Feelings for Knowing

Lauren Olin
University of Missouri – St. Louis
Department of Philosophy
Center for Neurodynamics
olinl@umsl.edu

ABSTRACT: Recent work in the psychology of human reasoning and judgment has emphasized the importance of affective dispositions in the pursuit of broadly epistemic ends. Philosophers, too, have recently become interested in empirical and conceptual questions about the connection between human affective states and distinctively epistemic ends like true belief and knowledge. Despite growing interest in the idea that some affective states are relevant to epistemology, however, questions about their natures, about the natures of the epistemic roles they play, and about the natures of the dispositions that undergird them, loom large. This paper provides a tentative distinction between epistemic feelings and epistemic emotions, and a theoretical framework designed to evaluate the roles of affective dispositions in the genesis and maintenance of belief systems.

1. Introduction: Epistemic feelings *vs.* Epistemic emotions

In recent years, both philosophers and psychologists have become interested in a cluster of affective states – what I'll here call *feelings for knowing* - that seem distinctively epistemic in character. I use the term feelings for knowing because epistemic emotions and epistemic feelings are not often carefully distinguished in philosophical or psychological literatures. Epistemic feelings like "feelings of knowing" and "feelings of certainty" -- staples in the metacognitive research programs that have flourished since Flavell's (1979) seminal paper on "thinking about your own thinking" -- are often discussed alongside other feelings for knowing, like curiosity and trust, that have status as epistemic emotions or virtues in more squarely epistemological circles.

Contemporary cognitive science, however, is awash with proposals for distinguishing between classes of affective states commonly glossed as emotions or as feelings. Accordingly, there are contemporary philosophical debates about whether epistemic emotions and epistemic feelings can be principally distinguished at all and, if so, how they ought to be distinguished. Carruthers (2017: 60) suggests that epistemically relevant affective states might just be categorized extensionally, for the purposes of investigation, "as the set that includes surprise, interest, curiosity, uncertainty, and feelings of knowing." Proust (2012), in contrast, insists on a sharp empirical distinction between epistemic emotions and epistemic feelings. And the weeds get

thicker yet: for it only takes a moment of reflection in this general sphere to generate a long list of affective states that might be claimed as candidates across the board. For instance, consider:

Surprise	Misology	Difficulty
Interest	Wonder	Déjà Vu
Curiosity	Tip-of-the-tongue states	Epistemic anxiety
Uncertainty	Interest	Drawing a blank
Feelings of Knowing	Feelings of Familiarity	Rightness
Intellectual courage	Feelings of (Dys)Fluency	Meaningfulness
Astonishment	Confusion	Frustration
The fear of the unknown	Certainty	Pastness
Insight	Confidence	Trust
Boredom	Understanding	Mirth

Even among those who argue that epistemic emotions and feelings can be differentiated, there are debates about how the resultant categories ought to be operationalized. Following Proust (2012), Meylan (2014) argues that epistemic feelings and epistemic emotions are differentiable along a hard line. But even after drawing that line, she argues that epistemic emotions don't themselves compose a natural psychological kind. Emotions, she reasons, can only qualify as epistemic if they are concerned with evaluative properties linked to the goal of somehow getting things right: "[a]n emotional episode is epistemic if and only if its formal object is an epistemic evaluative property." Candidate epistemic emotions like interest and surprise, she reasons, have very different relationships to the truth; grouping them together invites violations of what she calls the "formal object standard for epistemicity."

In addition to debates about what affective states are rightly counted as epistemic emotions and/or epistemic feelings, there are debates about what those states are *for*. In the history of the theory of rationality, feelings and emotions have been seen as obstacles to good reasoning and epistemic success. But there now exists a large body of empirical evidence suggesting that people deprived of certain emotional capacities are thereby deprived of certain rational capacities (cf. Damasio 1994), and philosophers have independently argued for the relevance of affective states in epistemology. De Sousa (2008), for instance, has argued that both epistemic feelings and epistemic emotions are centrally relevant to our abilities to solve difficult problems, including the frame problem, and problems related to epistemic justification. Morton (2010) has argued that epistemic emotions play a "conceptually vital" role in belief acquisition: on his view, unless people experienced certain epistemic emotions – motivating emotions such as worry, fascination, and curiosity - they would flounder in their pursuit and execution of epistemic projects, and end up lacking knowledge.

