish, 2009), where one either fully believes or disbelieves a proposition (e.g., The sky is blue). On the other hand, they may also vary in the degree of confidence we attach to them (e.g., It may rain tomorrow) (Ramsey, 1926). Additionally, propositions that are actively entertained (e.g., I need to make a right turn on Pine Street now), stored in memory, (e.g., It is safe to walk around my neighborhood at 5:00 p.m.), or generated on the fly (e.g., This guy just tried to cut me off!) all seem to intuitively qualify as beliefs. Even perceptual processes have been argued to be guided by beliefs, qua innate or early developing assumptions about features of the environment that aid in recognizing stimuli or in selecting the most likely percept (Goldman, 1986; Pinker, 1997; Spelke & Kinzler, 2007). To make things worse, people often behave as if they hold beliefs that they never explicitly consider. These implicit beliefs may come in the form of unconscious biases that yield behavior contrary to overtly expressed beliefs. Given this heterogeneous collection of phenomena, it is perhaps not surprising that the ontology of beliefs remains elusive.

For philosophers of a certain persuasion, beliefs do not even exist. Instead, beliefs may be regarded as explanatory fictions that account for observed behavior (Dennett, 2017) or mere linguistic expressions without meaningful referents (Churchland, 1981). Among those who accept the existence of beliefs, opinions regarding their nature also vary. Many argue that beliefs are a kind of "propositional attitude" (Quilty-Dunn & Mandelbaum, 2018) that contains at least two parts: a proposition, *P* (e.g., it is raining), and a mental stance, or attitude, *A*, that one holds with respect to *P*. When someone holds a belief – for example that it is raining – the attitude held towards *P* is its presumed veracity. This is to be contrasted with other attitudes one might hold toward *P*, such as desiring or hoping that *P*. Yet others suggest that belief must be identified with a type of behavioral disposition. One such view is that beliefs are a disposition to behave as if *P* is true or to assert *P* in the appropriate circumstances (Bain, 1872). For example, the belief that it is raining would simply mean possessing a disposition to assert as much or to grab an umbrella before going outside. More elaborate dispositional accounts identify belief with matching a stereotype for believing the proposition in question (Schwitzgebel, 2002). On this view, believing that there is beer in the fridge means possessing a set of dispositions that might include, looking in the fridge when thirsty, offering beer to visitors, not adding beer to one's shopping list, etc. Possessing all or any one of these dispositions is not necessary, as long as one's dispositions are similar to those stereotypically associated with the belief.

It should be noted that the longstanding difficulty in defining belief should not force us to adopt a particular view, nor should it necessarily be regarded as an obstacle to scientific investigation. As Fodor wrote, "pre-theoretically, we identify mental events by reference to clear cases. Post-theoretically, it is sufficient to identify them as those which fall under psychological laws" (1975, p. 4, fn 2; see also Fodor, 1968, pp. 10-11, 143). In other words, our pre-theoretical notions of beliefs are sufficient to focus our attention on certain questions and guide investigation, even if scientific notions of beliefs end up departing from these pre-theoretical conceptions once explanatory theories become available (see Stich, 1983).[1]

In gathering chapters for this book, we chose an inclusive definition of beliefs so as not to artificially impose a priori limitations on the scope of a more mature scientific theory of beliefs. Chapters in the present volume discuss representations ranging from explicit propositions to tacit Bayesian priors. As such, our broad definition includes any stored or generated information that is used to aid perception, action, or cognition.[2] We should be pleased to discover that this definition is in error, as this will mean that the science of belief has progressed toward sharper empirical distinctions. There are other reasons to adopt a broad view. A mature science of beliefs should not merely delimit the boundaries of the phenomenon; it should also be concerned with antecedents and consequences of beliefs. For example, if implicit representations or tacit priors are not deemed to be proper beliefs after all, they may still be of interest for their capacity to generate beliefs or assist them in guiding action (cf. Stich, 1978).

3. A cognitive perspective on beliefs

Though there is presently little agreement on a definition of beliefs, some aspects of the complex nature of beliefs might be better understood by examining beliefs through the lens of cognitive psychology. A cognitive approach raises the prospect that beliefs might differ due to the interaction of multiple psychological processes with computational principles underlying belief formation and storage. This perspective may help explain some of the complexity underlying beliefs discussed in the previous section.

Perhaps the obvious place to begin an analysis of the psychology of belief is with the question of what benefits beliefs confer on a cognitive system. The primary advantage of having beliefs seems to be the ability to construct and reason about internal representations of the external world instead of being forced to learn from direct experience (Campbell, 1974; Dennett, 2008). For example, thinking counterfactually (Evans & Over, 2004) and simulating outcomes can help us construct better plans and select adaptive actions. However, these abilities do not seem to account for the

---

[1] It is worth noting that our naïve intuitions about a broad range of natural phenomena have provided a very useful scaffolding for the development of scientific theories. However, once scientific concepts mature, they all too often become perplexing to our intuitions, an important point already noted by Hume. Commenting on Newton's ideas and the demise of the intuitive mechanical philosophy of the Cartesians, Hume (1778, p. 542) pointed out that "While Newton seemed to draw off the veil from some of the mysteries of nature, he showed at the same time the imperfections of the mechanical philosophy, so agreeable to the natural vanity and curiosity of men; and thereby restored her ultimate secrets to that obscurity, in which they ever did and ever will remain."

[2] A view of belief that is not explicitly propositional may account for animals' and pre-linguistic children's competencies (Churchland & Churchland, 2013). Note, however, that having language may not be necessary for the ability to express propositions in a "language of thought" (see e.g., Fodor, 1986, p. 19).

quantity of diverse beliefs that people come to hold, many of which are not readily associated with actions (Abelson, 1986).

One reason for possessing many beliefs is suggested by Newell's (1994) preparation-deliberation tradeoff: when one is faced with a problem, there are two broad options, either come prepared in advance (because say, the problem was encountered in the past), or calculate a solution. Newell (1994; see pp. 102-107) observed that two systems can achieve equivalent performance if one stores more knowledge and the other spends more time calculating. However, within a single system, deliberation cannot be easily improved because of fixed computational capacities like processing speed, but more knowledge can usually be stored (see also Sperber & Wilson, 1996, p. 47). This means that most improvements within a given system will come from learning and encoding new knowledge and procedures. This is especially true for an organism facing real-time constraints on decision-making and action, as computation is likely to take longer than accessing memory. We might therefore expect people to possess a large body of stored beliefs to reduce computational time and effort.

The preparation-deliberation tradeoff would seem to imply that larger the set of stored beliefs the better. However, there are also reasons to limit storage. An agent with a large set of *explicit* beliefs can face severe problems in updating (Janlert, 1987). Upon encountering new information, the agent will need to iterate through its beliefs, determine which are inferentially impacted by the new evidence, and revise each one. As the number of explicit beliefs grows, this process becomes computationally expensive and eventually intractable. Say an agent has a representation of several items on a table and explicitly stores beliefs about their directional relationships, e.g., a book is to the left of a glass and to the left of a pen, etc. If the book is moved to the right of the table, each of these explicitly stored relationships will become obsolete and require correction. In contrast, if directional relationships are left implicit, they can be calculated as needed, removing the need to keep the representations current (Bobrow, 1975).

