All Thinking is "Wishful" Thinking

Arie W. Kruglanski

University of Maryland

Katarzyna Jasko

Jagiellonian University

Karl Friston

University College London

Correspondence: kruglanski@gmail.com (A. Kruglanski)

Abstract

Whether one looks at ethology, social neuroscience, behavioral psychology, machine learning, or the societal implications of ‘fake news’, questions about epistemic motivation keep emerging. While people often seek new information and eagerly update their beliefs, at other times they avoid information or resist revising their knowledge. So, what are the principles behind our truth-seeking? We argue, from psychological and optimality principles, there is no ‘truth’: optimal belief updating has a mixed motivation base, where epistemic affordances contextualize prior beliefs. This perspective offers a nuanced picture of knowledge-seeking and updating – that can be underwritten by formal (active inference) accounts.

Keywords: epistemic motivation, active inference, surprise

*'No rational argument will have a rational effect on a man who does not **want** to adopt a rational attitude"* Karl R. Popper

Cognition vs. Motivation

As we go about our daily activities, we confront a perennial flow of information. Incoming news can be surprising when it violates previously held expectations, but it can also fit perfectly with what we anticipated. In some cases, it prompts vigorous action on our part; in others it elicits little if any response whatsoever. At times, the information is pleasing; it constitutes "good news" we are happy to receive. At other times, it is disturbing, affirming our deepest fears. What then explains how we react to information and what we do about it?

The answer to this question has often depended on a respondent's position in a debate between advocates of cognition vs. motivation. The dichotomy that separates humans into motivational and cognitive goes back thousands of years, at least to Aristotle's distinction between the rational soul – residing in the head (and responsible for cognition), and the appetitive soul – residing in the chest (pertaining to motivation). A heated debate over the primacy of cognition versus motivation has been ongoing in social psychology from the 1970s and was prominently exemplified by the debate between proponents of dissonance versus self-perception theory [e.g., 1-2], and the question of whether motivational biases in attribution are fact or fiction [3]. Advocates of cognitive dominance argued that information processing, while prone to errors due to limited capacity, is nonetheless predominantly guided by the accuracy principle. In reaction to such views of humans as cognitive machines, proponents of the motivational approach responded with a long list of motivated biases, which hinder accurate representations of reality. In this paper, we try to dissolve this dichotomy.

We approach the question of cognition and motivation from two theoretical perspectives represented by the notions of *epistemic motivation* [4] and *active inference* [5]. However, instead of

separating cognition and motivation into two separate systems, we propose that any epistemic activity is thoroughly suffused by motivation. The fundamental motivational substrate for any instance of epistemic behavior has been characterized in different ways. In the theory of epistemic motivation, a distinction is drawn between *nonspecific* and *specific certainty* (i.e., cognitive closure). In active inference, a distinction is made between *accuracy* and *complexity* – that together constitute Bayesian model evidence. These two perspectives on the role of motivation in cognition are juxtaposed and interrelated in the pages that follow. Both attest to indispensability of motivational considerations to an adequate theory of knowledge formation.

To preview our story, we first characterize the process of inference, the quintessential mechanism of knowledge construction. Next, we describe the role of motivation in this process. We then identify the parallels between active inference and epistemic motivation and explore their complementarities. Finally, we identify possible applications of their joint epistemic perspective.

The Process of Inference

What does the epistemic process entail? Basically, we construct new beliefs from prior beliefs by assimilating new evidence. We do so through an inference process probabilistically modeled by Bayesian principles [6-7]. According to that portrayal, relevant evidence—to which we are exposed—occasions an updating of our beliefs on the topic. In Bayesian belief updating, two components are crucial: 1) the strength of the prior belief; namely, the subjective probability of it being true, and 2) the cogency of the new evidence; namely, the degree to which it strengthens or weakens prior beliefs. In other words, people update their prior beliefs given new evidence, depending on whether the new evidence is perceived as precise, strong and relevant (vs. imprecise, weak and irrelevant) and whether their prior belief was held with high (vs. low) confidence. The change in prior beliefs—in light of the new evidence—is quantified by the informational gain or (Bayesian) surprise [8]. Substantial belief updating in response to a surprising event indicates that one's cognitive model of the world did not

match the reality, as it did not predict the event in question. Similarly, a lack of surprise means that an external event was fully predicted by the internal model.

However, people are not merely passive recipients of information, and belief updating based on incoming information is only part of the epistemic process—people also actively seek to obtain, avoid, or create new information about the world to increase the consistency between their models and the evidence at hand.

Epistemic Motivation

The Bayesian principle of belief updating, as such, is agnostic with regard to motivation. It addresses the *how* of knowledge formation rather than the *why* or *wherefore*. Yet the epistemic process in all of its phases is suffused by motivation as noted earlier (see Figure 1). Motivation affects the degree of confidence in one's prior beliefs about a subject [9]. It enters into what information is considered diagnostic and relevant, and thus serves as 'evidence' [10]. It affects the extent to which the evidence is seen to imply a given conclusion [11]. It influences the readiness to accept that implication uncritically and update one's beliefs in light of the new evidence [12]. It determines the magnitude of (positive or negative) affective reaction to the updated belief [13]. Finally, it affects the inclination to truncate the search and form a judgment versus continuing to seek further evidence [4, 14]. Motivation does so essentially by instilling a sense of doubt or quelling it. The former is affected by bringing up contrary evidence—inconsistent with the initial implication. The latter is accomplished by suppressing such evidence, inhibiting its entry into awareness, and/or producing supportive evidence that bolsters one's conclusion (see Figure 2). Four basic motivations energize and guide the process of knowledge construction: The imperatives to approach and avoid *nonspecific certainty* and to approach and avoid *specific certainty*. We discuss them next.