In summary, then, here's the messy starting point: feelings for knowing clearly matter for epistemology, but substantive disagreements about their natures characterize contemporary discussions in both philosophy and psychology. First, there are disagreements about what feelings for knowing are, and about whether epistemic emotions and epistemic feelings constitute

distinct categories or kinds of feelings for knowing. Second, there are debates about what roles feelings for knowing play in guiding epistemic inquiry, and corollary debates about whether epistemic emotions and epistemic feelings, if they can be distinguished as kinds, play distinctive epistemic roles.

What follows is an attempt to start tidying this mess: I will assume that feelings for knowing exist and are important, and argue that there is a strong case for distinguishing epistemic emotions and epistemic feelings. Rather than follow convention in surveying the motley crew of candidates for inclusion in either category, however, I'll reverse-engineer a distinction between epistemic emotions and epistemic feelings based on attention to the functional roles that epistemic feelings play in the "low-level" metacognitive monitoring mechanisms that psychologists have proposed give rise to them (Koriat & Nussinson 2009). The taxonomic framework developed suggests that epistemic feelings generate conscious, cue-based predictions about the contents of representational knowledge structures, and associated possibilities for mental action, while epistemic emotions have important motivational roles to play in gathering evidence that can be used to update those knowledge structures, or to create new ones, in cases where feelings-based metacognitive judgments go badly. The paper concludes with speculations concerning the relationship between metacognitive and mindreading capacities, and a possible role for mirth in negotiating the relationship between them.

2. Metacognitive evaluation & Epistemic feelings

Metacognition, in the most general terms, is cognition about one's own cognition, or thinking about one's own thinking. In Flavell's classic words (1979: 232):

> Metacognition refers to one's knowledge concerning one's own cognitive processes or anything related to them, e.g., the learning-relevant properties of information or data. For example, I am engaging in metacognition if I notice that I am having more trouble learning A than B; if it strikes me that I should double check C before accepting it as fact.

This definition is deliberately broad, and the term metacognition has, accordingly, since been used to refer to almost any process or capacity related to epistemic self-assessment. As another famous figure in the metacognitive tradition, A.L. Brown observed in rendering his influential model, "metacognition is not only a monster of obscure parentage, but a many-headed monster at that" (1987: 105). Moreover, emerging consensus dictates that capacities for metacognitive monitoring are not capacities that were specifically adapted to or designed for that purpose, but are rather "cobbled together" from other capacities (Fletcher & Carruthers 2012).

Still, despite the range of processes that have historically been treated as metacognitive, most contemporary theorists agree that all metacognitive processes for self-evaluation involve three distinct stages. First, when presented with a task, individuals *plan* by selecting appropriate

strategies and allocating cognitive resources accordingly. Second, once the task is underway, individuals *monitor* their performance. Finally, once the task has been completed – or at least after a specified amount of time has passed, individuals *evaluate* their performance on the task.

A second feature of metacognition that claims consensus involves a distinction between "low-level" metacognitive processes that implicate feelings, are non-analytic, and experience based, and metacognitive judgments, that are more analytic, and information-rich (Koriat & Levy-Sadot 1999). In particular, the research suggests that the judgments issued at planning stages of any metacognitive process rely on cue and heuristic based predictions about whether or not some item will be available in memory, or about whether one is likely to be possessed of other cognitive resources that are required to complete some task. For example, it is widely agreed that low-level metacognitive judgments utilize availability heuristics (Tversky & Kahneman, 1973): people assume that the more examples hat come to mind easily, the more frequent, typical, and probably true information contained in them is. Subsequent monitoring and evaluation processes, in contrast, are more analytic, and appear to prompt the activation of epistemic emotions with distinctive motivational and behavioral profiles.

There is now a large body of research supporting this general distinction between types of metacognitive processes, and I will here explicate the difference with reference to a few important examples. In a seminal study, Kelley & Rhodes (2002) demonstrated that memory retrieval is not simply a matter of the activation of traces in an episodically bound network, but rather a subjective experience that is guided in large part by attributional processes. Fluency of processing plays a large role in how people make metamnemonic judgments, or judgments of how likely it is that they will engage in successful attempts at memory retrieval. Judgments of learning also appear to be underwritten by broadly feelings-based judgment processes. Castel, McCabe, & Roediger (2007) presented subjects with related, unrelated, and identical word pairs and obtained judgments of learning immediately after presentation. Judgments of learning were higher for identical pairs, suggesting an important role for fluency and perceptual cue similarity in judgments of learning as well.