Similarly, if many beliefs are generated from a few core beliefs, altering just one of these will de facto update all the beliefs it composes. For example, if one learns that conservative politicians are opposed to large government, rather than applying this new belief to every known conservative, only a single belief about conservative positions needs to be updated. This belief can generate the information for application to any specific politician (assuming this knowledge is later activated). This approach also gracefully deals with learning about new conservative politicians. If the belief about opposition to large government is explicitly applied to each known conservative politician, it may not be applied when a new conservative is encountered. On the other hand, generating this information from a single belief about conservatives can flexibly apply it to new cases (Woods, 1975, p. 73). Implicit beliefs may represent exactly this solution, as they can be inferred from other knowledge and do not need to be stored in explicit form (cf. Gallistel & King, 2009, pp. 58, 208). Explicitly storing only a subset of possible beliefs may make updating more tractable (Sperber & Wilson, 1996, p. 85).

Which beliefs should be stored then? Those that can generate many other beliefs are good candidates for storage. In other words, one might store premises rather than conclusions. Additionally, stored conclusions may cause problems if the premises that generated them are later revised, leaving the false conclusions preserved in memory. For example, say conclusion q is initially inferred from premise p, but at a later time, p changes such that it no longer implies q. If q was committed to memory initially, it may persist later, even if it does not follow from p at that

later time. However, there may be occasions where storing a conclusion is worthwhile. If a conclusion is derived from disparate sources of knowledge which might not be assembled in working memory on later occasions, it may be worth committing to memory (see Sperber & Wilson, 1996, pp. 106-107). For example, if premises *p1*, *p2*, *p3*, and *p4* are all required to infer *q* and are not likely to simultaneously come to mind, one might want to store conclusion *q*. Such conclusions may persist in memory after the premises that generated them have been forgotten, leaving burned-in (Doyle, 1979) beliefs that are held for reasons that have been lost (see also Harman, 1986, pp. 41-42). Additionally, beliefs might be stored in accordance with their use, with a preference for frequently relevant beliefs. This may be achieved for free because commonly referenced beliefs will present themselves to memory more often, increasing their probability of storage (Sperber & Wilson, 1996, p. 77). Belief usage might also be expected to correlate with their generality, as more general beliefs will tend to be relevant across a wider range of situations.

In the case of extremely common situations, beliefs may become overlearned, receding out of conscious awareness, as do most habits. This phenomenon is analogous to the "chunking" of behavior (see Simon, 1996, p. 90) which allows frequent behaviors to be compiled and activated by environmental cues (cf. Schneider, Dumais, & Shiffrin, 1982). For example, routines like tying one's shoes are laborious and explicit when first learned, but become automatic with practice, perhaps in a process akin to running a program through an optimizing compiler (Pylyshyn, 1986; see also Abelson, 1973, p. 310).

Chunked inferences may underlie Frankish's (2004) concept of *implicit belief*. Frankish makes an important distinction between two closely related notions that he calls *tacitly* and *implicitly* accepted beliefs. *Tacit beliefs* are those that can be generated from prior knowledge even if they have never been explicitly considered, e.g., the proposition "1,013 is less than 8,927," which would likely be accepted as a belief by someone presented with it for the first time. In contrast, *implicit beliefs* serve as chunked background inferences. For example, when thinking about going out for dinner at night, a person familiar with their neighborhood can assume that it is safe to do so without consciously considering the issue – a compiled inference. However, while away on vacation or in an unfamiliar location, the safety of the neighborhood may not be taken for granted. In this case, more explicit reflection may be necessary.

The picture that emerges from these considerations is that an agent may be expected to have a set of overlearned and chunked beliefs that remain implicit and guide behavior in extremely common situations. For less recurrent events, explicit beliefs are predicted to be those that cannot be easily generated from prior knowledge, that generate inferences themselves, and that are applicable across situations. For rare problems, tacit beliefs may be freely generated and then forgotten to avoid problems of too many explicit beliefs. If this emerging picture holds, the diverse nature of beliefs might be due to the variety of problems faced by the system along with principles of knowledge organization such as compiling common inferences into chunks and avoiding unnecessary explicit storage. In a similar vein, which cognitive processes are appealed to, e.g., whether one is referring to memory retrieval or the generation of novel inferences, may yield different beliefs with different qualities, making a single definition difficult to find.

4. The mechanics and functions of beliefs

If a cognitive approach sheds light on the complexity of beliefs, the next question that may be asked is how should these beliefs work? A reasonable assumption might be that beliefs are meant to accurately represent the world in order to appropriately guide adaptive behaviors (Fodor, 2000, pp. 66-68). However, much of the psychological literature on beliefs has yielded results that appear to conflict with this assumption. The belief literature has largely been concerned with questions of how people search for and process evidence, whether reasoning is motivated, and whether belief updating is proportional to the strength of the evidence.[3] Many findings in this literature are suggestive of biased search, processing, and assimilation of evidence (for reviews, see Kunda, 1990; Nisbett & Ross, 1980), which raises further questions about whether humans are rational or irrational. In light of the prevalence of apparently irrational behavior, some have proposed theories suggesting that the purpose of beliefs is not to form an accurate model of the world. These include several instrumental or functional notions of belief according to which beliefs serve personal and social goals which may not be truth oriented. In this section, we briefly discuss the mechanics of belief updating, empirical evidence suggestive of irrationality, as well as the evolutionary case for functional theories of belief.

4.1 The mechanics of beliefs

The mechanics of belief includes how beliefs are updated, how people search for new evidence, as well as how they reason about evidence that supports or opposes their prior beliefs. We first consider how beliefs are updated. Beliefs should, of course, be sensitive to evidence. The normative model for updating a hypothesis upon receiving new evidence is Bayes' rule. Given some prior hypothesis H and new evidence E, Bayes' rule computes how much more likely H is to be true after learning of E. Say the hypothesis is that it rained the previous night, and the evidence is that the ground is wet. Belief in H will depend on three factors. First is the *prior probability* of H – if the setting is a desert, it is a priori less likely to rain than in a forest. Next is the *likelihood* of the evidence if H is true. Given that it rained, how likely is it that the ground would be wet? This may be quite likely, but not certain, as the rain could have been blocked by a barrier. Finally, evidence is less diagnostic when E is expected regardless of whether H is true. For example, if someone nearby is watering the lawn, the ground could be wet even if it had not rained.

Formally, Bayes' rule calculates the *posterior probability* that a hypothesis is true given the evidence, denoted p(H|E). The prior probability of H is represented as p(H); the likelihood of the data *given* that H is true, as p(E|H); and the probability of E is p(E). Bayes' rule can be given as:

$$p(H \mid E) = \frac{p(E \mid H) * p(H)}{p(E)} \quad (1)$$

Given the axiomatic rationality of Bayesian belief updating (Cox, 1961; de Finetti, 1970/1974), as well as the remarkable match between human behavior and the predictions of Bayesian models

[3] In the remainder of the chapter, we refer to these processes collectively as belief processes.

across cognitive domains (Anderson, 1990; see Chater et al., 2010 for a review), it might be expected that human belief updating should also be governed by Bayes' rule.

