CHAPTER 1

The Motivational Processes of Sense-Making

Zachary Wojtowicz, Nick Chater, and George Loewenstein

> Humboldt thought . . . A hill whose height remained unknown was an insult to the intelligence and made him uneasy . . . A riddle, no matter how small, could not be left by the side of the road.
>
> Kehlmann, *Measuring the World*

1.1 Introduction

Our innate drive to make sense of things is one of the most powerful forces shaping both individual human cognition and collective societal progress. Consider the huge impetus behind the accumulation and critique of knowledge, which touches on all subjects – whether they be scientific, historical, or cultural – and proceeds at a grand scale to fill every corner of life, from the lectures of academic halls to the chatter of coffee houses. Sometimes knowledge is sought with some immediate objective in mind, but this makes up, on the whole, a surprisingly small part of our intellectual life. The force driving us to identify the causes of the Bolshevik revolution, map the deep oceans or the surface of the moon, chart the history of jazz, and understand the origins of life is powerful enough to drive millions of hours of scholarly activity – often without obvious direct application and even without pay. Daily life, too, is filled with myriad activities that provoke our interest, from exploring new cities, music, or cuisine to tracing our family history, becoming intrigued by gossip at the next table, and following the news. Indeed, these pleasures are so great that vast sectors of human activity are devoted to creating objects whose primary purpose is to stimulate the delights of sense-making: novels, movies, works of art, puzzles, and many more.

Although we generally take our undirected urge to make sense of the world for granted, it may seem strange upon reflection, especially because it frequently does not confer obvious near-term benefits. One might expect

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

that the brutal logic of natural selection would have favored creatures interested only in practical concerns that directly enhance survival and reproduction. One might imagine, too, that societies with a laser-like focus on knowledge with immediate utility, rather than those promoting apparently purposeless inquiry, would be the ones to get ahead. Yet the opposite seems to be true: just "figuring stuff out" often yields unpredictable, but enormous, practical benefits. Indeed, the aimlessness of human curiosity may, paradoxically, be the secret of our species' success (if it can be called that). This chapter focuses on the rationale for, and nature of, the motivational processes underlying the drive for sense-making: the intrinsic human desire to make sense of the world. We explore why the drive for sense-making is so valuable and, crucially, how particular features of its implementation can at times lead us astray into systematically incorrect beliefs.

In Section 1.2 ("The Drive for Sense-Making"), we start by discussing why sense-making generates a drive, similar to those associated with the primary reinforcers of food, water, sleep, sex, shelter, and air. The essence of our argument is that the drive for sense-making helps us balance the immediate benefits of satisfying tangible wants against the delayed benefits of investing in knowledge about ourselves and the world around us. The task of optimally making such trade-offs, which are incumbent upon all agents capable of self-directed learning, presents a formidable challenge because explicitly forecasting the beneficial consequences of each and every potential cognitive investment is often more trouble than it is worth. For many of the decisions we are faced with every day, such calculations would require a great deal of computational effort and yield inaccurate results, if they are even possible at all.

The drive for sense-making circumvents this problem by directly incentivizing our ability to make sense of the world in the here and now. It operates under the general assumption that "knowledge is power" – that is, that an enriched understanding of the world will benefit us in the future even if we cannot foresee exactly how. In the absence of a drive for sense-making, a limited ability to prospectively evaluate, and hence appreciate, the benefits of cognitively enriching activities would lead us to persistently underinvest in them. In this way, the drive for sense-making fills a critical gap that arises in purely goal-oriented cognition.

An economic framing of this argument reveals that the motivational incentive generated by the sense-making drive is analogous to the monetary incentive generated by a subsidy on knowledge-producing activities. We glean insights from this analogy by discussing why societies do in fact

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

subsidize what is called basic research: "systematic study directed toward greater knowledge or understanding of the fundamental aspects of phenomena and of observable facts without specific applications."[1] Analogous to our argument that the drive for sense-making exists to enhance future pay-offs, the quoted definition continues by noting that basic research is "farsighted high payoff research that provides the basis for technological progress."

While Section 1.2 examines reasons why humans have a drive for sense-making, Section 1.3 ("The Objectives Governing Sense-Making") examines three different factors that guide the particular form sense-making takes: (1) the practical utility of accurate beliefs for attaining concrete goals, (2) the desire to make sense of the world in a way that feels good, and (3) the impact of computational limitations on the sense-making process, including our limited ability to explicitly predict what information will turn out to be useful.

Of note, only the first of these categories is accounted for by standard rational theories of human behavior. Standard economics treats cognition as strictly a means to material ends. Accordingly, it holds that both cognitive states (e.g., knowledge, understanding, beliefs) and functions (e.g., information acquisition and processing) are only valuable to the degree that they are "instrumental" in helping us achieve concrete goals, such as increasing consumption or reducing labor. According to this view, because a rational agent is better prepared to maximize utility when they have an accurate understanding of their environment (Blackwell, 1953), the goal of information acquisition and processing should be to arrive at beliefs that are as accurate (and hence useful) as possible.

Some cognitive scientists, for their part, have recently proposed that correctly predicting the environment is all that matters to agents – essentially inverting the classical economist's long-standing position by entirely subordinating material objectives to cognitive ones (Friston, 2010). Such "predictive processing" accounts take a different conceptual and mathematical form than rational economic models, but they share the fundamental conclusion that our cognition is exclusively aimed at generating accurate predictions about the future.

These perspectives yield important insights, but they also leave out critical aspects of sense-making. First, theories that exclusively focus on instrumental value (e.g., standard economics) fail to explain why we so fervently pursue activities, such as solving puzzles or reading mystery

[1] www.law.cornell.edu/cfr/text/32/272.3

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

novels, that seem to yield little instrumental value relative to other readily available uses of our time.[2] On the other hand, theories that exclusively focus on inferential value (e.g., predictive processing models) do not readily explain the purposeful, goal-oriented nature of much of our cognition: the obvious fact that we do care about eating, sleeping, attracting the attention of potential mates, and achieving innumerable other material objectives. Predictive processing theories also seem to make the implausible prediction that agents should seek a maximally predictable environment and stay there forever (known as the "dark room problem"; see Friston, Thornton, & Clark, 2012; Sun & Firestone, 2020).

Both the standard instrumental and the predictive processing theories of sense-making also leave out the fact that motivation and beliefs frequently interact with one another. In recent years, however, economists have begun to recognize that certain cognitive states and processes seem to be valued in themselves and confer strong motivational significance for agents. Work on "belief-based utility" (Loewenstein & Molnar, 2018) has shown that the desire to make sense that feels good plays a significant role in determining how we seek, interpret, and act upon information. In a similar vein, psychologists outside of the predictive processing tradition have long recognized the importance of motivated reasoning in shaping our beliefs (Kunda, 1990).

Motivational factors are also crucial for ensuring we make the best use of our limited cognitive resources when gathering and processing information. For example, the motivational signals of flow and curiosity direct us toward the most valuable new information we might gather through reading, observing, discussing, or experimenting (Wojtowicz, Chater, & Loewenstein, 2020; Wojtowicz & Loewenstein, 2020), and the sense of "cognitive dissonance" (Festinger, 1957) alerts us to inconsistencies in our beliefs that require further analysis and scrutiny. As a result, understanding what interpretation an individual will arrive at requires, at least in part, accounting for the motivational factors that guide our uptake and processing of information.