While there are debates about how to best conceptualize these feelings-based predictions, one approach is suggested by attention to the literatures on schema theory and predictive modeling. Human beings are constantly making predictions about the world, and about how the objects and individuals that populate it are likely to behave. According to schema theory, the information that guides these predictions is organized in dynamic knowledge structures called schemas (Bartlett 1932; Rumelhart & Ortony 1977; Mandler 1984). These knowledge structures represent situations, objects, events, individuals, and actions at multiple levels of abstractness. They function both to facilitate the comprehension of input from the environment, and to make predictions about future events and states of affairs on the basis of that input. In Rumelhart's (1980: 41) words:

schemata can represent knowledge at all levels - from ideologies and cultural truths to knowledge about the meaning of a particular word, to knowledge about what patterns of excitations are associated with what letters of the alphabet. We have schemata to represent all levels of our experience, at all levels of abstraction. Finally, our schemata are our knowledge. All of our generic knowledge is embedded in schemata.

Schemas can be built at varying levels of specificity, but always describe the general characteristics of an object or situation, etc., and always have variables that can assume a limited range of values across circumstances; "default" values correspond to the prototypical features of whatever is being represented. Schemas are built up, typically non-consciously, on the basis of experience, and generate future expectations of how the world ought to be, and of how things in it ought to behave.[1] To take an example drawn from Martin (2007: 86), a schema for birds would include a number of bird-related variables, for things like feet, beaks, wings, tail types, etc., and those variables can shift values to accommodate experiences with different kinds of birds – from canaries, to eagles, to ostriches, to pelicans.

Schemas have been studied for decades in psychology, although questions about their roles in learning and memory, and about the details of their neurological implementation, have only recently become hot topics. A growing literature in memory consolidation and its relationship to schema theory suggests that schemas are best conceptualized as networks of strongly interconnected neocortical representations that facilitate predictions about future states of affairs on the basis of similar past instances (van Kesteren et al. 2012).[2]

---

[1] *Scripts* (Abelson 1981; Shank & Ableson 1977) are particular kinds of schemas tailored to describing routine activities like making coffee, or social activities like going on a date, while *frames* (Minsky 1977) are schemas devoted to knowledge of specific physical environments.

[2] While there isn't much neuroscientific work on the mechanisms undergirding schematic representation, it is widely agreed that parts of the vmPFC, MTL, and mesolimbic reward centers are all implicated in schema instantiation and schema revision. According to the standard "systems view" of schema consolidation, new information is initially dependent on MTL structures, including especially the hippocampus, but over time and given repeated exposures it becomes represented independently – schematically - in vmPFC, which is also responsible for the integration of cognitive and affective information and for the generation of emotionally valenced, future-directed event representations. The vmPFC is important for making online predictions, as for instance in the case of perception, but also in the social domain (van Kesteren et al 2012; Summerfield & Koechlin 2008). MTL structures, in contrast, are important for identifying schema-incongruent, novel information; recent investigations have revealed that MTL structures are tightly coupled with vmPFC networks. This close coupling is thought to facilitate schema-related memory enhancement, and vmPFC is known to be crucially involved in integrating new information into existing schemas (Tse et al. 2011).

Moreover, it is widely agreed that there are competing demands on the instantiation of representational schemas. The variables in schemas can take a range of values because they are supposed to flexibly accommodate particular instances, but the predictive power of the schemas attaches to their generality and robustness: they can't include too many variables, or they would become too specific to deploy across contexts. The instantiation of schemas must be constrained; the goal of the representational system is to include all and only variables corresponding to elements that are likely to be relevant to future instances. Low-level metacognitive monitoring processes issue feelings-based predictions about the utility of representational schemata as applied in particular circumstances; positively valenced epistemic feelings provide a conscious signal to the individual that they are likely in the possession of the resources needed to successfully complete some task, while negatively valenced one's prompt adopting alternative epistemic strategies.

When experiencing a feeling of knowing (FOK), subjects feel like they will be able to retrieve some piece of information from a schematic representation in memory before they are in fact able to retrieve it. Feelings of knowing are states thought to be grounded in the cue-based recognition of components or elements present in a target memory schema, though the states can also arise on the basis of familiarity with the terms present in a question, even in cases where the answer to that question is totally unknown to an experimental subject (Dunlosky & Metcalfe 2009). While multiple studies have demonstrated that individuals can rely upon their feelings of knowing in order to accurately predict their memory performance with surprising reliability (Reder 1987; Reder & Schunn1996), it is widely agreed that the feeling is the result of an unconscious heuristic process that monitors aspects of memory performance that are only indirectly related to knowledge, and can sometimes be misleading: strong feelings of knowing do not always correlate with success in retrieval.