It is important to note that most Bayesians are not committed to the brain implementing Bayes' rule at Marr's algorithmic level (Anderson, 1990; Griffiths, et al., 2012), a level of analysis which attempts to specify the nature of the algorithm employed by the cognitive system (Marr, 1982). Instead, the assumption is that Bayesian explanations operate at Marr's computational theory level, which asks about the goals of the system and attempts to explicate the problem that the system is trying to solve. The algorithms underlying human behavior need not be explicitly Bayesian, but could approximate the normative Bayesian solution (Anderson, 1990; Chater et al., 2020).

In contrast to the theoretical rigor of Bayes, decades of findings suggest that human beings do not update their beliefs rationally. People selectively search for evidence that supports their previously held beliefs (Snyder & Swann, 1978; Wason & Johnson-Laird, 1972). They ignore evidence that should be relevant (Smedslund, 1963), and when presented with information that runs counter to their beliefs, people often engage in motivated reasoning to selectively disconfirm the incoming evidence (Kunda, 1990). Additionally, when people are exposed to mixed evidence supporting two sides of an issue, they show biased assimilation of evidence, accepting the evidence that supports their views and disregarding opposing information (Lord, Ross, & Lepper, 1979; Taber & Lodge, 2006).

Even in the face of apparently overwhelming evidence, people sometimes continue to believe (Anderson, Ross, & Lepper, 1980; Ross, Lepper, & Hubbard, 1975). In "belief perseverance" studies, participants are often presented with false feedback, suggesting they have performed poorly on a task (e.g., distinguishing between real and fake suicide notes). They are then let in on the initial deception and told that the feedback they received was fabricated. In spite of that, participants often cling to the belief that the feedback was an accurate reflection of their ability.

Findings such as these have led many to conclude that humans are fundamentally irrational (but see Anderson, 1990; Gigerenzer, 2000; Oaksford & Chater, 2007 for different conclusions). Others have even argued that these findings indicate that beliefs are not meant to accurately reflect the world, but are better understood as serving some instrumental, rather than epistemic function.

4.2 The functions of beliefs

If the purpose of beliefs is not to form an accurate model of the world, then to echo Smith, Bruner, and White (1956, p. 1), "Of what use to a man [sic] are his opinions?" It is beyond the scope of this chapter to fully survey the approaches dedicated to answering this question, but we present a brief overview below.

From an evolutionary perspective, beliefs matter indirectly and only insofar as they are tied to actions. Evolution cannot select based on an organism's internal states, but only on their resultant behaviors. If a true belief does not result in adaptive actions, it offers no advantage to its holder. Likewise, if an irrational belief reliably leads to actions that increase evolutionary fitness, it may be selected for, regardless of its truth value (but see Fodor, 2000, pp. 66-68 for an opposing argument).

There are various ways in which the evolutionary benefits of an action may be decoupled from a belief's correspondence to reality. For example, if one has unjustifiably high self-confidence, this may result in increased mating opportunities. If everyone in a community, or a particularly prestigious group member, has arrived at an incorrect belief, it may be more advantageous to conform than to dissent. As Goldman (1986, p. 98) observes, the opposite is also true: there are true beliefs that are not necessarily aligned with survival, such as an accurate understanding of nuclear physics, which can be used to create weapons of mass destruction. Therefore, in sharp contrast to what Kitcher (1998) calls Huxley's Credo, namely that "truth is better than much profit," Wilson (2010) argues that epistemic rationality is only beneficial to the extent that it helps achieve instrumental goals.

Beyond the evolutionary argument, there are at least three other reasons to think that beliefs may have functions that are not epistemic. First is the apparent belief-action gap, (Wicker, 1969; see Zanna, Higgins, & Herman, 1982). Numerous studies have found that there is a surprisingly low correlation between expressed beliefs and actions. These findings led Abelson (1972) to remark that "we are very well trained and very good at finding reasons for what we do, but not very good at doing what we find reasons for" (p. 25).[4] If beliefs do not guide actions, perhaps they serve some other purpose, such as post hoc justifications that paint one's actions in a positive light (Haidt, 2001). Second are the empirical findings, mentioned above, documenting biases in people's evidence search and evaluation, failures of reasoning, and use of heuristics and cognitive biases. People often form bizarre and seemingly irrational beliefs, including beliefs that are apparently inconsistent with each other (Wood & Sutton, 2012). Additionally, people are intransigent when confronted with evidence that opposes their views (Ross, Lepper, & Hubbard, 1975) and possess reasoning strategies that often lead to deviations from normative procedures (Kuhn, 1991) and results (e.g., Tversky & Kahneman, 1974). Third, though perhaps weakest, is the existence of a wide range of beliefs on virtually any topic imaginable, from the esoteric (e.g., Is the Many-Worlds interpretation of Quantum Mechanics correct?) to the much more mundane (Is the Earth round?). If the point of having beliefs is to form approximately accurate models of the world, why do people so often appear to fundamentally disagree or to come to such different conclusions?

If beliefs are not held because of their truth value, what other purposes might they serve? Functional approaches to attitudes (of which beliefs were traditionally thought to be a component) posit that attitudes serve psychological needs. Social psychologists historically categorized theories into several functional proposals (Katz, 1960; Smith, Bruner, & White, 1956). According to Knowledge or Object Appraisal theories, people have a need to categorize, or attribute meaning to, information without spending too much time integrating evidence. Ego-Defensive theories

[4] Since Wicker's (1969) influential paper, there have been strong responses to the belief-action gap. Among these are Fishbein and Ajzen (1977), who find that the gap largely disappears if measures of belief and behavior are taken at appropriate levels of analysis. These may be illustrated by their principles of compatibility and aggregation. The former notes that behaviors consist of *actions*, *targets*, *locations*, and *times*. The general belief that one likes going to the movies will not predict attendance at film X in theater Y at time Z, but a specific intention is likely to be more predictive. The latter principle suggests that to predict behavior from general beliefs, the measure of behavior must be correspondingly general. For example, predicting voting from the belief that "voting is important" should entail a repeated measure of voting across many elections.

Additionally, behavior is rarely guided by single beliefs, but usually depends on many conditions being met simultaneously, as has been long-recognized in philosophy (Braithewaite, 1932). These conditions are likely dependent on personal construals of the situation and are difficult to recognize from the outside (Salancik, 1982). See Grandin, Boon-Falleur, & Chevallier (Chapter 26, this volume) for more on the belief-action gap.

propose that information can be threatening to self-esteem and that people may form beliefs to protect themselves. Utilitarian theories suggest that people form beliefs to maximize reward and minimize punishment. Social Adjustive theories view attitudes as oriented toward conformity with a social group, e.g., one might adopt beliefs that are similar to those of friends or family members (see, Kahan, 2017; Williams, 2020). Value Expressive theories rest on the idea that attitudes may serve as a public signal of one's social or political allegiance (e.g., Simler & Hanson, 2017). This category may include proposals that beliefs can function as a form of strategic self-deception, which would allow people to better convince others of otherwise dubious claims by first convincing themselves that the claims are true (Trivers, 2011).