Approaching and Avoiding Nonspecific Certainty

The approach of nonspecific certainty entails the search for a firm, precise answer to a question, regardless of its specific content. For instance, a traveler might wish to be certain about the departure gate of her flight without any preference for a particular gate. A student may want to be certain about the date of a historic event without favoring a particular date over others. The complementary avoidance of nonspecific certainty pertains to cases where certainty is eschewed, and judgmental non-commitment is at premium. At times, uncertainty is valued as a means of keeping an open mind and avoiding premature (erroneous) closure. Judgmental non commitment also affords freedom from constraints and the associated sense of possibility and adventure that epistemic commitment eschews. For example, one may not wish to know what the future would bring so as to be able to indulge in various fantasies [15] or prefer not to learn the ending of a movie by avoiding spoilers, because knowing how events unfolded requires giving up on alternative scenarios. Certainty that one option is true means that all others are false. If someone knows what happens, alternative endings would become impossible to entertain and the epistemic affordance—of continuing to watch the movie—is vitiated.

Approaching and Avoiding Specific Certainty

Often, individuals crave specific certainty concerning beliefs they find reassuring, flattering or otherwise pleasing. A student may prefer to know that she passed an exam, a patient may prefer to receive a clean bill of health, a suitor may prefer to have their affections returned. Similarly, one may avoid specific certainties that are troubling or threatening. Not knowing that one failed an exam is more pleasant than knowing that one did. Agnosticism concerning the alleged misconduct of one's child is preferable to unpleasant certainty in this matter. Avoidance of specific certainty can sometimes lead to motivated ignorance persons may occasionally cherish [16]. Motivations for specific outcomes, or specific approach or avoidance goals (e.g., passing an exam, or avoiding illness) correspond to the epistemic motivation to *believe* that those outcomes indeed materialized. Such motivations may

selectively affect the foraging for information on these matters with an eye to increasing the likelihood of the desired conclusion and decreasing that of arriving at undesired beliefs.

The Structure of Motivation

Each of the foregoing epistemic motivations has a magnitude or intensity. One may approach (or avoid) specific or nonspecific certainty more or less fervently. Motivational magnitude in turn depends on the perceived costs and benefits of having or lacking confident beliefs as such or beliefs of a particular content. To illustrate that point, a police detective may want to be certain about the identity of a criminal but less so than the victim.

There is a sense in which the quest for certainty (whether of specific or nonspecific kind) implies the desirability of *simplicity* or *parsimony*. In questing for certainty, one wishes one's knowledge to be unqualified, that is, uncomplicated by moderating conditions. The wish to know that one's bill of health was clean, for instance, implies the wish that it would be so unconditionally, plain and simple.

Typically, an individual's epistemic process is driven by a motivational 'cocktail' of epistemic drives. For instance, if the imperative for nonspecific certainty was stronger than for specific certainty, then an undesirable inference (i.e., inconvenient truth) would be likely. Much as one would cherish the consequence that one's medical exam yielded a clean bill of health, one's wish for nonspecific certainty on these matters may be even higher. Willy nilly, one would accept the undesirable conclusion that one was diagnosed with an illness. If, however, the need to avoid a specific certainty (i.e., that one failed or was terminally ill) was more dominant, a state of uncertainty would be preferred if it helped to maintain 'blissful ignorance'. In such cases, someone might avoid new information or engage in fantasy or rationalization in order to maintain their prior beliefs. For instance, based on the same evidence, fans of opposing teams might conclude that theirs played a fairer, more sportsmanlike, game than their adversary [17]. These processes are illustrated by discussion of the relative strength of self-verification vs. self-enhancement motives. In some situations, the self-verification motive may override the self

enhancement motive. A person with a strong prior negative view of self may prefer negative information about the self over positive information because such information is considered to be more certain [18]. In other situations, the self enhancement motive may have the upper hand and a person would prefer a positive information about the self that would gratify their need for specific certainty.

In summary**,** human epistemic behavior is shaped by four types of motivation, representing the needs for approaching versus avoiding nonspecific and specific certainty. These motivations exert their ubiquitous effects, "sotto terra" as it were, below conscious awareness. This is so because, motivation (wishing something) is not a recognized reason for believing it ("wishing does not make it so"). Our beliefs are expected to be based on evidence rather than on our wishes. One implication of this notion (amenable to empirical verification) is that making an individual aware that their beliefs were influenced by their wishes, should reduce their confidence in those beliefs.

Whereas the term 'wishful thinking' is typically reserved for the motivation for specific certainty, the motivations to avoid it, or to have (or avoid) nonspecific certainties can be no less motivating or "wished for", supporting our claim that "all thinking is wishful thinking"; in other words, all thinking is *motivated*. In that sense then, there is no dichotomy between a "rational" cognitive process and motivated cognition.

Active Inference

To lend our arguments formal depth, we will show that their tenets emerge from 'first principle' accounts of self-evidencing [19] that apply to any creature faced with the problem of surviving in a capricious and changing world. Epistemic behavior is at the heart of active inference, which casts all action and perception as an optimization process. To be actively foraging for information requires a driving force that initiates and maintains behavior. Epistemic motivation is one such driving force.

According to the active inference framework, the brain creates an internal model of external reality and uses it to form predictions about what will happen next. This model is updated on the basis

of sensory evidence, from which brain infers external 'states of affairs' [20]. In other words, when people perceive the world, they act like a scientist who is trying to build a model of how causes (i.e., external or hidden states) generate sensory evidence or data [21-22]. Much like a scientist, the quality or evidence for a particular hypothesis (i.e., model) depends upon the accuracy of the predictions it makes and, importantly, how parsimonious those predictions are. Technically, the (log) evidence for a model is accuracy minus complexity; implying that the brain is trying to provide an accurate account of the sensory evidence it encounters that is minimally complex. Note the affinity of this conception to the quest for certainty construct in the epistemic motivation model: To feel that one's knowledge is accurate is to be certain of it. Certainty, in turn, implies simplicity or the unqualified nature of the particular knowledge at hand.