The standard instrumental and predictive accounts also generally overlook the impact of computational constraints on the sense we can and do make of the world. In particular, these accounts leave out the fact that considering each of the myriad possible interpretations of a given body of information as prescribed by Bayes' rule is often intractable (Jeffrey, 2004),

[2] Notably, most people spend a shockingly small fraction of their free time purposefully investing in economically valuable forms of human capital.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

even for relatively simple problems (Kwisthout, 2011; Van Rooij, 2008). Evidence suggests that our cognitive system instead approximates this normative standard by sampling interpretations one at a time (e.g., we see the duck-rabbit as either a duck, or a rabbit, but not both at once; see Figure 1.1). As we will argue, this has huge ramifications for how sense-making operates (Chater, 2019; Pashler, 1999).

Perhaps the most important practical limitation of both the standard instrumental and predictive processing accounts of sense-making, however, is that they fail to explain the troubling predominance of nonnormative belief patterns in society or to provide adequate guidance as to how they might be addressed. Recent developments – such as the precipitous growth of online radicalization, conspiracy theory communities, religious extremism, political polarization, anti-science rhetoric, climate change skepticism, antivaccination sentiment, COVID-19 denial, and hate groups – have heightened concerns about the descriptive adequacy of rational frameworks. Such phenomena are especially puzzling for rational theories given that their growth has coincided with (and, arguably, been fueled by) the rise of the Internet, which enables free and instantaneous access to much of human knowledge. According to a purely rational conception of belief-formation, such a dramatic increase in access to high-quality information should have resulted in a commensurate increase in the accuracy of popular beliefs, contrary to recent events. Finally, Section 1.4 ("Implications") shows how the alternative perspective we lay out in the preceding sections can be used to better understand these phenomena.

The core argument of this chapter is that analyzing the multiplicity of objectives governing sense-making can help to explain the scientific and practical puzzles that vex current theories. According to our account, instrumental, inferential, and computational factors work together to guide our decisions. The drive for sense-making is primarily directed at maximizing predictive accuracy, but the other above-noted factors – belief-based utility and cognitive efficiency – also shape the sense we make. The interaction of these (sometimes competing) factors gives rise to characteristic patterns of irrationality, which leave us vulnerable to seductive mistruths that are increasingly amplified, both passively (by technologies that spread misinformation with unprecedented speed) and actively (by social movements dedicated to propagating abnormal patterns of beliefs). A comprehensive picture of how sense-making fits into the broader psychology of motivation explains characteristic distortions in our relationship with truth and, in turn, sheds new light on these concerning trends.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

1.2 The Drive for Sense-Making

In this section, we develop a functional account of the drive for sense-making that explains its characteristic features by analyzing the cognitive problem it solves. Our account starts with the general observation that many – if not all – motivational states exist to address the boundedness of our rationality (Hanoch, 2002; MacLeod, 1996; Muramatsu & Hanoch, 2005; Samuelson & Swinkels, 2006; Sorg, Singh, & Lewis, 2010). Immediate drives, feelings, and urges help us make decisions quickly and cheaply by circumventing the need to prospectively calculate the costs and benefits of each potential option explicitly. More specifically, these visceral states circumvent the (often intractable) task of forecasting the consequences of our actions arbitrarily far into an uncertain future (Bechara & Damasio, 2005; Damasio, 2006) by encoding the expected survival value associated with evolutionarily significant behaviors, such as consuming key nutrients, copulating, nurturing offspring, and avoiding bodily harm (Cabanac, 1971; Cosmides & Tooby, 2000).

A subset of these states specifically function to shape our information seeking and processing behavior: boredom, flow (Wojtowicz et al., 2020), curiosity (Wojtowicz & Loewenstein, 2020), mental effort (Kurzban et al., 2013; Shenhav et al., 2017), and, as we will argue, the drive for sense-making. Although these states are psychologically distinct, they share many theoretical connections and overlap operationally due to the interrelated nature of their underlying functions. Indeed, we have argued elsewhere that curiosity may in fact be a special case of the drive for sense-making (Chater & Loewenstein, 2016; Wojtowicz & Loewenstein, 2020), and that flow and boredom partly reflect deviations from the amount of cognitive enrichment one has come to expect from similar environments (Wojtowicz et al., 2020).

As we have suggested, explicitly appraising the value of an increase in information, knowledge, or understanding is computationally intractable and would exhaust our finite cognitive resources in most situations. In most cases, our models of the external world are so underspecified that they do not provide meaningful answers to the question of how useful a particular piece of information is likely to be. But even if such models were available and could in principle yield well-defined answers, the computational costs of generating accurate predictions would still be prohibitive in most circumstances. This is because explicitly assessing the value of a piece of information or knowledge requires that we consider the many instances where a piece of information or knowledge would be applied. In general, the number of potential futures grows exponentially with the time horizon one considers; because cognitive resources can be applied arbitrarily far in the future, this explosion can be difficult to contend with (Bellman, 1957; Savage, 1972; Sutton & Barto, 2018). Planning the optimal sequence of information-acquisition behaviors also requires that one anticipate how information gained at each stage will impact the interpretation and usefulness of information gained at all later stages (Meder et al., 2019).

Our hypothesis is that the brain circumvents these computational challenges by directly incentivizing actions that result in increased understanding using a motivational state that we experience as the drive for sense-making (Chater & Loewenstein, 2016). This approach avoids the need to prospectively calculate the potential usefulness of knowledge explicitly because "sense" is quantified using a contemporaneous measure of our ability to explain empirical regularities in the world. This is principally a backward-looking appraisal that operates on fixed data and, critically, does not require us to simulate the exponential number of diverging possible futures where that sense might be applied.

The exact nature of how the brain quantifies sense is still an area of active research, but one hypothesis is that sense measures our ability to *compress* the information we encounter into explanations. Data can be compressed to the extent that patterns can be found in that data, so the degree of compression achieved provides a natural measure of how well patterns in that data have been uncovered, irrespective of whether those patterns will turn out to help achieve any practical goal. Viewed in this way, the amount of *sense* we make out of a particular piece of information corresponds to the reduction in representational code length that we can achieve when we discover successively better (i.e., compressive) explanations for it. *Sense-making* occurs when we strike upon insights or critical pieces of new information that help us to resolve ambiguities or recognize regularities in an existing set of facts, thereby enabling us to compress them further.

As an example, consider the text "GNIKAMESNES." While this might at first appear to be meaningless, it acquires more sense – especially in the context of this chapter – once we recognize it as "SENSEMAKING" spelled backwards. Under the compression hypothesis, this insight *makes sense* of the original text precisely because it reduces an unfamiliar and unwieldy jumble of letters to two simple cognitive operations: recalling a familiar word ("sense-making") plus applying a familiar transformation (left–right transposition), enabling us to cognitively represent, manipulate, encode, and recall the string more efficiently. If, for illustration, we imagine all "units"

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

are equal (whether letters, words, or transpositions), then we can see that spotting this new representation of "GNIKAMESNES" counts as definite progress. For a hypothetical cognitive system that encoded text using such a system – that is, either by storing it letter by letter or by applying a transformation to previously stored text – detecting this pattern would reduce the representational length of "GNIKAMESNES" from eleven to just two units, thus yielding nine "units of sense."