Reder and Ritter (1992) provide evidence that FOK is involved in the evaluation of questions in epistemically undesirable ways. Their results show that we're more likely to think we know the answer to "47*54" if we've seen "47+54" lots of times, if we've seen "47" lots of times, or if we've seen "54" lots of times – even when we've never seen "47*54", or its answer. They claim that "feelings of knowing" the answers to problems has little to do with whether we actually know them, but much to do with whether we've seen superficially similar problems before.[3]

Reder & Ritter's experiments both involved an adaptation of the *Game Show Paradigm* (cf. Reder 1987), in which subjects are briefly presented with a math problem, and press a button indicating whether or not they think they can answer it. These subjects chose "retrieve" if they thought they knew the answer, and "calculate" if they didn't. The idea behind this choice paradigm is that

---

[3] Thomas et al. (2012) reviews evidence that feeling-of-knowing judgments are likewise cue based. See Schwartz and Metcalfe (2011) for evidence that TOTs are, similarly, heuristically driven.

changing variables in the engineered math problems would allow the researchers to dissociate the effects of frequency of exposures to the whole problems, and to the parts of problems, and the to use this data in assessing whether FOKs tracked actually knowing the answer, or just being familiar with parts of the problem. With this framework, and the additional assumption that response time can be used as a measure of degree of knowing, Reder and Ritter's results cast doubt on the assumption that FOK judgments in fact yield good predictions about whether some item is in fact stored in memory.

Feelings of confidence, feelings of familiarity, feelings of fluency, and feelings of knowing all appear to be positively valenced,[4] and directed towards objects that are represented much more clearly than those representations that negative epistemic feelings are typically directed towards. None of the positive epistemic feelings are thought to implicate arousal or to motivate changes of action. Rather, they appear to be implicated in choices for maintaining a strategy, or adopting different strategies for mental search. They can be understood as falling on a kind of dimensional continuum from feelings of certainty, to feelings of error, or of being likely to get something wrong. On the certainty side, the feeling prompts individuals to trust their judgment that they are representing some aspect of the world appropriately. Feelings of uncertainty, in contrast, might prompt an individual to hesitate and continue searching and, eventually, perhaps to abandon a given search in favor of acting in the world in others ways that will facilitate acquiring the information they are after in some other way.

Other epistemic feelings that feature in evaluation stages of metacognitive judgments are less positive. In tip-of-the-tongue (TOT) states, individuals feel that they have memory knowledge of something that they nonetheless have fail to retrieve. They have access to certain of the features of the item they are attempting to recall and name, and studies have shown that individuals in TOT states are also able to discriminate between different objects they are presented with and make judgements about which is the most likely candidate (Schwartz 2002). As Arango-Munoz (2014) has suggested, tip-of-the-tongue states are in this way directed to the objects or pieces of information an individual is trying to recall.

TOT states are negatively valenced and, like other negatively valenced epistemic feelings, are directed toward mental objects and dispositions that are to some extent opaque. Feelings-of-forgetting (FOFs) – the nagging feelings that you are forgetting to do something you-know-not-what, are similarly negatively valenced, and directed towards objects that are similarly opaque. Feelings of uncertainty are also negatively valenced, but appear to be directed towards objects that are less opaque. In experiencing uncertainty about whether some item has been appropriately recalled from memory, for instance, there is less doubt about what the object in question or piece of information *is*, there is rather just doubt about whether or not *that* item was

---

4 Stepper & Strack (1993: 211) have suggested that epistemic feelings "have no fixed valence," but the evidence overall overwhelmingly supports the view that metacognitive emotions are affective in character. See Proust (2015 sec. 7) and Oppenheimer (2008) for reviews.

correctly called to mind, or whether the item called to mind has been correctly remembered. In these cases, an agent might continue to search for an alternative candidate representational schema, but do not appear to be motivated to stop thinking about their own thinking and adopt alternative strategies for obtaining or confirming the accuracy of the information that they feel uncertain about – at least not across the board.

In summary, then, epistemic feelings like confidence and uncertainty that are implicated metacognitive self-evaluations appear to be the inputs and outputs of low-level, probabilistic assessments about the contents of existing knowledge representations, and the likelihood that they can be called upon in order to facilitate good epistemic outcomes. In cases where those feelings are positively valenced, and subsequent evaluations reveal a mistake, then an agent might be prompted to adopt an alternative epistemic strategy for obtaining the needed information. Likewise, adopting a different strategy might called for whenever the outputs of low-level processes implicate feelings of anxiety or uncertainty.