This is a revealing because it means to be Bayes optimal, not only do we have to find accurate explanations for data, these explanations have to be parsimonious (c.f., Occam's principle: see Box 1). In short, we seek explanations that are as simple as possible, where simplicity reflects the unqualified, hence uncomplicated, nature of our knowledge. In this sense, optimal belief updating has little to do with 'truth' *per se*; indeed, one could argue that there is no 'true' model; just the 'best' model, under some prior beliefs. Crucially, this optimization process can be accomplished in one of two ways: we can either update our beliefs to provide a better model of—or explanation for—sensory evidence. Alternatively, we can act on—or sample—the sensorium to solicit sensory evidence that fits with our predictions [23-24]. This is the important challenge of active inference; namely, the problem of knowing which data to sample, even before we make sense of those data.

Active inference has proved itself in many domains within cognitive neuroscience, ethology and beyond (see Box 1). However, at first glance, it fails to account for cases where we find Bayesian surprise attractive. We often seek out experiences that are salient and novel, and we relish 'pleasant surprises' that defy our pessimistic expectations. These phenomena may seem paradoxical because self-evidencing

requires our experiences to be unsurprising, whereas epistemic behavior would appear to maximize Bayesian surprise. We will see below that this apparent paradox has a straightforward and revealing solution, which rests upon the fact that we actively choose how we sample the world. In this, enactive setting, we forage for information that minimizes *expected inaccuracy* and *expected complexity*; known as *ambiguity and risk,* respectively (see Box 3). A key aspect of active inference is the prescription of choices via a single imperative; namely, to minimise expected free energy or uncertainty. Crucially, expected free energy combines epistemic and pragmatic imperatives in the form of ambiguity and risk. This means that goal-seeking and information-seeking become two sides of the same coin.

Epistemic Motivation in Active Inference

Ambiguity and Approaching Nonspecific Certainty

As already noted**,** the concept of epistemic motivation is fundamental to the very notion of active inference. Minimizing *ambiguity*; namely, maximizing the subjective certainty of an outcome maps nicely onto need for *nonspecific certainty*. Individuals animated by the need for nonspecific certainty would sample any relevant evidence of which they became aware. Issuing as it does from the realm of visual neuroscience, active inference naturally highlights the need for nonspecific certainty. After all, in orienting oneself in physical space what matters is nonspecific certainty that would allow one to navigate one's environment without stumbling and bumping into objects. In short, the first thing we do on entering a dark room, is to switch on the light, minimizing the ambiguity of visual sensations—and resolving nonspecific uncertainty.

In active inference, the motivation for nonspecific certainty would lead the actor to choose a policy that minimizes expected free energy – or maximizes expected model evidence. The stronger the actor's motivation for nonspecific certainty, the stronger their tendency to select such an ambiguity resolving action. These notions are formalized by the expected free energy under our (predictive) beliefs about the consequence of a given course of action. Mathematically, this means that the certainty seeking

actor would choose policies that maximize expected accuracy (i.e., certainty), while minimizing expected complexity (i.e., ego-dystonic consequences). There are a number of neuronal process theories that describe this kind of self-evidencing. Perhaps most popular is predictive coding, in which prediction errors from the sensory cortex are propagated up cortical hierarchies to elicit descending predictions, which try to suppress prediction errors at lower hierarchical levels [see 25, 26-29].

Risk and Approaching and Avoiding Specific Certainty

Individuals animated by the need for *specific certainty* may selectively sample data that confirm their desired conclusions (specific certainties) and avoid data that would be inconsistent with those conclusions. Such biased information foraging is allowed in the active inference model particularly when it protects an individual from a radical revision of their models, which would increase the predicted complexity of those models. In statistics and machine learning, minimizing complexity ensures generalization; in other words, it precludes overfitting the data at hand. When people hold strong beliefs, information that is consistent with those beliefs will be sampled and information that is inconsistent and, because of that, risky will be avoided. In other words, when there is little uncertainty about states of affairs in the world, there will be little interest in further epistemic foraging—and instead the agent would engage in goal-pursuit based on their subjectively precise knowledge about the outcomes they prefer. In such a situation pragmatic, goal-directed, risk-minimizing behavior predominates—driven by our prior preferences and beliefs [30-31].

Consider the example of Sally, who entertains two hypotheses; namely, the planet will be destroyed by climate change or 'everything is fine'. How would she act? That depends on the strength or precision of her beliefs about the consequences of action. If Sally does not have precise preferences (i.e., prior beliefs) about whether the world will end—and she perceives scientific authorities as a reliable source of information—she will forage for information and update her beliefs accordingly. However, if Sally had precise beliefs that climate catastrophe will not materialize, she will find foraging

for information a risky business—and tend to ignore evidence that is inconsistent with her prior beliefs. The situation changes subtly if she (believes she) is in a position to make a geopolitical difference that matters. In this setting, her policies may include outcomes that are consistent with her prior preferences that 'everything is fine'—if she can change things. In other words, instead of ignoring evidence, she could actively create new evidence making her preferences self-fulfilling prophecies.

Importantly, while both models assume that an actor may end up with a belief that is biased by their preferences there is a difference in how they conceptualize those preferences. Specifically, in the lay epistemic model, the two concepts – prior beliefs about states of the world and preferences – are separable; in the sense that one can believe *a priori* that something is highly likely but still prefer that it did not occur. Thus, sometimes prior beliefs are in line with desired beliefs (e.g., 'I think it's going to be sunny and I want it to be sunny'), but at other times prior beliefs can diverge from desired beliefs (e.g., 'I think it's going to be rainy but I want it to be sunny'), in which case a person would be motivated to change their beliefs. In short, according to the lay epistemic model, approaching specific certainty may mean affirming one's prior beliefs (if they are pleasing or desirable) but it could also mean disconfirming them (if they are unpleasant and unwelcome).