Still other treatments assume that people have preferences over their beliefs (Caplan, 2001; 2011) or that beliefs have direct utility for those who hold them (Molnar & Loewenstein, Chapter 14, this volume). These positions are similar to Abelson's (1986) suggestion that beliefs are like possessions, which may differ in value to their holders, just the way material possessions do.

In sum, there appear to be evolutionary arguments for processes of belief formation that are not concerned with truth except as a means to an instrumental end.[5] Additionally, people's empirical belief updating and reasoning abilities seem difficult to understand as truth-oriented. This might suggest that beliefs are sensitive to instrumental factors. However, forming and updating beliefs may be more difficult than it appears.

5. The practical difficulty of belief

In the sections 2 and 3, it became apparent that beliefs are complex and that this complexity may stem from the variety of problems people face and computational considerations governing storage and updating. Section 4 illustrated apparent departures from optimality, which might imply that people are irrational. Alternatively, as we have seen, one can posit that beliefs serve functional needs other than truth. If both of these conclusions are unpalatable, one can attempt to demonstrate that the irrationality is only apparent. In fact, there are several explanations that can account for irrational looking beliefs, indicating that a retreat to functional explanations may be premature.

Aside from functional considerations, beliefs and the behaviors they entail may appear to deviate from rationality for three reasons. First, even belief processes with the goal of arriving at true beliefs may be forced into epistemic tradeoffs that compromise accuracy. Second, some problems present such extreme computational constraints that forming optimal beliefs is impossible regardless of the goals of belief processes or the tradeoffs they implement. Finally, rational processes influenced by incorrect prior knowledge can yield false beliefs. In this section, we highlight the difficulty of arriving at true beliefs given these epistemic tradeoffs, computational constraints, and effects of prior knowledge.

[5] Note that Fodor (1992) distinguishes between beliefs and belief-forming *processes* having functions. It might be that only processes have functions (e.g., of producing true beliefs, ego-protecting beliefs, etc.), but that beliefs themselves are often functionless. For example, the belief that one has not recently painted a picture may not have its own function, even if the inferential process that creates it has the function of deriving novel true beliefs from stored knowledge (see pp. 65-66, but see also p. 129).

5.1 Epistemic tradeoffs

When people form beliefs, they should be sensitive to changing facts of a given situation (Pinker, 2010). For this reason, flexible processes[6] that can readily acquire new beliefs should be selected for. While one might assume that these processes can maximize for truth, or at least for evidence or justification, this may not be practically achievable. Critically, even processes with the goal of attaining true beliefs might be forced into trading off justification for other epistemic values, with the consequence of forming some inaccurate beliefs. As De Sousa (1971) noted, truth is unlikely to be the sole consideration underlying belief formation, as "we would rather have a largish set of propositions labelled 'true' – even though their title to that label may not be impeccable – than a very small set labelled 'absolutely certain' and an immense, unwieldy set consisting of all other propositions" (p. 56).

Though possessing only beliefs that are certain to be true might seem ideal, in the real world, even truth-seeking belief processes may be expected to trade off some measure of validation for speed, power, and reliability (Goldman, 1986, pp. 26-27). Speed is intuitive enough – a process that arrives at new true beliefs at a rate of one every ten years is unlikely to be selected for. Of course, as Fodor (1987, p. 140) put it, "the price of economy is warrant." In other words, the faster beliefs are obtained, the less justified they will be. A process's power is its capacity to produce many true beliefs, while its reliability is given by the ratio of true to false beliefs it achieves. A powerful but unreliable process could cause one to believe every possible proposition, thus learning every available truth – but also every possible falsehood. By contrast, a process with low power but high reliability might yield so few beliefs that it would benefit from accepting new beliefs even if some are likely to be false. This may be viewed as a tradeoff between minimizing *errors*, or mistaken beliefs, as opposed to *ignorance* – the absence of true beliefs (Goldman, 1986, p. 26). A powerful process with low reliability will gain some knowledge at the cost of accruing many errors, while a reliable process with low power will remain ignorant, though it will produce few errors. Some compromise that puts reasonable limits on errors while allowing some knowledge to be accumulated is better than setting a threshold for error so high that nothing is learned.

In sum, it is likely that even if the processes involved in belief formation and updating were selected for epistemic goals, practical limitations preclude achieving completely justified beliefs. Arriving at some false or unwarranted beliefs might not be the result of selection for functional beliefs, but for processes that actually function under limiting factors like time pressure. If this is the case, apparent irrationality may be understood as the result of necessary tradeoffs between different epistemic virtues, such as reliability and speed.

5.2 Computational constraints

In Section 5.1, we observed that under real world conditions, belief processes may have to compromise among epistemic goals to attain fairly accurate beliefs in a reasonable amount of time. The implication is that with more time or effort, better beliefs might be achieved. However, as we

[6] There is much debate in cognitive science over whether evolution selects for processes or content (Fodor, 1975). In domains such as intuitive ontological beliefs (e.g., folk physics, biology, psychology, etc.), some content may be innately built into the organism (Pinker, 1997; Spelke & Kinzler, 2007; see also Shtulman, Chapter 15, this volume).

will see in this section, the problems involved in belief formation and updating are so difficult that no amount of time or effort can be traded off to acquire true beliefs.

The distinction between epistemic tradeoffs and computational constraints is exemplified by a difference between the heuristics and biases tradition in decision-making and Herbert Simon's notion of bounded rationality. Kahneman and Tversky (1974) suggested that people take the easy way out of an "effort-accuracy tradeoff," and rely on heuristics despite problems being solvable through more effortful explicit reasoning. Simon (1955) had a very different class of problems in mind: those that cannot be solved by computation within a human lifetime. The only recourse is to find some heuristic that approximates a solution (see Faust, 1984, p. 106 for a similar distinction). An example of an effort-accuracy tradeoff might be assuming a news headline is roughly accurate without reading the whole article. While a more careful reading might catch a few inaccuracies, it may not be worth the effort. A much more difficult problem might be achieving a complete understanding of a complex matter in foreign policy, which could require a lifetime of research to learn all the relevant history and important facts. Here, it is necessary to aim for a lower standard of understanding if the goal is to be achieved. Belief formation and updating often involve problems of this second kind. Under these circumstances, the computational constraints are so severe that reliably coming to true beliefs is not possible.

A primary constraint on belief processes is the issue of computational intractability. First, Bayesian updating is computationally intractable in the real world (Griffiths et al., 2010). Second, evidence search can, in principle, never end – one can always look for more information, and no set of confirmatory observations suffices to establish a fact. This is the classic problem of induction. That the sun has risen every day for billions of years does not establish that it will rise forever, as indeed it will not. Third, even maintaining consistency among a large set of beliefs is intractable (cf. Sommer, Musolino, & Hemmer, in preparation). If any subset of one's beliefs may have inferential relevance for any other subset, the ensuing combinatorial explosion makes maintaining consistency intractable. As Cherniak (1984) points out, checking a list of only 138 logically independent propositions for consistency at a rate of one line per "the time a light ray takes to traverse the diameter of a proton" will take twenty billion years!