While more research is needed to determine what form the representations underlying a fully domain-general measure of sense-making might take, a variety of candidates have been proposed that range from the most comprehensive model of computation – programs compiled by a Turing complete language (Chater, 1996; Chater & Loewenstein, 2016; Chater & Vitányi, 2003) – to less powerful automata capable of expressing more restricted grammars (i.e., ones at a lower level of the Chomsky hierarchy; see Griffiths & Tenenbaum, 2003; Simon, 1972). For now, the question of how these mathematically abstracted computational-level measures might be implemented in the brain is a largely unexplored – but exciting – topic for future research.

According to this perspective, the *drive for sense-making* is an innate source of motivation that rewards us for each marginal increase (and, perhaps, punishes us for each marginal decrease) in our ability to compress information into efficient representations (Chater & Loewenstein, 2016). While the goal of compressing the information we encounter is certainly valuable for its own sake (e.g., because it enables us to store information more efficiently in the brain), its primary benefit is that it directs our cognitive machinery to actively search for regularities in the phenomena we observe, thus enabling us to better describe, predict, and control the world.[3]

Given that sense-making and the classical drives serve similar psychological functions, they also share many basic characteristics. For example, classical drives consist of both a "carrot" and a "stick": pleasure when we fulfill the drive's target behavior and pain when we abstain from it. For example, eating when hungry feels good, but failing to do so for long

[3] This hypothesized correspondence between sense-making and compression may also help explain why memorization is such a critical component of pedagogy. In many educational contexts, no one truly expects that students will retain most of the information they learn after the course is finished. Nevertheless, the challenge of memorizing a large domain of related facts efficiently enough to reproduce them on a test forces students to search for the underlying connections, structures, and regularities that are the true marrow of knowledge. Even if the particulars are themselves forgotten, the concepts which bind them together are generally retained, and these are often the most useful.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

periods of time becomes highly aversive, especially while in the presence of food. Paralleling these mechanisms, a few studies have shown that curiosity activates the same areas of the brain that process extrinsic rewards (Jepma et al., 2012; Kang et al., 2009), suggesting that sense-making considerations may enter into standard reward calculation as an intrinsic reward (or punishment) signal (Gottlieb et al., 2013; Kidd & Hayden, 2015).

In the case of sense-making, the carrot corresponds to the pleasure we experience when we succeed at uncovering regularities that generate new sense. In moments of profound insight, the sudden rush of sense-making pleasure can be quite intense (Gopnik, 1998), as exemplified by Archimedes' famous exclamation of "Eureka!" upon discovering the principle of buoyancy. Less acute instances of sense-making pleasure also permeate many aspects of our daily life and range from the delight of discovering the answer to a riddle to the satisfaction of arriving at a mystery novel's grand reveal. The stick, on the other hand, consists of the unpleasant sense of deprivation we feel when we are faced with a salient lack of understanding, as exemplified by the torment of leaving a riddle unanswered or a mystery novel unfinished. This deprivation is stronger the more apparent the gap in our understanding becomes, and the less easily it can be closed (Golman & Loewenstein, 2018; Loewenstein, 1994).

The drive for sense-making is related to, and may even entirely subsume, other motivational states that guide how we gather and process information. The most obvious example is curiosity, which shares the same drive-like features (Loewenstein, 1994), solves the same cognitive problem (Wojtowicz & Loewenstein, 2020), and has overlapping behavioral implications (Chater & Loewenstein, 2016) as sense-making. Other examples include boredom, which redirects our attention away from understimulating activities when more promising opportunities seem to exist in our environment, and flow, which keeps our attention focused on the task at hand when other, better opportunities seem unlikely to exist. Both of these states emerge from a counterfactual comparison between the current and anticipated value of engagement, which is largely determined by the degree of sense-making achieved (Chater & Loewenstein, 2016; Wojtowicz et al., 2020). Sense-making is also closely related to our preferences for creating and resolving uncertainty (Ruan, Hsee, & Lu, 2018) and may underpin the states of suspense and surprise (Ely, Frankel, & Kamenica, 2015). Finally, the explanatory values we use to evaluate everything from scientific hypotheses to quick excuses – such as how simple, descriptive, or unifying an account is – are key implements of sense-making and arguably exist to further the same overall inferential objective (Wojtowicz & DeDeo, 2020).

According to our account, the drive for sense-making makes up for our limited ability to appreciate the true long-term value of investing in knowledge. This parallels the way in which governments use subsidies to overcome the inherent tendency of private enterprise to underinvest in knowledge-generating activities. In a social setting, it is virtually free to include, and very difficult to exclude, others from using knowledge once it has been created. Knowledge is therefore an example of what economists refer to as "public goods," which are chronically undersupplied relative to the socially efficient optimum because potential producers cannot capture the full value they create by investing in them.

Modern societies address this problem through government funding of public universities, scientific institutions, and basic research. Just as the drive for sense-making is necessary to motivate undirected inquiry, this funding is necessary to sustain learning for its own sake, without any immediate expectation of profit. As it turns out, however, such research often lays the groundwork for a variety of unforeseen applications that more than pay for the initial outlay through increased long-term economic growth. Also like sense-making, our inability to predict which types of knowledge will eventually be useful for particular problems means that continued broad investment in basic research often turns out to be the best way of ensuring we eventually solve them. Moreover, heavy-handed attempts to override research curiosity and narrowly optimize the direction of their work often end up backfiring because the process of justifying the value of scientific projects (including, sometimes, their practical value) through grant writing and related activities takes up time that could be used for actual research. In much the same way, forecasting the future value of sense-making uses up the very mental resources one needs to make sense of the world.

The function of the drive for sense-making is also illustrated by an analogy to education. Students perpetually complain that what they learn has no obvious value or relevance to their daily lives or future careers. While out-of-date education is certainly a problem, these critiques are often overstated, especially in young children who have no conception of what adult life is like and consequently cannot accurately gauge the importance of the knowledge and skills they are learning. Indeed, the distinction between *education* and *training* nicely captures the difference between the provision of knowledge which has no immediate application and that which is focused on learning an applicable skill. While *training* is, of course, extremely important, a school and university system focused purely on immediately applicable skills would fail to cultivate the growth of general knowledge that is crucial to long-term development. The main goal of *education*, therefore, is to provide a broad base of

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

fundamental knowledge that helps students get a sense of the overall "geography" of knowledge in its broadest outlines. As students get older and their particular interests, proclivities, and goals become more clear, a greater degree of specialization is gradually introduced, but education is not, and cannot be, perfectly tailored.

If, as this analysis suggests, the purpose of formal education is to ensure that students acquire skills that are *unexpectedly* useful (and therefore would not seek out themselves), initiatives to shift the curriculum toward more apparently useful material may miss the point entirely. In much the same way, sense-making drives us to enrich our cognitive capacities in numerous directions, only some of which will turn out to be useful. Like a good teacher, the sense-making drive encourages us to engage in enriching activities, even in the absence of foreseeable benefits. Given that sense-making functions as "nature's endogenous teacher," it is not surprising that its derivative states, most notably curiosity, play a critical role in supporting learning, both in and out of the classroom (Deci & Ryan, 1981; Litman, 2005; Markey & Loewenstein, 2014; Pluck & Johnson, 2011; Wade & Kidd, 2019).