In contrast to lay epistemic model – that refrains from assuming that prior beliefs as such have a positive valence – according to the active inference model, *some* prior beliefs do reflect preferences; these are prior beliefs about the consequences of an action. The distinction between preference in psychology and prior preference in active inference is subtle but fundamental. Active inference makes a formal distinction between prior preferences about outcomes and posterior beliefs about (preferred) courses of action. For example, a priori, I may prefer to pass my driving test. However, after assessing the evidence that I am likely to pass, I may, a posteriori, conclude I will fail – and postpone it. This is a straightforward consequence of active inference – that manifests as a form of motivated ignorance; in the sense that I avoid putting my driving competence to the test. In short, I would prefer (a priori) to

pass but I select a policy (a posteriori) that appears contrary to my preferences. This example highlights the role of risk in policy selection. Here, if I imagine the future under a policy that involves 'taking the test', my posterior belief that I am likely to fail diverges from my prior preference that I am likely to pass. This divergence is risk in active inference – and renders the posterior preference for 'taking the test' very small (see Box 3).

Risk and Avoiding Nonspecific Certainty

Whereas the active inference model appears to emphasize the epistemic quest for *certainty* (specific or nonspecific), there are cases where *ambiguity* is attractive. This is because the expected free energy comprises both risk and ambiguity – and ambiguity is the price that is often paid to avoid risk. We have already discussed why people sometimes avoid information in order to prevent the risk of revising their desired beliefs. However, people might also choose to avoid information, irrespective of its content. For instance, they may prefer to avoid movie spoilers. On the surface, this might seem inconsistent with epistemic motivation. However, if a person believes that watching a movie will provide them with two pleasant hours of epistemic foraging, they will avoid spoilers that nullify the epistemic value of a movie watching policy. Conversely, if someone does not have time to watch a movie, they might choose to watch a spoiler to resolve uncertainty about its denouement. People also may choose a policy that conserves ignorance, if obtaining new information is too costly. In active inference terms, this reflects the expected complexity cost or risk; e.g., 'if I attend this medical checkup, I may come away with a diagnosis of cancer'. In short, actions that may appear to be motivated by the preservation of ignorance are in fact means to satisfy the right balance of epistemic and pragmatic imperatives.

Motivational Substrate of Epistemic Behavior: Some General Implications

The formal consilience between epistemic motivation and active inference brings several opportunities to the table. For example, it offers a way of articulating social psychology and ethological

constructs in terms of Bayesian computations [32-33]; of the sort that could be installed into artificial intelligence. From a neurobiological perspective, the neuronal process theories that accompany active inference enable empirical predictions about epistemic motivation during belief updating—as manifest in things like event related potentials and phasic dopamine responses in the brain [34-35]. Furthermore, this convergence delineates a general process at work across all the manifold contents and instances of knowledge formation. Given such breadth, one would expect it to yield both theoretical and practical implications. We briefly exemplify one such application in the context of cognitive consistency theories.

Relevant to the role of motivation—in the epistemic process—is the recent theoretical debate about the affective reactions to cognitive consistency and inconsistency [36-38]. The notion that people universally prefer cognitive consistency to inconsistency, and that they react to inconsistency with negative affect has been a mainstay in the field of social cognition and the staple of the cognitive dissonance theory [39], one of the most impactful and highly cited frameworks in all of psychology [40]. But our present portrayal of the inference process questions the assumption of a universal human need for cognitive consistency.

Briefly, cognitive 'consistency' and 'inconsistency' between prior and posterior beliefs corresponds to the degree of Bayesian surprise: Perfectly consistent information is information that reaffirms prior beliefs. In contrast, inconsistent information requires revision of one's prior beliefs, creating a discrepancy between prior and posterior beliefs. Yet, and this is the crucial point, the affective reactions to such updates are completely determined by nature of the epistemic motivations that drive active inference in a given instance. For instance, if (consistent or inconsistent) information increased uncertainty about a given states of affairs, this would be upsetting to a knower who was motivated to approach nonspecific certainty (i.e., reduce ambiguity). In contrast, if (consistent or inconsistent) information resolved uncertainty, the (nonspecific) certainty questing knower should be happy. A failure to make (subjectively) precise inferences; either about states of affairs or the policies that 'I am

currently pursuing' can be readily associated with emotional constructs such as stress, anxiety and the like. This formulation of self-consistency in self-evidencing terms can, on one account, be unpacked in terms of psychopathology; leading to fairly detailed models of psychiatric conditions [41-42]. In short, there appears to be a tight coupling between affective aspects of epistemic behavior, stemming from irreducible uncertainty *if* the knower craved nonspecific certainty on a given topic.

But she might not. She might prefer a state of ignorance (uncertainty) on this issue that would leave her 'options open' and foreclose binding commitment to a judgment. (See [43] for a worked example and [44] for an empirical demonstration). In such a case, the knower would be pleased rather than upset by cognitive inconsistency. She might be similarly pleased if the ensuing uncertainty prevented the formation of an undesirable certainty; i.e., preventing *risk* that the latter certainty would entail. And she might be unhappy if the uncertainty precluded the formation of a specific pleasing certainty she was motivated to pursue.

All these affective reactions are moderated by the relative strength or precision of the epistemic motivations involved. In these terms, the voluminous research on cognitive consistency and inconsistency has confounded (1) the epistemic impact of belief updating and (2) the affective value of those beliefs to the knower: Reduced belief strength in a proposition that denotes a positive state of affairs (for the knower) will induce negative affect proportionately to strength of the desire to have that state of affairs come true. Increased belief strength in such proposition will induce positive affect, again proportionately to the desire to have this state materialize. Similarly, increased belief strength would induce positive affect for someone who desired certainty on a topic, and negative affect for someone who shunned such certainty again proportionately to strength of those desires. As the foregoing implies, the affective consequence of belief updating derives from the degree to which the knower's epistemic motivations were served or undermined by consistent or inconsistent information; rather than by consistency or inconsistency *as such*.