Perhaps, then, we should not be surprised that humans are not deductively closed (Nozick, 1983).[7] People are not capable of inferring the consequences of all their beliefs. Just as all subsets of one's beliefs cannot be checked for consistency, they cannot be examined for their inferential relevance to each other. This is why Eratosthenes' calculation of the circumference of the Earth from the angle of a shadow and the distance between two cities is considered an extraordinary feat rather than a trivial consequence of knowing geometry. If people were deductively closed, this calculation would be a natural occurrence upon learning all the relevant information; instead, it is a feat still remarked on 2,000 years later. The impossibility of closure is only exacerbated if it includes generating all possible implicit beliefs and then computing their consequences as well (Woods, 1975, p. 48). Incidentally, Goldman (1986) notes that if belief-fixing processes are imperfect, such that they allow some erroneous beliefs to be adopted, closure is not actually desirable, as it would result in the deduction of many new beliefs from any false premises. Instead

[7] Note that if it is surprising, this is because despite prior experiences of missing the implications of one's beliefs, the inference to lack of deductive closure is not readily made.

of closure, people are likely restricted to reasoning from a few "salient beliefs" activated in working memory (Fishbein & Ajzen, 1977; Holland et al., 1986).

In sum, evidence search and reasoning can never be considered complete. Beliefs may have obvious implications that go unnoticed by their holders. This may even include situations where people have beliefs that conflict with each other. Note that these problems require heuristic solutions. No amount of effort suffices to form a perfectly justified belief (barring cases like logical deduction). Optimizing for truth is a wonderful ideal, but as Simon (1997, p. 154) wrote, "If you let me determine the constraints, I don't care who selects the optimization criterion." In other words, problem constraints often have more of a say than one's goals in determining outcomes. Even if the purpose of belief processes is to acquire true beliefs, the constraints they face are so difficult that they may fail to meet this goal.

5.3 Effects of prior knowledge

Section 5.1 suggested that people may arrive at inaccurate beliefs because their belief processes trade off some reliability for faster or more powerful belief acquisition. Section 5.2 noted that false beliefs may also result from processes that implement fallible heuristic solutions to computationally intractable problems. In both cases, the implication is that irrational beliefs are the result of limitations of belief processes, even if these limits are caused by understandable tradeoffs and constraints. However, a person may come to possess inaccurate beliefs not because their belief processes are taking necessary shortcuts, but because these processes are provided with incorrect data. For example, given a set of false prior beliefs, rational processes will not result in true conclusions. From the axioms, "water contains arsenic" and "arsenic is toxic," the conclusion that "water is toxic," is valid, but logical *validity* is no guarantee of *soundness*. The problem here is not in the inferential process, but in the set of beliefs.

Interpreting behavior as irrational often assumes a defect in the processes of belief acquisition and revision. As discussed above (section 4.1), these processes are not limited to Bayesian updating and include retrieval of relevant beliefs, inferential methods for recombining beliefs, evidence search and evaluation procedures, and reasoning strategies. For example, motivated reasoning is often portrayed as a flawed process. However, Kunda (1990) suggests that it might be better understood as directed by expectations based on prior knowledge[8] rather than motivation per se (see also Henle, 1955; Henle & Michael, 1956). Indeed, it may be reasonable to defend a strong belief more effortfully than a weak belief (Koehler, 1993). Likewise, in many environments, it is rational to search for evidence that is consistent with beliefs one already holds (Klayman and Ha, 1987). Certainly, if one assumes that most beliefs (e.g., "the sky is blue," "water is safe to drink," etc.) are true, attempted disconfirmation would be a large computational cost with little prospect of gain. Along these lines, Pennycook and Rand (2021) found that political partisans often have large differences in their factual beliefs. When these differences are controlled for, they account for much of what might otherwise be attributed to motivated partisanship (Tappin, Pennycook, & Rand, 2020).

[8] The term prior knowledge is used here as a synonym for beliefs, it is not used in the philosophical sense of justified true beliefs. For an attempt to distinguish between knowledge and belief in psychology, see Abelson (1979).

In a similar vein, a Bayesian updating process operating on incorrect beliefs may produce results identical to an irrational process operating over accurate beliefs (see Hastie & Dawes, 2010, pp. 188-189). As Anderson observes, "it is entirely in the spirit of a Bayesian analysis to have priors obscure the effect of regularities in the data" (1990, p. 132). Bayes' rule is tautologically true if its inputs are true statistics of the environment and valid likelihoods. However, human psychology bears little resemblance to this ideal case: priors (Feldman, 2017) and likelihoods (Fishbein & Ajzen, 1977, p. 188) are both subjective. Without knowledge of a person's beliefs, Bayes is of little help in understanding their updating. The ideal experiment instructs participants to ignore all their prior knowledge, to accept all information provided, and then measures updating. In reality, people will dismiss evidence that is provided to them, counterargue for why it is unlikely, or discount it based on other beliefs. Of course, if any of these processes were flawed, similar results could obtain, but this inference is not warranted solely by an apparent failure to update. As Simon (1980, p. 36) remarked, "It is a common experience in experimental psychology… to discover that we are studying sociology – the effects of the past histories of our subjects – when we think we are studying physiology – the effects of properties of the human nervous system."

The role of prior knowledge may often be overlooked in psychology due to the assumptions underlying certain approaches to studying the mind. To understand the effects of prior knowledge on belief processes, it is helpful to examine the assumptions made at Marr's (1982) computational theory and algorithmic levels of analysis (see section 4.1). Studies at the computational theory level exploit the fact that successful behavior can be informative about the problem the cognitive system is trying to solve while abstracting away from details about cognitive processes (see Newell, 1994, pp. 150-152; Simon, 1996, pp. 7-11). For example, if the task is to multiply 3 by 5, the same answer will result from taking three fives, five threes, retrieving the answer from memory, asking a friend, etc. Reporting the correct answer will rarely indicate which algorithm generated the response. Rather, it is informative about the problem the cognitive system is trying to solve. In effect, a problem acts as a filter for processes that can solve it: any process that arrives at an incorrect solution will be filtered out, but many different processes might succeed. These successful processes may operate in different ways, but their outputs will resemble each other precisely because they fit through the filter (cf. Fodor, 1992, p. 100). Looking at outputs alone will not disambiguate which process produced them, but if the outputs solve the problem, they are indicative of the shape of the filter – i.e., of what the problem requires of its solution.