3.   Inquiry & epistemic emotions

The foregoing interpretation of the literatures in metacognition and predictive modeling suggests that there are important roles for epistemic feelings to play in probabilistic self-evaluations representational schemata or knowledge structures. I've suggested that the metacognitive judgments associated with epistemic feelings are those involved in evaluations of whether an activated schematic model is likely to facilitate epistemic success. In this section, based on that analysis, I will suggest a number of dimensions along which it seems appropriate to distinguish epistemic emotions and epistemic feelings.

3.1   *Motivation & Arousal*
The feelings implicated in metacognitive evaluations don't appear to implicate the sorts of physiological arousal that are associated with action. Rather, in cases where individuals experience negative metacognitive feelings, or in cases where a metacognitive judgment is revealed to be mistaken, individuals will then experience other epistemic emotions Unlike epistemic feelings, epistemic emotions appear to be intrinsically arousing or motivational, and thereby aid the practical sorts of inquiry needed to gather the evidence needed to appropriately update schematic knowledge representations.

It might be objected that some epistemic feelings are associated with motivations for action. And to be sure, epistemic feelings *can* be associated with particular kinds of action. For instance, as Nagel (2007: 809) has suggested, if one experiences a protracted feeling of knowing some answer, they might characteristically feel like reporting to others *"I know that I know this one"* or something of the sort. The claim is rather that, unlike epistemic emotions, epistemic feelings do not appear to be associated with *particular* motivational states that dispose an individual to pursue particular sorts of actions.

It is worth noting that the point here concerns the intrinsic motivational structure of the states in question, not about whether those states are in any interesting way volitional. One might very well experience a strong feeling of knowing without any attendant motivation to act, and yet still choose to act for other reasons. By comparison, consider the experience of feeling surprised. Surprise is associated with characteristic facial expressions that involve, among other things, a widening of the eyes. The emotion is also associated with motivations to engage in a range of behaviors that involve attending to whatever individual, object, or event elicited the surprising reaction, approaching it, and exploring it. Curiosity, likewise, characteristically involves a motivation to explore or take a closer look at whatever captured one's attention.

Emotions in general are understood as *motivating* states; The importance of the motivational structure of epistemic emotions is highlighted well by Morton (2010), who asks that we imagine

> an extraordinarily well trained and malleable young scientist. From early on in her career she has been mentored by older scientists who not only are top researchers in her field, but are also pedagogically sharp and sensitive. The result is that she has a superlative grasp of research techniques, is aware of the live problems at the cutting edge of her subject, and has the patience and intelligence to do very good work. There is one flaw, however. She does not care about the subject. She has no curiosity. She wants a career, and she knows that with her background she is more likely to succeed by pushing some lines of theory than others, so she is capable of a form of scientific partisanship. But she does not find herself wanting the truth to turn out one way rather than another in more than this instrumental way. She does not sometimes wonder whether lines of inquiry that are, with good reasons, disparaged by her research group might not in the end give important clues to the underlying processes she is investigating.

According to Morton, this scientist lacks certain emotions, like wonder and curiosity, that might motivate her to explore the sorts of scientific questions that characterize eminent, rather than mediocre, careers in knowledge gathering.

3.2   Valence
A second observation that comes from considering the characteristics of epistemic feelings that are implicated in metacognitive self-assessments of knowledge representations is that while epistemic feelings are always valenced – that is, always experienced as positive or negative - epistemic emotions can lack valence. Experiences of being surprised, for instance, can be good or bad, depending on the overall context.

3.3   Consciousness
As Michaelian & Arango-Munoz (2014) and Meylan (2014) have also pointed out, while epistemic feelings are always consciously apprehended – even if only on the "fringes" of consciousness – emotional experiences are often unconscious.

3.4   Content

Finally, attention to the role of epistemic feelings as probabilistic indicators of the appropriateness of one's schematic knowledge representations makes it clear that epistemic feelings always have these knowledge representations, or aspects of their contents, as their objects. In cases where the feelings are valenced negatively these contents can be construed as opaque, as for example in experiences of the feeling of forgetting, or the feeling of uncertainty. In making low-level metacognitive judgments, individuals self-ascribe mental states or capacities without having access to their contents, as for instance in cases where someone claims to know the answer to a memory query before actually retrieving the answer.