Concluding Remarks

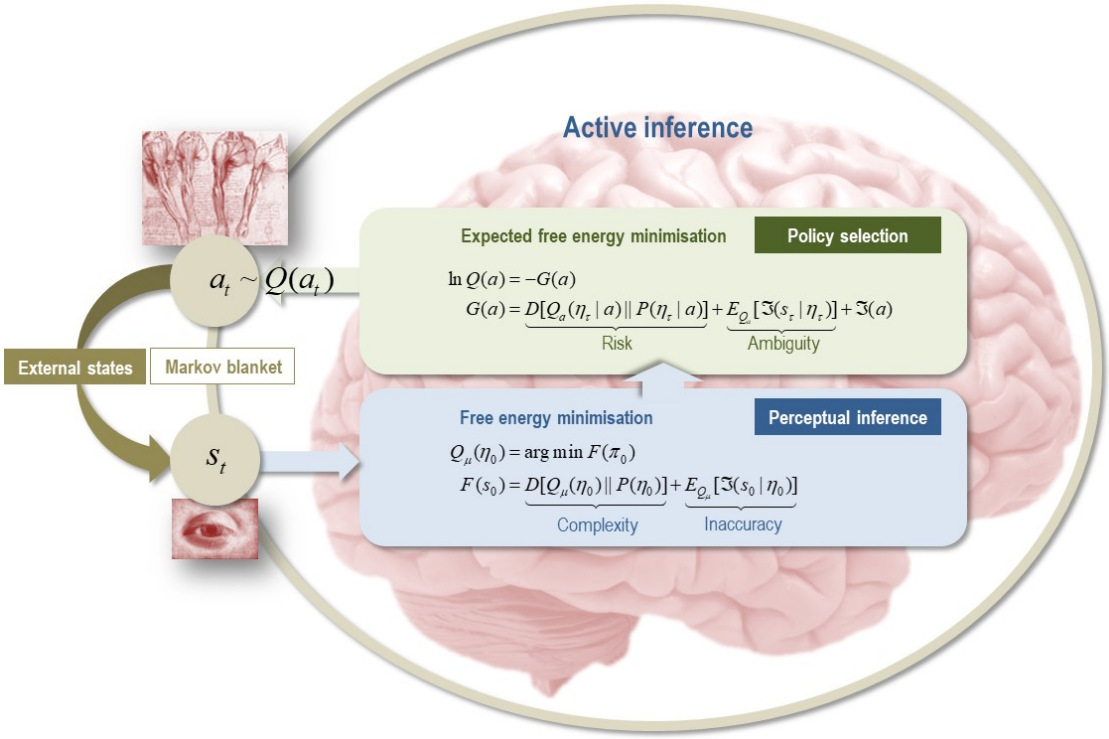
The role of epistemic motivations in shaping our judgments and world views is paramount. Because motivations vary widely across persons, situations and time, people form divergent beliefs, subscribe to opposed world views and embrace disparate ideologies. Such diversity has served humanity well in promoting creativity and precluding intellectual stagnation. Yet it also fostered conflict and rampant violence in the name of opposite subjective truths and *sine qua non* idealisms. Understanding the role of motivation is an essential step in getting people to do the 'right thing' (however defined). Talking people out of their socially destructive conceptions (e.g., in the realm of xenophobia and violent extremism) through 'rational arguments' alone will hardly work unless the motivational underpinnings of their inferences are considered and appropriately addressed.

Figure 1: **Upper panel**: this graphic describes the functional architecture that underwrites epistemic motivation in terms of the synthesis of new evidence with prior beliefs to produce posterior beliefs and subsequent affective reactions. The key point made by this figure is the circular causality induced by epistemic action—both on the prior beliefs and the way in which new evidence is solicited on the basis of affective reactions. Crucially, motivation gets into the game at nearly all levels of this process. This is illustrated by the pervasive effects of motivation on prior beliefs and the sampling of relevant information with high diagnostic value, in relation to those beliefs and their attendant uncertainty. Furthermore, motivation effects integration of evidence during the formation of posterior beliefs and the affective reaction to those posterior beliefs. Finally, motivation is also manifest in the selection of actions based on affective reactions to posterior beliefs. In turn, it determines how epistemic foraging or exploration seeks new evidence for subsequent belief updating. **Lower panel**: this shows the same architecture but described in terms of active inference. In this setting, the affective reaction has been associated with belief updating, not about hidden states (i.e., latent causes) in the world but about the policies being pursued – and their precision (i.e., confidence about 'what I am doing'). Finally, motivation has been decomposed into *risk* and *ambiguity* that corresponds to the motivational aspects of specific and nonspecific [un]certainty, described in the main text.

Table 1. Epistemic choices: the nature of uncertainty, ambiguity and risk

Certainty	**Epistemic motivation**	**Example**	***Ambiguity (expected inaccuracy)***	***Risk (expected complexity)***
Nonspecific	***Approach***	Finding out the departure gate of one's flight	minimized	–
	Avoid	Avoiding knowledge of a film's end	minimized	–
Specific	***Approach***	Hoping for a clean bill of health on the annual check up	–	minimized
	Avoid	Avoiding listening to a TV commentator opposed to one's views	–	minimized