The success of the Bayesian approach in cognitive science stems from its ability to explain behavior by specifying the nature of the problems that the mind must solve. If a task can be formulated as a Bayesian problem and people solve it, behavior will look Bayesian, because successful performance conforms to the shape of the problem. Though this does not individuate the exact process that was used, as multiple processes may achieve the same results, it is evidence for the Bayesian characterization of the problem. As Bayesian researchers often take pains to point out (e.g., Anderson, 1990, p. 251; Oaksford & Chater, 2007, pp. 263-265), the conclusion to be drawn from successful performance is not that the mind is employing a Bayesian algorithm. Instead, success is evidence that the Bayesian characterization of the problem is what the system was evolutionarily selected to solve. For example, a naïve view of memory is that recall of all experienced events should be perfect. When the task of the memory system is so characterized, performance seems poor. However, an alternative view of memory is more akin to a search engine, which must retrieve the most relevant memories from a massive database of stored information. Additionally, it must do so given only noisy cues about what is needed from the current

environment (Anderson, 1990, p. 42). If human memory approximates optimal performance on this task, that is evidence that the noisy retrieval problem is what the system is designed to solve (cf. Zhang, Chapter 4, this volume). Many algorithms might produce equivalent performance,[9] but the system can be usefully understood at the computational theory level without detailing specific algorithms.

The same principle is also the strength of Dennett's (1971) intentional stance and Newell's (1982) knowledge level, which allow effective prediction of others' behavior from their beliefs and desires, without requiring a deep understanding of psychological mechanisms. One can assume that other people's behavior will rationally adapt to their goals and, to the extent that they succeed, this will obscure the details of their psychological processes. For example, if someone's goal is to buy groceries, our predictions can ignore details about their visual perception or memory capacity. Indeed, part of their rational goal pursuit may involve putting on glasses and writing down a list, thereby ensuring that limitations of vision or memory are not discernable in their behavior.[10] However, if psychological limitations cause the person to fail, this offers evidence about the nature of their psychological processes (Newell & Simon, 1972, p. 55).[11] If our shopper forgot to add a crucial item to the list or misread a product label, appeals to memory or visual processing limitations become more explanatory than knowledge of their goals.

Because failures can convey information about processes, studies designed to find evidence about processes at Marr's algorithmic level often induce the cognitive system to fail at a given task. If failure is induced by imposing a time limit, this suggests limits of processing speed; if too much information is presented for the memory system to keep up, this is informative about the system's (short term) storage capacity. Tversky and Kahneman's (1974) heuristics and biases program may be the most famous example of this approach, highlighting the heuristics people use by demonstrating how and when they fail. For example, people overestimate the frequency of uncommon risks that are highly salient, such as plane crashes. This error suggests that likelihood estimation uses a heuristic which judges probability based on how easily events come to mind (Tversky & Kahneman, 1974).

In short, the computational theory level abstracts over successful algorithms that produce similar adaptive behavior on a given problem. The algorithmic level attempts to differentiate algorithms based on their idiosyncratic failure conditions. Neither of these approaches explicitly accounts for prior knowledge. The computational theory level effectively filters for accurate knowledge as it filters for algorithms that can succeed at its specified problems. At the algorithmic level, failure is used to individuate processes, so when a failure is observed, it is often taken as information about a cognitive process. In both cases, incorrect knowledge may be overlooked.

---

[9] This is part of the reason that the Bayesian approach is openly agnostic about and even skeptical of the possibility of isolating cognitive algorithms (Anderson, 1978; 1990, p. 24).

[10] Similarly, evolutionary explanations can abstract over biological details on the assumption that adaptations will serve their bearers well in a given environment (Simon, 1996, p. 7). Just as adaptive behavior conforms to the shape of the problem, evolution molds organisms to the shape of their niche. For example, the prediction that a polar bear's white coat is useful camouflage does not need to refer to the biochemical processes governing fur color. However, if a genetic mutation caused blotchy fur and made bears easy prey, adaptive explanation would fail, and a lower-level account would be required. Likewise, convergent evolution arrives at the same solutions to a problem via different biological structures (Fodor, 1986, p. 81).

[11] Kurt Lewin made many of these same observations in 1936 (pp. 26-27, 148).

For many domains of cognition, ignoring prior knowledge is not an issue. In much of cognitive science, successes and failures can reliably be attributed to processes, and not to prior knowledge, for (at least) three reasons. First, because the processes under examination are closer to the cognitive architecture (e.g., edge detection in visual perception) and therefore make use of less knowledge (Newell, 1994); second, because knowledge differences in the domain are roughly uniform (e.g., knowledge of language); third, because any knowledge differences are eliminated with meaningless stimuli such as nonsense syllables in memory studies (see Bartlett, 1932, pp. 3-6 for an early criticism of this practice). However, if naturalistic stimuli are used in the study of higher cognitive domains, such as belief, results are less straightforwardly the outcome of processes alone. Beliefs are (usually; see footnote 6) formed out of the interaction between cognitive processes and prior knowledge.[12] This knowledge includes stored beliefs and their generated inferences as well as reasoning strategies, or what Stanovich (2011), following Perkins (1995), calls "mindware" (e.g., probabilistic reasoning). When prior knowledge is accurate or when no prior knowledge is required to solve a problem, observations are tests of the system's processes. However, in a system with some false beliefs, as more beliefs become relevant to a problem, false beliefs will come to dominate performance.

For example, imagine two people with identical belief processes. Person A is presented correct information while Person B is lied to and provided with propaganda. A will succeed in forming correct beliefs while B will fail via the same processes. If confronted with information that contradicts their views, they may both engage a selective search for arguments to bolster their opinions. This may be rational for each of them, given the strong belief that they are correct and that such arguments can be found. However, assuming they are both able to preserve their prior beliefs, they will remain polarized. A less neutral observer might say that A followed a healthy media diet, came to true beliefs, and successfully debunked fake news, while B bought into propaganda and used motivated reasoning to persist in believing falsehoods. To the extent that this characterization is accurate, it is due to differences in knowledge, not belief processes.

In conclusion, higher cognitive processes such as those involved in belief seem to be strongly influenced by prior knowledge as opposed to new evidence alone. This presents another explanation for people arriving at irrational beliefs: through no error or compromise of belief processes, false beliefs may still be adopted if they are consistent with erroneous existing beliefs. If, as we have seen, human belief processes are subject to tradeoffs and constraints, and are strongly influenced by prior knowledge, achievable performance will be limited by these factors. In light of similar considerations, Nisbett and Ross (1980) argued that proposing novel explanations for irrationality may be "parsimonious only to the extent that normal or unimpassioned intellectual functioning is held to approach perfect rationality... Given [human intellectual] shortcomings, the rule of parsimony would seem to require that we hesitate before postulating additional, extra-intellectual agencies in accounting for judgmental failures" (p. 229). A parallel line of reasoning implies that apparent irrationality might not necessitate an appeal to functional theories of belief.

5.4 A final note on irrationality

[12] A similar notion is McCarthy and Hayes' (1969) distinction between heuristics and epistemology in artificial intelligence.