These points are corroborated by research in machine learning, which has shown that intrinsically generated sense-making rewards help to foster robust learning by encouraging structured exploration. Schmidhuber (1991, p. 222) points out that these incentives not only instill a desire for an artificial system to improve its understanding of the world, but also "to model its own ignorance, thus showing a rudimentary form of self-introspective behavior." Lopes, Lang, Toussaint, and Oudeyer (2012) further show that such rewards can be generated using heuristic online estimates of learning progress, closely matching our conception of the drive for sense-making as rewarding gains in our ability to compress existing information. In a similar vein, Pathak, Agrawal, Efros, and Darrell (2017) demonstrated the benefits of combining standard reinforcement learning with an "intrinsic curiosity module" that learns to predict which actions might expose the shortcomings in an agent's model of the environment. They show that adding these predictions to the stream of extrinsic reward feedback an agent receives from the environment speeds up learning; in fact, their agents learn to successfully navigate video games when motivated by intrinsic curiosity alone (see also Burda et al., 2018).

1.3 The Objectives Governing Sense-Making

In this section, we describe three objectives that shape sense-making, either directly, through motivational signals that orient sense-making, or indirectly, through constraints on the cognitive processes that underlie it.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

1.3.1 Instrumental Objectives

The most obvious goal of sense-making is to help people make decisions that reliably lead to desired outcomes. From the perspective of decision theory, a rational agent acting in isolation is better equipped to pursue concrete ends when armed with more accurate beliefs (Blackwell, 1953), so the instrumental objective of sense-making often boils down to developing beliefs that are as accurate as possible. Indeed, in extreme cases, holding beliefs that are too divorced from reality (e.g., believing that one knows how to swim when one does not) can be fatal. Consequently, there seem to be strong constraints on the sense-making process: we cannot simply believe whatever we wish, and we labor to justify even our most fanciful beliefs to ourselves and others.

However, the fact that humans are both boundedly rational and highly social adds several important caveats to the truth-orienting function of sense-making, such that inaccurate beliefs may sometimes be advantageous when other psychological factors are taken into account. For example, the autonomic effects of nervousness evolved because they are usually adaptive, but they have the unintended consequence of degrading performance in some circumstances, such as test taking (Zeidner, 2010), public speaking (Beatty 1988), high-stakes games (Ariely, Gneezy, Loewenstein, & Mazar, 2009), athletic performance (Kleine, 1990), and sexual function (McCabe, 2005). Conditional on being subject to these autonomic forces, overconfidence in our objective abilities might benefit us in situations in which performance anxiety would otherwise hold us back. Systematic cognitive errors can also be beneficial if and when they compensate for other types of errors. As one example, Kahneman and Lovallo (1993) argue that overconfidence can be beneficial to the degree that it compensates for the conservatism and extreme avoidance of risk that would, in its absence, arise from loss aversion.

The beliefs we hold also change the way others regard us. For example, people are more likely to trust the leadership and advice of those who are self-confident. Anderson, Brion, Moore, and Kennedy (2012) present a series of studies showing that overconfidence leads other people to view an individual as more competent, and generally enhances their social status. Some strategic interactions, such as the game of chicken, also favor those who can convince others of their irrational commitment to undertake risky actions (Colman, 2003; Rapoport & Chammah, 1966; Schelling, 1980). While it is always possible, in principle, for a well-calibrated individual to fake confidence, such an act can, in practice, be

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

difficult to sustain. In some domains, the most effective way to convince others of one's exceptional abilities may be to first convince oneself.

Mercier and Sperber (2011) go even further, advancing the provocative hypothesis that the principle function of reasoning is to develop arguments that will be convincing to others. Needless to say, it is not always in one's personal best interest to reason in good faith when the objective is to sway someone else. According to this account, many apparently irrational aspects of cognition are actually driven by the benefits associated with successfully influencing others.

1.3.2 Hedonic Objectives

Although beliefs primarily function to help us achieve desired outcomes, people also care about what happens purely in their own minds. In other words, beliefs are not merely a means to an end, but can also become an end in themselves. While this general phenomenon – known in economics as belief-based utility – is at odds with basic tenets of rational thought (and, as we will describe, can undercut the instrumental function of beliefs outlined in the preceding section), it nevertheless performs an indispensable cognitive function by motivating us to pursue complex goals that would be hard to define without the aid of sense-making.

Evolution has, as we have noted, endowed us with a variety of motivational mechanisms that encode the value of various goals and push us to pursue beneficial actions, most notably the visceral feeling states and hedonic signals associated with classic drives that incentivize us to maintain homeostasis and satisfy various biological imperatives. While the satisfaction of many basic physiological goals can be determined automatically and without conscious awareness (e.g., monitoring the blood for a satisfactory glucose level), measuring progress on other goals, especially social goals, depends on nuanced inferences that must be assessed using higher-level cognitive processing. For instance, we care about considerations such as our standing in the world – whether we are liked and respected by others – and, with especially obvious evolutionary significance, whether we are found attractive by potential mates. Our motivational system induces us to pursue these goals by making certain belief states directly valuable. The pleasure associated with, for example, believing that others view us favorably provides an incentive for us to behave in a fashion that makes it true. This may, in turn, lead to substantial long-term benefits, such as the cooperative support of others, although the

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

specific nature of these benefits will, of course, be difficult to foresee precisely.

Belief-based utility exists to motivate behaviors that bring about desirable situations, but it is an imperfect mechanism for achieving this goal from a purely hedonic point of view. After all, simply believing what makes us feel good, irrespective of reality, would be a more direct route to unlocking the pleasures of the mind. Fortunately for the survival of our species, there seem to be significant limitations on our ability to believe whatever makes us feel good (Loewenstein & Molnar, 2018). So, for example, we cannot, by sheer force of will, perceive low teaching feedback scores as high praise – though we may be able to avoid looking at our teaching ratings entirely. However, despite such constraints on our ability to see what we want to see, the motives induced by belief-based utility can, in some instances, distort our relationship with truth and undermine our ability to achieve material goals. For example, an overestimate of our ability might feel good, but it can also lead us to expend time, energy, and money on endeavors where we are overwhelmingly likely to fail.

Exactly how motivational forces influence the direction that sense-making takes is an interesting and underexplored question. The influence of motivational processes on sense-making is undoubtedly aided by the fact that sense-making is, like most cognitive processes, sequential. That means that motivations can influence the *direction* that information processing takes. As Epley and Gilovich (2016, p. 133) note, "People don't simply believe what they want to believe . . . People generally reason their way to conclusions they favor, with their preferences influencing the way evidence is gathered, arguments are processed, and memories of past experience are recalled. Each of these processes can be affected in subtle ways by people's motivations." "For propositions we want to believe," Gilovich (2008, pp. 83–84) writes in his classic *How We Know What Isn't So*, "we ask only that the evidence not force us to believe otherwise. . . For propositions we want to resist, however, we ask whether the evidence compels such a distasteful conclusion. . . For desired conclusions, in other words, it is as if we ask ourselves, 'Can I believe this?', but for unpalatable conclusions we ask 'Must I believe this?'." Or, as Kunda (1990, pp. 482–483) expressed it, people "draw the desired conclusion only if they can muster up the evidence necessary to support it."