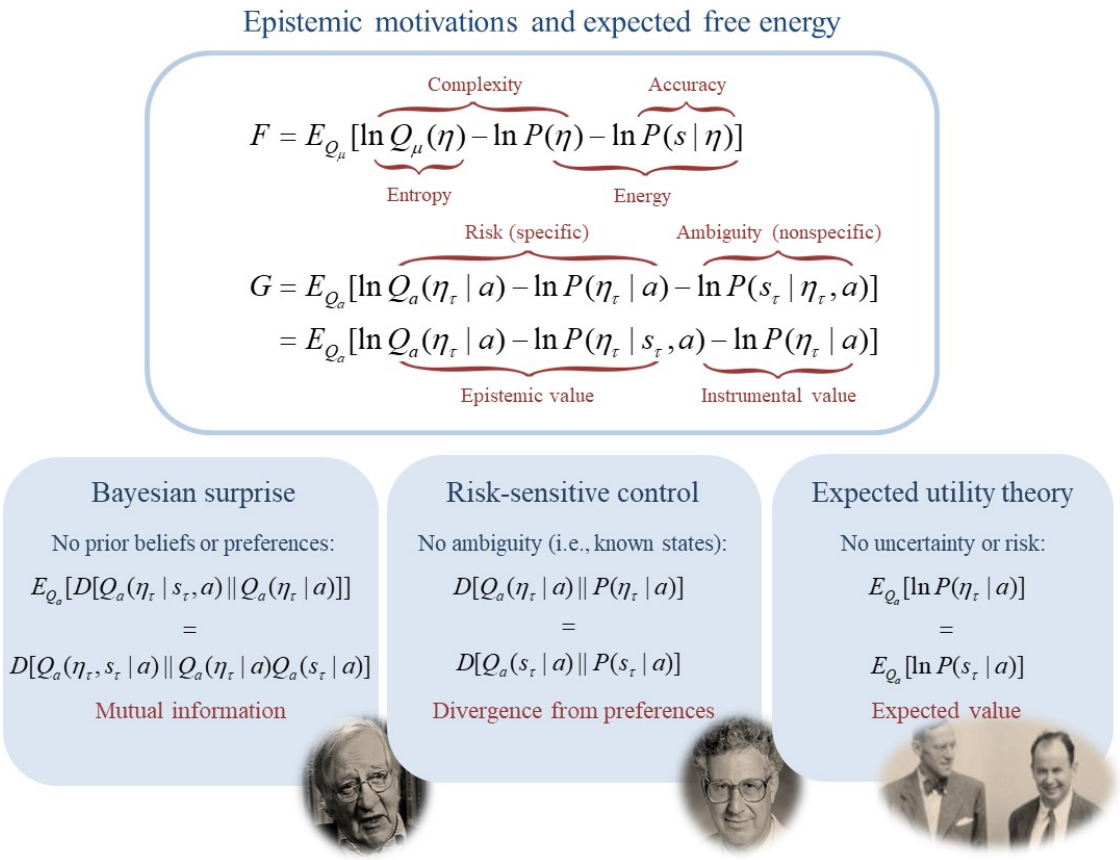
Box 2: Active inference and epistemic affordance



Bayesian belief updating and active inference. This graphic summarizes the belief updating implicit in the minimization of variational and expected free energy. It summarizes a generic (active) inference scheme that has been used in a variety of applications and simulations; ranging from games in behavioral economics [48] and reinforcement learning [43] through to language [49] and scene construction [50]. In this setup, actions solicit a sensory outcome (or evidence) that informs approximate posterior beliefs about hidden or external states of the world – via minimization of variational free energy, under a set of plausible policies (i.e., *perceptual inference*). The approximate posterior beliefs are then used to evaluate expected free energy and subsequent beliefs about action (i.e., *policy selection*). Note a subtle but important move in this construction: the expected free energy furnishes prior *beliefs about policies*. This is interesting from several perspectives. For example, it

means that agents infer policies and, implicitly, their action. In other words, beliefs about policies – encoded by internal states – are distinct from action per se. In more sophisticated schemes, agents infer hidden states under plausible policies with a generative model based on a Markov decision process. This means the agent predicts how it will behave and then verifies those predictions based on sensory samples. In other words, agents garner evidence for their own behavior and actively self-evidence. In this setting, variational free energy reflects the surprisal or evidence that a particular policy is being pursued. In sum, this means the agent (will appear to) have elemental beliefs about its enactive self – beliefs that endow it with a sense of purpose, in virtue of the prior preferences that constitute risk. Please see the glossary of terms for explanation of the variables in the equations.

Box 3. Epistemic motivations and expected free energy



Expected free energy and epistemics: This graphic illustrates the relationship between variational and expected free energy – and how the minimization of expected free energy relates to well-known schemes in the neurosciences and behavioral economics literature. The upper panel shows the variational free energy decomposed into its constituent parts. These can be variously rearranged to be interpreted in terms of *energy* minus *entropy* – or *complexity* minus *accuracy*. The second equalities are the expectations or averages of variational free energy under a predictive belief about outcomes and their causes (hidden states) in the future. These have been written to show that *expected inaccuracy* corresponds to *ambiguity*, while *expected complexity* corresponds to *risk*. We can rearrange these terms to show that this is equivalent to a mixture of epistemic value – of the sort considered in visual foraging

– and instrumental value – of the sort considered in expected utility theory. The lower panels illustrate some special cases of expected free energy to show how it relates to standard constructs in the visual neurosciences. Here, in the absence of prior preferences, the remaining epistemic value is also known as expected Bayesian surprise or salience in the visual search literature [51-52]. This is the same as the mutual information between sensory consequences and their causes – that underwrites the principle of maximum efficiency or information [53-54]. If we remove uncertainty about external states of the world, we end up with a quantity used in optimal control theory and economics called *risk* [55-56]. This is basically the difference between predicted and preferred outcomes. Finally, if we remove the last source of uncertainty; namely, the consequences of action, expected free energy reduces to expected utility found in economics [57]; where utility corresponds to the logarithm of prior preferences. In summary, standard constructs in cognitive neuroscience and economics obtain as special cases of minimizing expected free energy – as we remove successive sources of uncertainty.

Box 4. Relations between accuracy, complexity, ambiguity, and risk.

From a statistical or (active) inference perspective everything we do is in the service of maximizing the evidence for our generative models of the world. This is sometimes referred to as self-evidencing [19]. The logarithm of evidence (i.e., free energy) is accuracy minus complexity, which means that perception is in the game of maximizing accuracy, while minimizing complexity. Similarly, actions are selected to maximize accuracy and minimise complexity following an action. The expected consequences of action; namely, expected inaccuracy (i.e., ambiguity) and expected complexity (i.e., risk) are combined into expected log evidence (i.e., expected free energy). This means that Bayes optimal behavior minimizes ambiguity and risk. The relative contribution of accuracy (resp. ambiguity) and complexity (resp. risk) depends on the precision of prior beliefs (resp. preferences). Technically, accuracy is just the expected log likelihood of some data, under posterior beliefs about how those data were generated. Complexity is the degree of belief updating incurred by observing some data, as scored by the KL divergence between prior and posterior beliefs (i.e., probability distributions before and after seeing the data). Intuitively, complexity corresponds to the degrees of freedom used to explain some data, where a complex explanation uses many degrees of freedom and implicit belief updating.

