Cambridge University Press & Assessment
978-1-316-51590-7 — The Drive for Knowledge
Edited by Irene Cogliati Dezza , Eric Schulz , Charley M. Wu
Excerpt
More Information

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 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 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

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.

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.

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" 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.

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