As the above-quoted passages hint, the processes that people use to achieve sense-making that feels good bear a striking resemblance to the biased processes that scientists use to collect and analyze data in a fashion that supports the conclusions they want to arrive at (c.f. John,

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011) – a set of practices that have come to be known, collectively, as "p-hacking." Much as scientists may collect just enough information to support their favored hypothesis and no more, people who want to behave selfishly without perceiving themselves as such will avoid collecting new information about the consequences of their actions when their current data supports the conclusion that their actions will not hurt others (Chen, et al., 2020).

One potential consequence of motivated processing is a phenomenon known as belief polarization, which occurs when exposure to the same new piece of evidence causes individuals who hold different beliefs to diverge even further (Batson, 1975; Liberman & Chaiken, 1992; Lord, Ross, & Lepper, 1979). Some have rightly pointed out that this pattern of updating is not necessarily irrational given that it can result from Bayes' rule in certain circumstances (Cook & Lewandowsky, 2016; Jern, Chang, & Kemp, 2014). However, by the same token, consistency with Bayes' rule does not, on its own, necessarily preclude the influence of motivational factors. Indeed, the many degrees of freedom available to a mischievous Bayesian – what evidence to consider and how, exactly, to interpret that evidence – provide a variety of opportunities for motivation to influence an otherwise mechanical application of Bayes' rule (c.f., Rabin & Schrag, 1999).

In Cook and Lewandowsky (2016), for example, Bayes' theorem is made to accommodate belief polarization through the addition of variables such as "a pro-market worldview" and "trust in scientists" that influence an agent's priors and interpretation of evidence about global warming. This, however, only pushes the question of motivational influence up a level: while it may indeed be Bayesian for someone to reject the academic consensus on global warming conditional on the belief that climate science is a communist conspiracy to undermine the free market, this hypercritical approach to evidence may itself not be warranted, especially if it were only adopted to protect a cherished belief.

The question of what normative constraints, if any, can be said to limit how a rational agent constructs their likelihood function and prior is a fascinating, complex, and as-yet unresolved problem; our goal here is simply to point out that the apparent use of Bayesian inference at one level does not necessarily preclude the influence of motivational factors at another. At the same time, it is important to note that post-hoc rationalization and other forms of motivated reasoning are not necessarily irrational once the many practical considerations that constrain cognition

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

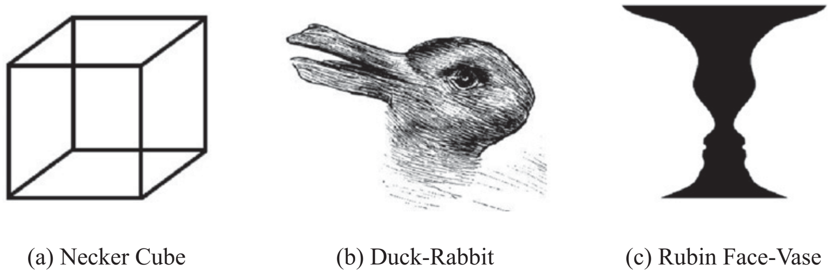
have been taken into account. Cushman (2020), for instance, argues that adjusting beliefs to rationalize our actions is a functional mechanism that transfers knowledge from the decision processes that underlie instinct, intuition, and habit to our rational mind. We turn to the influence of cognitive constraints next.

1.3.3 Cognitive Efficiency

Sense-making, like all cognitive processes, is subject to computational constraints. This has the implication that we generally cannot consider every potential interpretation of information in strict accordance with Bayes' rule (Aragones, Gilboa, Postlewaite, & Schmeidler, 2005; Kwisthout, 2011; Van Rooij, 2008). Recent work suggests that our brain instead approximates Bayesian inference using a step-by-step process of sampling – that is, we adopt a single working hypothesis and update it over time as we think of alternative hypotheses and gather new data (Bramley, Dayan, Griffiths, & Lagnado, 2017; Dasgupta, Schulz, & Gershman, 2017; Dayan, 1998; Gershman, Vul, & Tenenbaum, 2009; Griffiths, Vul, & Sanborn, 2012; Levy, Reali, & Griffiths, 2009; Sanborn & Chater, 2016; Vul, Goodman, Griffiths, & Tenenbaum, 2014; Vul & Pashler, 2008). Randomly selecting potential hypotheses for consideration is inefficient because most of them would turn out to be nonsensical, so the alternatives we consider are usually local, piecewise modifications of the working hypothesis we already have and are heavily informed by prior beliefs (Bramley et al., 2017; Tenenbaum, Griffiths, & Niyogi, 2007). This strategy brings otherwise intractable problems within reach, but it also leads to a number of biases (Chater et al., 2020; Sanborn & Chater, 2016).

The ambiguous visual patterns of Figure 1.1 demonstrate the constraint that we only perceive one interpretation at a time. Each image appears to flip back and forth between two distinct percepts, exhibiting "multistability." Both interpretations are equally correct, but we experience one – and only one – as the concrete, definitive, and exclusive truth at each point in time. Thus, rather than being able to average over possible perceptual interpretations, we seem limited to sampling interpretations one at a time (Moreno-Bote, Knill, & Pouget, 2011). In the examples of Figure 1.1, these shifts happen quickly enough that the subjectivity of our perception is made apparent and can be tempered by rational self-reflection, but this is likely the exception, not the rule. When extrapolated to higher-order epistemic (rather than perceptual) beliefs, it is easy to imagine how such

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press



(a) Necker Cube (b) Duck-Rabbit (c) Rubin Face-Vase

Figure 1.1 Three examples of perceptual multistability. Each image can be interpreted in two ways, but we only perceive one interpretation at a time. Evidence suggests that this phenomenon extends to higher cognitive processes such as explanatory inference. See text for details.

illusions of objectivity might exacerbate interpersonal conflict between people who arrive at different conclusions from common information.

Another implication of this cognitive strategy is that sense-making proceeds much like evolution, which "tinkers" by creatively repurposing existing biological machinery to address new environmental challenges rather than redesigning each species, cell-type, organ, physiological process, or behavior from scratch. Jacob (1977) illustrates this with the examples of the woodpecker and the aye-aye (a type of lemur), both of which exploit the same ecological niche of extracting insects from small crevices in dead wood. Each has adapted to accomplish this task using features already possessed by their evolutionary forebears: the woodpecker, whose ancestors possessed beaks but lacked hands, developed an elongated beak; the aye-aye, whose ancestors lacked beaks but had hands, developed unusually long and thin fingers. While there may, in principle, be an optimal solution to this ecological problem involving neither beak nor hand, evolution cannot discover it except by way of intermediate forms. Each evolutionary lineage is, in the near term, fated to build on what it has already developed.