It might seem that the present treatment excuses all apparent irrationality or relaxes the definition of rationality to accommodate the most fringe beliefs, resulting in a view that people are always rational. Such a criterion of rationality would hardly seem worth having. As Oaksford and Chater (2007, pp. 24-25) note, the intuitive sense of rationality is a normative notion and entails that some inferences and decisions are better than others. If everyone is rational, any normative standard would be lost. On the other hand, rationality should be judged relative to one's actual capabilities; as in ethics, 'ought implies can.' A standard of rationality that is impossible to achieve appears as worthless as one which deems all behavior rational. It seems that a standard of rationality should be normative as well as achievable. The findings indicative of irrationality, surveyed above, do seem suboptimal and, as Stanovich (2011) has documented, there are individuals who do report the normative solutions on such tasks. This appears to meet our criteria for irrationality, however, the question of where this irrationality is located may be more subtle than it appears.

In section 5.3, we discussed the role that prior knowledge plays in belief formation and updating, including intervening in the processes underlying evidence search and evaluation. This suggests that irrationality may be understood as resulting from problematic beliefs, rather than irrational cognitive processes. On a related note, Newell (1994, p. 92) distinguished between performance on a task and the separate problem of acquiring knowledge related to the task. Newell argued that judgments of intelligent performance should be made relative to the knowledge a person actually possesses, even if they are ignorant of task-relevant information. For example, if a person is attempting to buy groceries and has never learned math, their abilities will be constrained by their lack of knowledge, but an assessment of their performance should account for this ignorance. This is not to say that such a lack of knowledge is not a problem, but that it is better viewed as a failure on the separate task of learning arithmetic. Similarly, while belief updating may be "optimal" given an individual's current knowledge and reasoning strategies, these might be improved with training, allowing incremental progress toward an achievable standard of rationality.

This view of irrationality as originating from ignorance, rather than biased cognitive mechanisms is an optimistic one. It suggests that education and instruction might ameliorate apparently irrational processing and lead to improved beliefs (see Scopelliti, Chapter 29, this volume; Metz, Chapter 30, this volume). Instruction in knowledge and mindware that aid accurate reasoning, and perhaps in attenuating information or methods that are less helpful, could allow people to achieve beliefs that are not just the best they can do given their current knowledge, but which approach a higher standard of rationality. Accomplishing this will require not just a deeper understanding of people's cognitive processes, but also of their prior beliefs. In other words, we must open the black box of the mind. This is the task of a cognitive science of belief.

6. The Cognitive Science of Belief

In this final section, we offer some brief remarks about the structure of the book. We organized the ideas presented in this volume into three main sections. The first section is concerned with the nature of beliefs and believing, a perennial set of questions about which much remains to be learned. This portion of the book introduces linguistic and philosophical perspectives on the nature of beliefs, presents normative models of optimal beliefs, explores the distinction between implicit and explicit beliefs, and discusses the psychology of beliefs from an evolutionary perspective. The goal of Section 1 is to provide readers with an overview of some of the main questions that arise regarding the ontology of beliefs as well as some of the tools used in different areas of cognitive

science to study and model beliefs. With these considerations in mind, Section 2 examines different domains of beliefs, including scientific beliefs, political and economic beliefs, religious beliefs, as well as beliefs about race and morality. In some cases, the beliefs under consideration are those of lay people, and in others, they are those of professional scholars working in the relevant areas of inquiry. By juxtaposing different domains of belief, our hope is that questions as well as answers that cut across domains might be brought into sharper focus and lead to possible integration between different areas of research on beliefs. Finally, Section 3 explores variation in beliefs. Here, chapters consider individual differences in beliefs, pathological beliefs, as well as environmental effects on beliefs. The book ends on what we take to be a hopeful note with two chapters that discuss ways to achieve more accurate beliefs.

References

Abelson, R. P. (1972). Are attitudes necessary. In B. T. King & E. McGinnies (Eds). *Attitudes, Conflict, and Social Change*. Academic Press.

Abelson, R. P. (1973). The structure of belief systems. In R. Schank & M. Colby (Eds). *Computer models of thought and language*. W. H. Freeman and Company.

Abelson, R. P. (1979). Differences between belief and knowledge systems. *Cognitive Science*, *3*(4), 355-366.

Abelson, R. P. (1986). Beliefs are like possessions. *Journal for the Theory of Social Behaviour*.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: the role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, *39*(6), 1037.

Anderson, John R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*(4), 249–277.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Psychology Press.

Bain, A. (1872). *Mind and body: The theories of their relation (Vol. 4)*. Appleton.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bobrow, D. G. (1975). Dimensions of representation. In *Representation and Understanding* (D. G. Bobrow & A. Collins, Eds). Morgan Kaufmann.

Braithwaite, R. B. (1932). The nature of believing. In *Proceedings of the Aristotelian Society (Vol. 33*, pp. 129-146). Aristotelian Society, Wiley.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed). *The Philosophy of Karl Popper*. Open Court.

Caplan, B. (2001). Rational ignorance versus rational irrationality. *Kyklos*, *54*(1), 3-26.

Caplan, B. (2011). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies-New Edition.* Princeton University Press.

Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. Wiley Interdisciplinary Reviews: *Cognitive Science*, *1*(6), 811-823.

Chater, N., Zhu, J. Q., Spicer, J., Sundh, J., León-Villagrá, P., & Sanborn, A. (2020). Probabilistic biases meet the Bayesian brain. *Current Directions in Psychological Science*, *29*(5), 506-512.

Cherniak, C. (1984). Computational complexity and the universal acceptance of logic. *The Journal of Philosophy*, *81*(12), 739-758.

Churchland, P. M. (1981). Eliminative materialism and propositional attitudes. *The Journal of Philosophy*, *78*(2), 67-90.

Churchland, P. S., & Churchland, P. M. (2013). What are beliefs? In F. Kruger & J. Grafman (Eds). *The neural basis of human belief systems*. Psychology Press.

Colby, K. M. (1964). Experimental treatment of neurotic computer programs. Archives of *General Psychiatry, 10*(3), 220-227.

Colby, K. M. (1973) Simulations of Belief Systems. In K. M. Colby & R. Schank, (Eds). *Computer Models of Thought and Language*. W. H. Freeman & Co.

Connors, M. H., & Halligan, P. W. (2015). A cognitive account of belief: A tentative road map. *Frontiers in Psychology, 5*, 1588.

Cox, R. T. (1961). *The Algebra of Probable Inference*. Oxford University Press.

de Finetti, B. (1970/1974). Teoria delle Probabilita 1. Translated by A. Machi and A. Smith, as *Theory of Probability 1*. John Wiley and Sons.

de Sousa, R. B. (1971). How to give a piece of your mind: or, the logic of belief and assent. The *Review of Metaphysics, 25*(1), 52-79.

Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy, 68*(4), 87-106.

Dennett, D. C. (2008). *Kinds of minds: Toward an understanding of consciousness*. Basic Books.

Dennett, D. C. (2017). *Brainstorms: Philosophical essays on mind and psychology*. MIT press.

Doyle, J. (1979). A truth maintenance system. *Artificial intelligence, 12*(3), 231-272.

Egan, O. (1986). The concept of belief in cognitive theory. In *Annals of Theoretical Psychology*. Springer, Boston, MA.

Evans, J. S. B., & Over, D. E. (2004). *If: Supposition, pragmatics, and dual processes*. Oxford University Press.

Faust, D. (1984). *The limits of scientific reasoning*. University of Minnesota Press.