Epistemic emotions, while also directed, appear to take as their objects particular, identifiable things in the world, not internal capacities, knowledge structures, or opaque representations. Both epistemic feelings and epistemic emotions are intentional and directed towards objects, but only epistemic feelings take as their objects (aspects of) internal representational states. As Koriat & Nussinson (2009), have emphasized, while metacognitive feelings implicitly inform choices about how to allocate cognitive resources, but do so in ways that involve no access to the semantic contents of the individual's representations. Epistemic feelings, on this view are mediated by the application of heuristics that "pertain to global, structural aspects of the processing of information" – cues like fluency, and familiarity, rather than the contents of the representation in question. Epistemic feelings, on this view are mediated by the application of heuristics that "pertain to global, structural aspects of the processing of information" – cues like fluency, and familiarity; they are not directed towards the contents of the representation in question.

In summary, then: epistemic feelings appear to function as the inputs and outputs to low-level metacognitive processes that are widely agreed to function on the basis of cues, and never to implicate purely propositional content. Epistemic emotions, in contrast, appear to be directed, intentional states that sometimes involve propositional content, and that sometimes implicate sophisticated cognitive states like beliefs and desires.

4. Metacognition & mindreading

A final debate in the literature on epistemic emotions concerns the relationship between capacities for mindreading and capacities for epistemic self-evaluation. It is widely agreed that there are some relationships between the capacities for mindreading and capacities for metacognition, though there are debates about the particular nature of this relationship. *Self-ascriptive* theorists have argued that there are no specially adapted mechanisms for metacognition. Rather, capacities for epistemic self-evaluation piggy-back on more general mechanisms that evolved to facilitate inferences about the mental states of others (Carruthers 2009, 2016).

There are multiple sources of evidence for self-ascriptive views. One is that there exist substantial individual differences in capacities for metacognitive monitoring and control: some people reflect on their reasoning all the time, while others only do so when culturally prompted by *e.g.* their

parents, or professors at school. The degree of variation implies that there is no evolved mechanism devoted to engaging in such evaluations. Another source of evidence starts with the observation that the development of capacities for epistemic self-evaluation run more or less in parallel with the development of capacities for mind-reading: when asked about whether they have known something they just learned, 3 year olds regularly report that they've always known it! As Proust (2012) summarizes:

> when asked to verbally report about what they know, what appears to them, what they can remember, etc., children seem unable to offer reliable answers before they are able to read their own minds. However, once they have acquired, through verbal communication, the concepts for the basic mental states, and thereby become able to understand how other agents can be wrong about the world, children learn to attribute errors and misrepresentation to themselves as well.

According to *self-evaluative* theorists, in contrast, while capacities for mindreading are relevant to metacognitive processes, they are just one factor: low-level metacognitive capacities make predictions about the likelihood or certain kinds of epistemic success (or failure), while conceptual mindreading capacities provide individuals with a second-order understanding of their cognitive dispositions that, sometimes, will prompt overriding lower-level, feelings-based metacognitive judgments.

Distinguishing between the details of these competing proposals is difficult, not least because, as I've already mentioned, the definitions of metacognition that feature in the literature are often poorly rendered. But developmental research indicates overwhelmingly that as early as 9 months of age, infants deploy intentional schemas in making predictive inferences about the behavior of agents (Csibra et al. 1999). Indeed, while researchers agree that these initial intentional attributions mark only the beginning of a long developmental trajectory that results in a full-fledged theory of mind – a theory that facilitates meta-representational inferences - some theorists hold that a basic set of intentional schemas is relatively "hard-wired" by natural selection, and that children default to using these schemas and over-attribute agency in their early predictions about the behavior of objects in the world.

Attention to the distinction proposed between epistemic feelings and epistemic emotions, however, suggests an implicit argument in favor of an account according to which low-level metacognitive processes determine courses of mental action much of the time, and without the aid of higher order mindreading processes, but that in cases where feelings-based predictions go awry that these further resources are engaged. Moreover, it appears that they are engaged in the same way, and for many of the same reasons, that they are engaged when low-level evaluations of the behavior of other agents generate similar epistemic feelings. This section offers an exploratory elaboration of this general analogy.

Epistemic feelings like feelings of knowing and confidence, for instance, might themselves be sufficient to motivate attempts to recall an answer, for instance, totally without any support from higher cognitive or emotional systems. It is only in these circumstances that the mindreading system, the primary function of which is to attribute mental states to other agents, can be turned inward on one's own mind, making it possible to take advantage of the informational structures and inferential principles that are typically employed in the course of attributing false beliefs and other mental states third-personally.