In much the same way, reconsidering our entire worldview in light of every new piece of information would be computationally prohibitive: we do not update globally. Instead, we take the bulk of our beliefs as given and re-examine only those that bear most directly on new evidence we encounter. In Otto Neurath's famous metaphor, knowledge is like a boat that is always on the high seas, never able to rebuild from scratch but always forced to repair using whatever existing materials are to hand (Bramley et al., 2017; Cat, 2021). In cases when our deeply held beliefs and evidence come into conflict, the latter is usually made to yield. Only rarely do we reflect upon and take stock of "core" beliefs, foundational assumptions, and axiomatic commitments; crises of faith are the exception, not the rule. Such phenomena are, to a degree, already a feature of hierarchical Bayesian models, where priors descending from higher levels can override the bottom-up flow of information (Friston & Kiebel, 2009; Tenenbaum et al., 2007; Yuille & Kersten, 2006). Note, however, that these dynamics are greatly exacerbated by the process of local updating that makes a sampling approximation to Bayesian inference computationally efficient.

The ambiguous patterns in Figure 1.1 demonstrate that we do not control certain aspects of how our brain performs inference; no matter how hard we try, it is simply impossible to see both interpretations of a multistable image simultaneously. There are, however, other aspects of sense-making that we are able to influence through deliberative choice. To the degree that these choices bear upon our use of scarce cognitive resources, we might expect them to be mediated by motivational signals. In the same way that physical exhaustion exists to force us to reckon with the physiological consequences of continued exertion, some mental states may exist to "price in" the cognitive costs of sense-making operations.

This observation suggests a possible functional reinterpretation of dogmatic thinking, which Christensen (1994, p. 69) paraphrases as the epistemological attitude that "I happen to believe it – and that's all the justification I need for continuing to believe it." The process of updating beliefs imposes costs that range from the physical (e.g., rewiring neurons) to the practical (e.g., sapping cognitive resources from other important uses). Moreover, if belief updating is a serial, step-by-step process, it will inevitably be both limited and slow (although changing one belief may then, of course, have a cascade of implications for others). This means that the tendency to maintain one's beliefs in the absence of any reason to do otherwise may be unavoidable – and possibly also normatively justified (a philosophical position that Gilbert Harman calls "general foundationalism"; Harman, 2003). In the presence of such considerations, a motivational force that pushes back against the free revision of belief might be beneficial, even if it led to epistemic distortions. Such an account might also help to explain why we are more dogmatic about certain types of beliefs than others; for example, a higher real cost of uprooting more fundamental beliefs would explain our greater resistance to questioning them.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

1.4 Implications

The fact that sense-making is a drive helps to explain why humans are so enthusiastic about many activities that make no obvious contribution to survival or reproduction – for example, reading fiction, watching films, and solving puzzles that we ourselves create. Sense-making, like other drives, originally emerged to promote biological fitness, but operates even when this function is obviously nonoperative, similarly to when we have sex using birth control or consume "empty calories." Indeed, forms of entertainment that provide the least amount of informational enrichment (and are therefore arguably the most heavily driven by sense-making, e.g., mystery novels) are often precisely structured to build and release suspense artificially. This characteristic strategy of drive buildup and consummation is seen across the many other drives we cultivate for pleasure. Take, for instance, hunger, which we actively protect by avoiding snacks that will "ruin our appetite," then tease over many courses before finally indulging. Viewed from this perspective, much of what we call culture appears, indeed, to be a grand collection of machines that produce the pleasure of sense-making through the origination, elaboration, and resolution of fascinating complexities.

All drives are, of course, imperfect regulators. A starving person exposed to unlimited amounts of food will overconsume to the point of sickness or even death; extreme levels of pain and fear can, in some cases, become counterproductive. Sense-making, likewise, is not always perfectly calibrated to the provision of long-term benefits in every situation. The mass appeal of conspiracy theories and pseudoscientific frameworks, as exemplified by the widespread rejection of life-saving vaccines against COVID-19, illustrate sense-making taken beyond the point of functionality. The potential for sense-making to reach dysfunctional levels is also vividly illustrated by delusional schizophrenia, which is marked by a tendency to attribute "too much" coherence to meaningless or inconclusive information while dismissing contrary evidence (DSM-5, 2013; McLean, Mattiske, & Balzan, 2017). Of note, individual differences in conspiracy-mindedness and schizotypal personality disorder are interrelated (Bruder, Haffke, Neave, Nouripanah, & Imhoff, 2013; Darwin, Neave, & Holmes, 2011), raising the intriguing possibility that shared cognitive foundations may help explain these excesses of sense-making.

The rapid proliferation of new digital information technologies poses both great promise and great peril for sense-making. On one hand, the

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

"information explosion" occasioned by the rise of the Internet has drastically expanded access to nourishment for sense-making, ranging from the most extensive encyclopedia in history (Voß, 2005) to tens of thousands of digitized books (Coyle, 2006) and millions of user-generated videos (Cheng, Dale, & Liu, 2008). At the same time, social media has also greatly increased our exposure to the sense-making produced by others, leading to a dense cross-fertilization of ideas and the almost instantaneous transmission of new insights between people, leading to a kind of globalization of knowledge.

On the other hand, changes to the topological structure of communication have profoundly disrupted how sense-making flows through society, often in troubling ways. Homophily (the preferential tendency for similar individuals to form network connections; McPherson, Smith-Lovin, and Cook (2001)) combined with the sense-making distortions introduced by belief-based utility (as discussed in Section 1.2) has led to concern about the emergence of online "echo chambers": massive networks of individuals who see and propagate information or explanations that corroborate their existing worldview with little critical feedback (Bakshy, Messing, & Adamic, 2015; Colleoni, Rozza, & Arvidsson, 2014; Sunstein, 2002; but see also the moderating evidence of Dubois & Blank, 2018; Flaxman, Goel, & Rao, 2016). These effects are, no doubt, exacerbated by the failure of individuals to take account of just how biased their media diet is (Enke & Zimmermann, 2019; Eyster & Rabin, 2014; Pronin, Lin, & Ross, 2002; Vallone, Ross, & Lepper, 1985), worsening the recalcitrance and illusions of objectivity already inherent to our inferential cognition.

Even outside echo chambers, certain of these belief dynamics threaten to reduce diversity in how we interpret the world. Rather than, quite literally, thinking for ourselves, the proliferation of public commentary has made it all too easy to simply adopt the explanations of the people and media we surround ourselves with. This is appealing to each individual in the short run, as doing so yields an immediate sense-making boost with little cognitive investment, but it is potentially disastrous to the health of social discourse as a whole, which depends on the diversity of public opinion.

For all its many benefits, the rapid democratization of information and mass communication has also had the side-effect of destabilizing mechanisms that societies have historically relied upon to filter information and vet explanations. Those who in ages past claimed privileged sense-making authority – most notably academic scholars, religious leaders, journalists, and political officials – are now frequently reduced to shouting their opinions over the din of popular commentary.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

Increasingly, the most valuable commodity in the marketplace of ideas is not a reputation for careful consideration, but rather the sheer ability to garner attention (Heath & Heath, 2007). Newly ascendant counter-normative belief communities fueled by these dynamics – antivaxxers, climate deniers, flat earthers, conspiracy theorists, and religious extremists – have begun to undermine the ability of social institutions to function properly by out-competing their traditional counterparts when it comes to harnessing public attention, and, with it, opinion.