Glossary

Accuracy: the expected log likelihood of some outcome, under posterior Bayesian beliefs about the causes of that outcome.

Active inference: the minimisation of variational free energy through approximate Bayesian inference and active sampling of (sensory) data. This sampling induces belief updating, under prior beliefs that sampling will minimise free energy expected in the future. This is equivalent to resolving uncertainty and maximising model evidence – sometimes called self-evidencing.

Ambiguity: the expected inaccuracy of future outcomes, under a particular policy – as measured by the conditional entropy (i.e., uncertainty) about outcomes, given their causes.

Bayesian belief: a posterior probability distribution over a random variable, such as a latent cause or hidden state of the world causing (sensory) data**.**

Belief updating: the process of statistical inference, in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available. Technically, a prior belief is updated to form a posterior belief.

Complexity: the divergence between prior and posterior beliefs; in other words, the degree to which Bayesian beliefs change before and after belief updating.

Epistemic value: the information gain expected under a particular policy. This is sometimes referred to as intrinsic motivation, Bayesian surprise or salience. Novelty is a form of salience that reflects the epistemic affordance of policies, which resolve uncertainty about the parameters of a generative model.

Expected free energy: an attribute of a policy that can be decomposed into epistemic and pragmatic value – or into risk and ambiguity (please see Box 3).

Generative model: a probability distribution over the causes of observable consequences. A generative model is usually specified in terms of a likelihood and a prior belief; namely, the probability of an outcome given its cause and the prior belief about the course.

Inference: the optimisation of beliefs by maximising model evidence. Approximate Bayesian inference corresponds to minimising variational free energy

Kullback-Leibler divergence: the difference between two probability distributions – as measured with their relative entropy.

Likelihood: the probability of observing some (sensory) data, given the causes of those data

Model evidence: the probability of some (sensory) data under a generative model. Also known as the marginal likelihood. The log model evidence is approximated by (negative) variational free energy.

Policy: an ordered sequence of actions.

Posterior belief: a belief about the causes of (sensory) data after belief updating.

Pragmatic value: the expected log likelihood of preferred outcomes, under a particular policy. In economics this is known as expected utility. In Bayesian decision theory it is known as (negative) Bayesian risk.

Prior belief: a belief about the causes of data, prior to sampling (sensory) data

Prior preference: a prior belief about an outcome in the future, which generally depends upon a policy.

Risk: the complexity expected under a particular policy. In other words, the expected divergence between predicted and preferred outcomes.

Variational Free energy: a functional of sensory data and posterior beliefs that approximates model evidence. Free energy scores the implausibility of some (sensory) data, given posterior beliefs about the causes of those data.

Expression	Description	Units
$\{\eta, s, a, \mu\}$	External, sensory, active and internal states	a.u.
$P(\eta, s, a)$	Generative model; i.e., a probabilistic specification of how external states cause sensory and active states	
$Q_\mu(\eta)$	Posterior (Bayesian) belief about external states, parameterized by internal states	
$Q_a(s, \eta \mid a)$	Predictive (Bayesian) belief about future sensory and external states, under a particular policy or action	
$F(s)$	Variational free energy – an upper bound on the surprisal of sensory states	nats
$G(a)$	Expected free energy – an upper bound on the surprisal of sensory states in the future	nats
$\Im(s) = -\ln P(s)$	Surprisal or self-information	nats
$D[Q(\eta) \mid\mid P(\eta)] = E_Q[\ln Q(\eta) - \ln P(\eta)]$	Relative entropy or Kullback-Leibler divergence	nats

References

1. Bem, D. J. (1972) Self-perception theory. Adv Exp Soc Psychol 6, 1-62.
2. Zanna, M. P. and Cooper, J. (1974) Dissonance and the pill: an attribution approach to studying the arousal properties of dissonance. J. Pers. Soc. Psychol 29, 703-709.
3. Miller, D. T. and Ross, M. (1975) Self-serving biases in the attribution of causality: Fact or fiction?. Psychol Bull 82, 213-225.
4. Kruglanski, A.W. (2004) The Psychology of Closed-Mindedness. New York: Psychology Press.
5. Friston, K. et al. (2015) Active inference and epistemic value. Cogn Neurosci 6, 187-214.
6. Chater, N. et al. (2006) Probabilistic models of cognition: Conceptual foundations. Trends Cogn Sci 10, 287 – 291.
7. Griffiths, T.L. et al. (2008) Bayesian models of cognition. In The Cambridge Handbook of Computational Psychology (Sun, R. ed), pp. 59-100. New York, NY, US: Cambridge University Press.
8. Itti, L. and Baldi, P.F. (2006) Bayesian surprise attracts human attention. Adv Neural Inf Process Syst, 547-554.
9. Weinstein, N. D. (1980) Unrealistic optimism about future life events. J. Pers. Soc. Psychol 39, 806-820.
10. Lord, C.G. et al. (1979) Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. J. Pers. Soc. Psychol 37, 2098-2109.
11. Kunda, Z. (1987) Motivated inference: Self-serving generation and evaluation of causal theories. J. Pers. Soc. Psychol 53, 636-647.