Epistemic emotions, accordingly, may only become relevant in cases where feelings-based predictions about the contents of representational schemata fail. When confronted with such failures, people sometimes adjust their strategies, *shifting* from one sort of schema to another. For example, people sometimes make mistakes about whether they are in fact dealing with an intentional system, and default to the physical or design stances in order to recover. Haunted house aficionados will be familiar, for instance, with the experience of being frightened by a bloodied gadget, then switching stances once evidence of mechanism accumulates. In cases of extreme behavioral patterns, such as those characteristic of mentally ill persons, Dennett suggests we may find "a species of the design stance" effective and adopt it instead, going forward (1971: 94).

These shifts may sometimes be pursued non-consciously, without requiring resources from reflective or emotional systems. One attractive hypothesis is that when the epistemic feeling associated with the failure – like a feeling of certainty or TOT state - the positive valence of the emotion will be enough to motivate a continued search program that will eventuate in the activation of an alternative schema. If enough time passes, however, without success, it seems likely that emotional dispositions will become engaged and motivate alternate avenues of inquiry.

In still other cases, negatively valenced epistemic feelings might prompt *condition-updating*, or revisions to one's view of the kinds of objects that are well represented by particular schemata. If one experiences a feeling of certainty that a word has been encountered before, for instance, but then fail to retrieve it, one might be prompted to revise the conditions under which the activated schema ought to be applied. Similarly, revisions to one's view of the kinds of systems whose behavior is well predicted by applying intentional representational schemas might be sensitive to such failures. For example, human beings naturally associate applications of the intentional stance with other signatures of agency – things like the possession of eyes, capacities for directed motion, and contingent interaction (Arico et al. 2011) - but in some circumstances the intentional stance is applied to other sorts of entities: corporations are sometimes treated as rational, despite lacking typical signatures of agency. In these cases, one attributes to themselves

a false belief, and is subsequently motivated to engage in the kinds of epistemic investigations associated with puzzlement, curiosity, and perhaps even fear.[5]

In still other circumstances when confronted with negatively valenced epistemic feelings, people may be prompted to engage in *schema-revision*. With *enough* evidence, for example, it may be worth updating one's knowledge structures to accommodate new instances. With *enough* evidence, for example, it may be worth updating one's intentional schemas in order to accommodate a view of intentional systems as, well, less intentional. Robbins and Jack (2006; cf. Dennett 1971) have suggested, for example, that while the "explanatory gap" intuitions commonly associated with feelings of uncertainty seem well explained by reference to conflict between the conceptual schemes associated with the intentional and physical stances, that intuitions about whether there is, in fact, an explanatory gap might vary depending on whether a person's experience has motivated changes to intentional conceptual schemas. Folk-intentional schemas possessed by psychologists and neuroscientists who claim to lack dualistic intuitions have perhaps been revised to admit of more mechanistic components. Alternatively, perhaps their folk-physical schemas have been altered to accommodate the presence of intentionality in the physical world.

Finally, one might experience failures of low-level metacognitive evaluations of representational schemas in ways that don't then prompt stance-switching, condition-updating, or schema-revision. In these cases, representational schemas will be retained and re-deployed in similar contexts without alteration. For instance, consider the experience feelings of uncertain when confronted with a word being used in a comic way, before the subject matter is revealed to be part of a joke. It would be a mistake to then update one's associated semantic knowledge structure in ways that make it able to accommodate future instances of the word being used in a comic way. This suggests that a subset of "neutral" emotions like mirth might play a role in the maintenance of robust knowledge representations. The low-level metacognitive failure is deemed serious enough to engage higher-order capacities and attribute to oneself a false belief about the contents of one's own mind, but not to then "do anything" about it.

When people parse sentences or events in the world most generally, they start by adopting a predictive schema that is then filled in holistically with perceptual, conceptual, and pragmatic details. In some cases those schemas are revealed to be inadequate for the purposes of predicting what will come next and, as a result, a new schema is activated that is compatible with the new information. Garden path sentences, also associated with feelings of uncertainty and experiences of mirth, lure their interpreters into adopting a schema that yields false expectations about what's going to happen next, and, as a result the utility of the schema initially deployed is called into question. However, it is rarely the case that in the future it would be wise to begin with a different

---

[5] Morton (2014) has argued that almost any virtue – or vice – can function in the service – or to the detriment – of epistemic ends in at least some circumstances. The framework sketched here is consistent with that possibility.

schema. Consider an individual who encounters the garden path sentence *She told me a little white lie would come back to haunt me.* In parsing the next sentence that begins "She told me a little white lie…" the same predictive schema should be employed – it's default value unchanged – despite the fact that the garden path sentence violated her expectations previously. Robust grammatical expectations are important to maintain in the face of countervailing evidence, just as are expectations about how complex intentional systems are likely to behave.[6]

5. Conclusions & Directions

This framework is speculative, but it still also involves several distinct advantages relative to other extant approaches towards taxonomizing the epistemic emotions and feelings in the literature. The first is that it explains the relationship between epistemic feelings and epistemic emotions, and the distinctive roles that they play in an individual's cognitive economy.