As the sense-making landscape has been upended, those who seek to influence society's understanding have adapted their strategies to take advantage of the new opportunities it provides. Technology has created an increasingly sophisticated set of tools that grant the ability to precisely target and massively amplify both the dissemination (Goldfarb, 2014; Kramer, Guillory, & Hancock, 2014) and the suppression (Bamman, O'Connor, & Smith, 2012) of information. These efforts are becoming increasingly sophisticated now that insights from behavioral and data science are being applied to predict what will engage and persuade us (Matz, Kosinski, Nave, & Stillwell, 2017; Zarouali, Dobber, De Pauw, & de Vreese, 2020).

These intra- and interindividual-level processes that determine the direction of sense-making can have profound consequences for society. As highlighted in George Marshall's (2015) insightful treatise *Don't Even Think About It: Why Our Brains Are Wired to Ignore Climate Change*, our collective ability to grapple with existential problems facing humanity depends on how we collect and make sense of information. Different nations' success in combating the coronavirus pandemic has likewise been affected by the sense that citizens have made of the virus and of interventions intended to stem its spread, often in ways that link to wider political attitudes and group affiliations.

The analysis of sense-making's cognitive foundations that we have pursued in this chapter is only the start of a much broader intellectual project, one which will involve an analysis of how the quirks of our fixed sense-making capacity can be deceived by shifting environmental forces, especially those created by technological advances. As has been illustrated by some of the recent trends that have emerged from this dynamic, the psychological foundations of sense-making have far-reaching consequences for society that we are only just beginning to understand.

References

Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status enhancement account of overconfidence. *Journal of Personality and Social Psychology*, 103(4), 718.

Aragones, E., Gilboa, I., Postlewaite, A., & Schmeidler, D. (2005). Fact-free learning. *American Economic Review*, 95(5), 1355–1368.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, 76(2), 451–469.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132.

Bamman, D., O'Connor, B., & Smith, N. (2012). Censorship and deletion practices in Chinese social media. *First Monday*, *17*(3). https://doi.org/10.5210/fm.v17i3.3943.

Batson, C. D. (1975). Rational processing or rationalization? The effect of disconfirming information on a stated religious belief. *Journal of Personality and Social Psychology*, 32(1), 176.

Beatty, M. J. (1988). Situational and predispositional correlates of public speaking anxiety. *Communication Education*, 37(1), 28–39.

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, 52(2), 336–372.

Bellman, R. E. (1957). *Dynamic programming*. Princeton University Press.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The annals of mathematical statistics*, 24(2), 265–272. www.jstor.org/stable/2236332.

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, 124(3), 301.

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy mentality questionnaire. *Frontiers in Psychology*, 4, 225.

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2018). Large-scale study of curiosity-driven learning. *arXiv preprint arXiv:1808.04355*.

Cabanac, M. (1971). Physiological role of pleasure. *Science*, 173(4002), 1103–1107.

Cat, J. (2021). Otto Neurath. In Zalta, E. N. (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entries/neurath/.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, 103(3), 566.

Chater, N. (2019). *The mind is flat*. Yale University Press.

Chater, N., & Loewenstein, G. (2016). The under-appreciated drive for sensemaking. *Journal of Economic Behavior & Organization*, 126, 137–154.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, 7(1), 19–22.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

Chater, N., Zhu, J., 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.

Chen, S., & Heese, C. (2021). "Fishing for Good News: Motivated Information Acquisition." CRC TR 224 Discussion Paper Series crctr224_2021_223v3, University of Bonn and University of Mannheim, Germany.

Cheng, X., Dale, C., & Liu, J. (2008). Statistics and social network of YouTube videos. In *2008 16th International Workshop on Quality of Service* (pp. 229–238). IEEE. https://ieeexplore.ieee.org/document/4539688.

Christensen, D. (1994). Conservatism in epistemology. *Noûs*, 28(1), 69–89.

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317–332.

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences*, 26(2), 139–153.

Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Topics in Cognitive Science*, 8(1), 160–179.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In Lewis, M. & Haviland-Jones, J. M. (eds.), *Handbook of Emotions* (pp. 91–115). Guilford.

Coyle, K. (2006). Mass digitization of books. *The Journal of Academic Librarianship*, 32(6), 641–645.

Cushman, F. (2020). Rationalization is rational. *Behavioral and Brain Sciences*, 43, E28.

Damasio, A. R. (2006). *Descartes' error*. Random House.

Darwin, H., Neave, N., & Holmes, J. (2011). Belief in conspiracy theories. The role of paranormal belief, paranoid ideation and schizotypy. *Personality and Individual Differences*, 50(8), 1289–1293.

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, 96, 1–25.

Dayan, P. (1998). A hierarchical model of binocular rivalry. *Neural Computation*, 10(5), 1119–1135.

Deci, E. L., & Ryan, R. M. (1981). Curiosity and self-directed learning: The role of motivation in education. In Katz, L. (Ed.), *Current topics in early childhood education* (Vol. 4). Ablex Publishing Co.

DSM-5. (2013). *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729–745.

Ely, J., Frankel, A., & Kamenica, E. (2015). Suspense and surprise. *Journal of Political Economy*, 123(1), 215–260.

Enke, B., & Zimmermann, F. (2019). Correlation neglect in belief formation. *The Review of Economic Studies*, 86(1), 313–332.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic perspectives*, 30(3), 133–140.

Eyster, E., & Rabin, M. (2014). Extensive imitation is irrational and harmful. *The Quarterly Journal of Economics*, 129(4), 1861–1898.

Festinger, L. (1957). *A theory of cognitive dissonance* (Vol. 2). Stanford University Press.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298–320.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1211–1221.

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology*, 3, 130.

Gershman, S., Vul, E., & Tenenbaum, J. B. (2009). Perceptual multistability as Markov chain Monte Carlo inference. In *Advances in neural information processing systems 22* (pp. 611–619). https://proceedings.neurips.cc/paper/2009/hash/692f93be8c7a41525c0baf2076aecfb4-Abstract.html.

Gilovich, T. (2008). *How we know what isn't so*. Simon and Schuster.

Goldfarb, A. (2014). What is different about online advertising? *Review of Industrial Organization*, 44(2), 115–129.

Golman, R., & Loewenstein, G. (2018). Information gaps: A theory of preferences regarding the presence and absence of information. *Decision*, 5(3), 143.

Gopnik, A. (1998). Explanation as orgasm. *Minds and Machines*, 8(1), 101–118.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, 17(11), 585–593.

Griffiths, T. L., & Tenenbaum, J. B. (2003). Probability, algorithmic complexity, and subjective randomness. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 25). https://escholarship.org/uc/item/6ts3j7bw.

Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science*, 21(4), 263–268.

Hanoch, Y. (2002). "Neither an angel nor an ant": Emotion as an aid to bounded rationality. *Journal of Economic Psychology*, 23(1), 1–25.

Harman, G. (2003). Skepticism and foundations. In Luper, S. (ed.), *The skeptics: Contemporary essays* (pp. 1–11). Routledge.

Heath, C., & Heath, D. (2007). *Made to stick: Why some ideas survive and others die*. Random House.