12. Taber, C.S. and Lodge, M. (2006) Motivated skepticism in the evaluation of political beliefs. Am J Pol Sci 50, 755-769.
13. Shepperd, J. A. and McNulty, J. K. (2002) The affective consequences of expected and unexpected outcomes. Psychol Sci 13, 85–88.
14. Ditto, P. H. et al. (1998) Motivated sensitivity to preference-inconsistent information. J. Pers. Soc. Psychol 75, 53-69.
15. Gigerenzer, G. and Garcia-Retamero, R. (2017) Cassandra's regret: The psychology of not wanting to know. Psychol Rev 124, 179-196.
16. Golman, R. et al. (2017) Information avoidance. J Econ Lit 55, 96-135.
17. Hastorf, A.H. and Cantril, H. (1954) They saw a game: A case study. J Abnorm Soc Psychol 49, 129-134.
18. Swann, W. B., Jr. et al. (1992) Depression and the search for negative evaluations: More evidence of the role of self-verification strivings. J Abnorm Psychol 101, 314–317.
19. Hohwy, J. (2016) The self-evidencing brain. Noûs 50, 259-285.
20. Hinton, G.E. (2007) Learning multiple layers of representation. Trends Cogn Sci 11, 428-434.
21. Gregory, R.L. (1980) Perceptions as hypotheses. Phil Trans R Soc Lond B. 290, 181-197.
22. Helmholtz, H. (1878/1971) The facts of perception. In The Selected Writings of Hermann von Helmholtz (Middletown, R.K., ed) pp. 384. Connecticut: Wesleyan University Press.
23. Clark, A. (2016) Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford University Press.
24. Friston, K. (2009) The free-energy principle: A rough guide to the brain? Trends Cogn Sci. 13, 293-301.

25. Bowers, J.S. and Davis, C.J. (2012) Bayesian just-so stories in psychology and neuroscience. Psychol Bull 138, 389-414.
26. Bastos, A.M. et al. (2012) Canonical microcircuits for predictive coding. Neuron 76, 695-711.
27. Friston, K.J. et al. (2017) The graphical brain: Belief propagation and active inference. Network neuroscience (Cambridge, Mass.) 1, 381-414.
28. Rao, R.P., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci. 2, 79-87.
29. Shipp, S., 2016. Neural Elements for Predictive Coding. Front Psychol 7, 1792.
30. Constant, A. et al. (2019) Regimes of expectations: An active inference model of social conformity and human decision making. Front Psychol 10, 679.
31. Friston, K. et al. (2015) Active inference and epistemic value. Cogn Neurosci 6, 187-214.
32. Constant A. et al. (2019) Regimes of expectations: An active inference model of social conformity and human decision making. Front Psychol 10, 679.
33. FitzGerald, T.H. et al. (2015) Active inference, evidence accumulation, and the urn task. Neural Comput 27, 306-28.
34. Friston K. et al. (2017) Active inference: A process theory. Neural Comput 29, 1-49.
35. Rao R. (2010) Decision making under uncertainty: A neural model based on partially observable Markov decision processes. Front Comput Neuroscience 4, 146.
36. Friston, K. (2018) Active inference and cognitive consistency. Psychological Inq 29, 67-73.
37. Kruglanski, A.W. et al. (2018) Cognitive consistency theory in social psychology: A paradigm reconsidered. Psychological Inq 29, 45-59.

38. Kruglanski, A.W. et al. (2018) All about cognitive consistency: A reply to commentaries. Psychological Inq 29, 109-116.
39. Festinger, L. (1957) A Theory of Cognitive Dissonance. Evanston, IL: Row.
40. Gawronski, B. and Strack, F. eds (2012) Cognitive Consistency: A Fundamental Principle in Social Cognition. New York, NY: Guilford Press.
41. Powers, A.R. et al. (2017) Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. Science 357, 596-600.
42. Van de Cruys, S. et al. (2014). Precise minds in uncertain worlds: Predictive coding in autism. Psychol Rev 121, 649-675.
43. (43) Schwartenbeck, P. et al. (2015) Evidence for surprise minimization over value maximization in choice behavior. Sci Rep 5, 16575.
44. (44) Guess, A. et al. (2019) Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Sci Adv 5, eaau4586.
45. Sharot, T. (2011) The optimism bias. Curr Biol 21, R941-R45.
46. Sharot, T. et al. (2012) How dopamine enhances an optimism bias in humans. Curr Biol 22, 1477-1481.
47. Sharot, T. et al. (2009) Dopamine enhances expectation of pleasure in humans. Curr Biol 19, 2077-2080.
48. FitzGerald, T.H. et al. (2015) Active inference, evidence accumulation, and the urn task. Neural Comput 27, 306-328.
49. Friston, K. et al. (2017) Deep temporal models and active inference. Neurosci Biobeh Rev 77, 388-402.

50. Mirza, M.B. et al. (2016) Scene construction, visual foraging, and active inference. Front Comput Neurosci 10, 56.

51. Itti, L. and Baldi, P. (2009) Bayesian surprise attracts human attention. Vision Res 49, 1295-1306.

52. Sun, Y. et al. (2011) Planning to be surprised: Optimal Bayesian exploration in dynamic environments. In Artificial General Intelligence: 4th International Conference, AGI 2011, Mountain View, CA, USA, August 3-6, 2011. Proceedings, ed. J Schmidhuber, KR Thórisson, M Looks, pp. 41-51. Berlin, Heidelberg: Springer Berlin Heidelberg.

53. Barlow, H. (1961) Possible principles underlying the transformations of sensory messages In Sensory Communication (Rosenblith, W., ed), pp. 217-34. Cambridge, MA: MIT Press.

54. Linsker, R. (1990) Perceptual neural organization: Some approaches based on network models and information theory. Annu Rev Neurosci 13, 257-281.

55. Fleming, W.H. and Sheu, S.J. (2002) Risk-sensitive control and an optimal investment model II. Ann. Appl. Probab 12, 730-767.

56. van den Broek, J.L. et al. (2010) Risk-sensitive path integral control. UAI 6, 1–8.

57. Von Neumann, J. and Morgenstern, O. (1944) Theory of Games and Economic Behavior. Princeton: Princeton University Press.