Feldman, J. (2017). What Are the "True" Statistics of the Environment? *Cognitive Science, 41*(7), 1871-1903.

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

Fodor, J. A. (1968). *Psychological explanation: An introduction to the philosophy of psychology*. Random House.

Fodor, J. A. (1975). *The Language of Thought*. Harvard University Press.

Fodor, J. A. (1986). *Representations: Philosophical essays on the foundations of cognitive science.* MIT Press.

Fodor, J. A. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Z. W. Pylyshyn (Ed). *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. Ablex Publishing Corp.

Fodor, J. A. (1992). *A theory of content and other essays.* MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. MIT Press.

Frankish, K. (2004). *Mind and Supermind*. Cambridge University Press.

Frankish, K. (2009). Partial belief and flat-out belief. In F. Huber & C. Schmidt-Petri (Eds). *Degrees of Belief*. Springer.

Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience*. John Wiley & Sons.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*. Oxford University Press.

Goldman, A. I. (1986). *Epistemology and Cognition*. Harvard University Press.

Grandin, A., Boon-Falleur, M., & Chevallier, C. (Chapter 26, this volume). The belief-action gap in environmental psychology: How wide? How irrational?

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012).

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.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review. 108*, 814-834

Harman, G. (1986). *Change in view: Principles of reasoning*. The MIT Press.

Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making.* Sage Publications.

Henle, M. (1955). Some effects of motivational processes on cognition. *Psychological Review, 62*(6), 423.

Henle, M., & Michael, M. (1956). The influence of attitudes on syllogistic reasoning. *The Journal of Social Psychology, 44*(1), 115-127.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. MIT Press.

Hume, D. (1778). *The History of England, from the Invasion of Julius Caesar to the Revolution in 1688* (Vol. 6). Liberty Classics.

James, W. (1889). The psychology of belief. *Mind, 14*(55), 321-352.

Janlert, L. E. (1987). Modeling change: The frame problem. In Z. W. Pylyshyn (Ed). *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. Ablex Publishing Corp.

Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition. *Cultural Cognition Project Working Paper Series* No. 164.

Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly, 24*(2), 163-204.

Kitcher, P. (1998). Truth or Consequences? In *Proceedings and Addresses of the American Philosophical Association* (Vol. 72, No. 2, 49-63). American Philosophical Association.

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

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational behavior and human decision processes, 56*(1), 28-55.

Kuhn, D. (1991). *The Skills of Argument*. Cambridge University Press.

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

Lewin, K. (1936). *Principles of topological psychology*. McGraw-Hill.

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.

Lycan, W. G., (1986). Tacit Belief. In R. J. Bogdan (Ed). *Belief: form, content, and function*. Oxford University Press.

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

McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. L. Webber & N. J. Nilsson (Eds). *Readings in artificial intelligence*. Morgan Kaufmann.

Metz, E. (Chapter 30, this volume). Building better beliefs through actively open-minded thinking.

Molnar, A. & Loewenstein, G. (Chapter 14, this volume). Thoughts and players: An introduction to old and new economic perspectives on beliefs.

Newell, A. (1982). The knowledge level. *Artificial intelligence, 18*(1), 87-127.

Newell, A. (1994). *Unified Theories of Cognition*. Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Nisbett, R. E., & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Prentice Hall.

Nozick, R. (1983). *Philosophical Explanations*. Harvard University Press.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Okabe, T. (1910). An Experimental Study of Belief. *The American Journal of Psychology, 21*(4), 563-596.

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences. 25*(5), 388-402.

Perkins, D. (1995). *Outsmarting IQ: The emerging science of learnable intelligence*. Simon and Schuster.

Pinker, S. (1997). *How the mind works*. Princeton University Press.

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences, 107(Supplement 2)*, 8993-8999.

Popper, K. (1978). Natural selection and the emergence of mind. *Dialectica*, 339-355.

Porot, N., & Mandelbaum, E. (Chapter 2, this volume). The science of belief: A progress report.

Pylyshyn, Z. W. (1986). *Computation and cognition: Toward a foundation for cognitive science.* The MIT Press.

Quilty-Dunn, J., & Mandelbaum, E. (2018). Against dispositionalism: Belief in cognitive science. *Philosophical Studies, 175*(9), 2353-2372.

Ramsey, F. P. (1926). Truth and Probability. In R.B. Braithwaite (Ed). *The Foundations of Mathematics and other Logical Essays*. Kegan, Paul, Trench, Trubner & Co.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*(5), 880.

Salancik, G. R. (1982). Attitude-behavior consistencies as social logics. In M. Zanna, E. T. Higgins, & C. P. Herman (Eds). *Consistency in social behavior: The Ontario symposium* (Vol. 2). Erlbaum.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1982). Automatic/Control Processing and Attention. Illinois Univ Champaign Human Attention Research Lab.

Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Noûs*, *36*(2), 249-275.

Scopelliti, I. (Chapter 29, this volume). Training can improve decision making.

Shtulman, A. (Chapter 15, this volume). How intuitive beliefs inoculate us against scientific ones.

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

Simon, H. A. (1980). Cognitive science: The newest science of the artificial. *Cognitive Science, 4*(1), 33-46.

Simon, H. A. (1996). *The sciences of the artificial*. 3rd edition. MIT Press.

Simon, H. A. (1997). *Administrative Behavior, fourth edition*. Simon and Schuster. (Original work published 1945).

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology, 4*(1), 165-173.

Smith, M. B., Bruner, J. S., & White, R. W. (1956). *Opinions and Personality*. Wiley.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36*(11), 1202.

Sommer, J., Musolino, J., & Hemmer, P. (in preparation). The trouble with consistency in belief.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89-96.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Second Edition. Cambridge, MA: Harvard University Press.

Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.

Stich, S. P. (1978). Beliefs and subdoxastic states. *Philosophy of Science, 45*(4), 499-518.

Stich, S. P. (1983). *From folk psychology to cognitive science: The case against belief*. The MIT Press.

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

Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Bayesian or biased? Analytic thinking and political belief updating. *Cognition, 204*.

Trivers, R. (2011). *The Folly of Fools: The logic of deceit and self-deception in human life*. Basic Books.

Tversky, A., & Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content*. Harvard University Press.

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, *25*(4), 41-78.

Williams, D. (2020). Socially adaptive belief. *Mind & Language*, *36*(3), 333-354.

Wilson, D. S. (2010). Rational and Irrational Beliefs from an Evolutionary Perspective. In D. David, S. J. Lynn, & A. Ellis (Eds). *Rational and Irrational Beliefs: Research, Theory, and Clinical Practice*. Oxford University Press.

Woods, W. A. (1975). Foundations for semantic networks. In *Representation and Understanding* (D. G. Bobrow & A. Collins, Eds). Morgan Kaufmann.

Zanna, M. P., Higgins, E. T., & Herman, C. P. (1982). *Consistency in social behavior: The Ontario Symposium* (Vol. 2). Erlbaum.

Zhang, Q. (Chapter 4, this volume). How And Why Does Schematic Knowledge Affect Memory.