Jacob, F. (1977). Evolution and tinkering. *Science*, 196(4295), 1161–1166.

Jeffrey, R. C. (2004). *Subjective probability: The real thing*. Cambridge University Press.

Jepma, M., Verdonschot, R. G., Van Steenbergen, H., Rombouts, S. A., & Nieuwenhuis, S. (2012). Neural mechanisms underlying the induction and

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

relief of perceptual curiosity. *Frontiers in Behavioral Neuroscience*, 6, 5. www.frontiersin.org/article/10.3389/fnbeh.2012.00005.

Jern, A., Chang, K.-M. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review*, 121(2), 206.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, 39(1), 17–31.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, 20(8), 963–973.

Kehlmann, D. (2009). *Measuring the world: A novel*. Vintage.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, 88(3), 449–460.

Kleine, D. (1990). Anxiety and sport performance: A meta-analysis. *Anxiety Research*, 2(2), 113–131.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790.

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

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, 36(6), 661–679.

Kwisthout, J. (2011). Most probable explanations in Bayesian networks: Complexity and tractability. *International Journal of Approximate Reasoning*, 52(9), 1452–1469.

Levy, R. P., Reali, F., & Griffiths, T. L. (2009). Modeling the effects of memory on human online sentence processing with particle filters. In *Advances in neural information processing systems* (pp. 937–944). https://cocosci.princeton.edu/tom/papers/sentencepf1.pdf.

Liberman, A., & Chaiken, S. (1992). Defensive processing of personally relevant health messages. *Personality and Social Psychology Bulletin*, 18(6), 669–679.

Litman, J. (2005). Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition & Emotion*, 19(6), 793–814.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116(1), 75.

Loewenstein, G., & Molnar, A. (2018). The renaissance of belief-based utility in economics. *Nature Human Behaviour*, 2(3), 166–167.

Lopes, M., Lang, T., Toussaint, M., & Oudeyer, P.-Y. (2012). Exploration in model-based reinforcement learning by empirically estimating learning progress. In Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q. (eds.),

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

Advances in neural information processing systems (pp. 206–214). Curran Associates, Inc.

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.

MacLeod, W. B. (1996). Decision, contract, and emotion: Some economics for a complex and confusing world. *Canadian Journal of Economics*, 29(4), 788–810.

Markey, A., & Loewenstein, G. (2014). Curiosity. In Linnenbrink-Garcia, L. (ed.) *International handbook of emotions in education* (pp. 228–245). Routledge.

Marshall, G. (2015). *Don't even think about it: Why our brains are wired to ignore climate change*. Bloomsbury Publishing USA.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, 114(48), 12714–12719.

McCabe, M. P. (2005). The role of performance anxiety in the development and maintenance of sexual dysfunction in men and women. *International Journal of Stress Management*, 12(4), 379.

McLean, B. F., Mattiske, J. K., & Balzan, R. P. (2017). Association of the jumping to conclusions and evidence integration biases with delusions in psychosis: A detailed meta-analysis. *Schizophrenia Bulletin*, 43(2), 344–354.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415–444.

Meder, B., Nelson, J. D., Jones, M., & Ruggeri, A. (2019). Stepwise versus globally optimal search in children and adults. *Cognition*, 191, 103965.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences*, 108(30), 12491–12496.

Muramatsu, R., & Hanoch, Y. (2005). Emotions as a mechanism for boundedly rational agents: The fast and frugal way. *Journal of Economic Psychology*, 26(2), 201–221.

Pashler, H. (1999). *The psychology of attention*. MIT Press.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 16– 17). http://proceedings.mlr.press/v70/pathak17a/pathak17a.pdf.

Pluck, G., & Johnson, H. (2011). Stimulating curiosity to enhance learning. *GESJ: Education Sciences and Psychology*, 2(19). ISSN 1512-1801.

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.

Rabin, M., & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1), 37–82.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

Rapoport, A., & Chammah, A. M. (1966). The game of chicken. *American Behavioral Scientist*, 10(3), 10–28.

Ruan, B., Hsee, C. K., & Lu, Z. Y. (2018). The teasing effect: An underappreciated benefit of creating and resolving an uncertainty. *Journal of Marketing Research*, 55(4), 556–570.

Samuelson, L., & Swinkels, J. M. (2006). Information, evolution and utility. *Theoretical Economics*, 1(1), 119–142.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, 20(12), 883–893.

Savage, L. J. (1972). *The foundations of statistics*. Courier Corporation.

Schelling, T. C. (1980). *The strategy of conflict: With a new preface by the author*. Harvard University Press.

Schmidhuber, J. (1991). A possibility for implementing curiosity and boredom in model-building neural controllers. In Meyer, J. A. and Wilson, S. W. (eds.), *Proc. of the international conference on simulation of adaptive behavior: From animals to animats* (pp. 222–227). MIT Press/Bradford Books.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, 40, 99–124.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366.

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 79(5), 369.

Sorg, J., Singh, S. P., & Lewis, R. L. (2010). Internal rewards mitigate agent boundedness. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (pp. 1007–1014). https://icml.cc/Conferences/2010/papers/442.pdf.

Sun, Z., & Firestone, C. (2020). The dark room problem. *Trends in Cognitive Sciences*, 24. https://doi.org/10.1016/j.tics.2020.02.006.

Sunstein, C. R. (2002). The law of group polarization. *The Journal of Political Philosophy*, 10(2), 175–195.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tenenbaum, J. B., Griffiths, T. L., & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In *Causal learning: Psychology, philosophy, and computation*, 301–322. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780195176803.001.0001/acprof-9780195176803-chapter-20.

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.

Van Rooij, I. (2008). The tractable cognition thesis. *Cognitive Science*, 32(6), 939–984.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press

Voß, J. (2005). Measuring Wikipedia. In *International Conference of the International Society for Scientometrics and Informetrics: 10th, Stockholm (Sweden), 24–28 July 2005* (pp. 221–231). http://eprints.rclis.org/6207/.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, 38(4), 599–637.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, 19(7), 645–647.

Wade, S., & Kidd, C. (2019). The role of prior knowledge and curiosity in learning. *Psychonomic Bulletin & Review*, 26(4), 1377–1387.

Wojtowicz, Z., Chater, N., & Loewenstein, G. (2020). Boredom and flow: An opportunity cost theory of attention-directing motivational states. *Available at SSRN 3339123*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3339123.

Wojtowicz, Z., & DeDeo, S. (2020). From probability to consilience: How explanatory values implement Bayesian reasoning. *Trends in Cognitive Sciences*, 24(12), 981–993.

Wojtowicz, Z., & Loewenstein, G. (2020). Curiosity and the economics of attention. *Current Opinion in Behavioral Sciences*, 35, 135–140.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, 10(7), 301–308.

Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2020). Using a personality-profiling algorithm to investigate political microtargeting: Assessing the persuasion effects of personality-tailored ads on social media. *Communication Research*, 0093650220961965.

Zeidner, M. (2010). Test anxiety. In Weiner, I. B., & Craighead, W. E. (eds.), *The Corsini encyclopedia of psychology*, pp. 1–3. John Wiley.

https://doi.org/10.1017/9781009026949.002 Published online by Cambridge University Press