Second, it is agnostic concerning the question of whether or not the epistemic emotions are a unitary psychological kind, leaving open the possibility that many different emotional states can be relevant to the pursuit of epistemic ends like truth and knowledge. At the same time, the category can be understood as constrained on functional grounds: epistemic emotions are those that come online on the tails of negatively valenced epistemic feelings, like the feeling of forgetting, or the feeling of uncertainty.

Finally, the account suggests a principled way of investigating and understanding the different roles that epistemic feelings and epistemic emotions play in structuring our epistemic lives, while at the same time acknowledging, as Goldie (2002: 235) reminds us, that:

> [o]ur emotions, moods, and character traits, broadly conceived, can interweave, overlap, and mutually affect each other. Even apparently insignificant traits, abilities, and habits can resonate through a person's psyche in such a way that their 'addition' or 'removal' could have dramatic and unforeseen consequences for the whole person…'tinkering' with personality can be an unpredictable business. A particular feature of the garden may, considered in itself, be attractive and undesirable; yet its removal might have adverse ramifications which overwhelm the advantages. Don't kill off the worms!

---

[6] That there exists a general mechanism designed to maintain the default-values attached to representational schemata makes sense from a practical perspective in cases where there are demands on the schemas to be widely deployed. But the idea also claims support from research on the difference between *adapted* and truly *cognitive* systems. Adaptive systems have access to representations of the world, but cognitive systems, in addition, manipulate representations that are detached from whatever happens to be going on in experience. It is the capacity to manipulate representations off-line that allows cognitive systems to evaluate possible future outcomes, and coordinate actions in ways that realize those outcomes.

Lots of discussion of the ways in which fluency-based and other heuristics have eventuated in poor epistemic evaluations have assumed that such attributions are inherently epistemically suspicious. The framework outlined in the above suggests that, in contrast, fluency-based knowledge attributions can in some circumstances to function in the service of mitigating self-doubt, and prompt the evaluation of schematic knowledge representations in a cost-effective way, bringing more cognitively expensive emotional evaluations online only where they are likely to substantially improve a subject's epistemic position overall.

The analysis above is consistent with the emerging consensus in the metacognitive literatures that metacognitive abilities are best understood as being composed of "a diverse hodge-podge of self-management strategies acquired through individual and cultural learning, which co-opt whatever cognitive resources are available to serve monitoring-and-control functions" (Fletcher & Carruthers 2012: 5). It is also broadly consistent with evidence suggesting that, in some cases, agents are better off trusting the outputs of feelings-based processes, for instance, in the case of complicated decision tasks that are concerned with all-things-considered judgments: epistemic feelings, but not epistemic emotions, can be better directed towards generally "good" and "bad" features in a complex choice, while epistemic emotions always take specific objects (Dijksterhuis 2004).

While work on the natures of epistemic feelings and epistemic emotions has often emphasized the ways in which they can lead people astray in the course of epistemic inquiry and self-evaluation, considering their complementary roles in the maintenance of schematic knowledge representations suggests that the best strategies for epistemic self-management will involve judgments based on epistemic feelings that are sensitive to extraneous cues *and* the reflective, emotion-based inquiries that are implicated in updating schematic knowledge structures. In this way, reflective forms of epistemic inquiry indirectly modulate the probabilistic assessments of knowledge structures, and the epistemic feelings that result from those assessments. In many cases, we *know* things that we would doubt on the basis of conscious reflection, but epistemic feelings provide us with positively valenced feelings of confidence to keep searching. And when those feelings are negatively valenced, they prompt the self-attribution of false beliefs associated with epistemic emotions like interest, curiosity, and surprise that prompt the sorts of epistemic inquiries needed to bring representations into line with the world.

References

Abelson, R. P. (1981). Psychological status of the script concept. *American psychologist, 36*(7), 715.

Arico, A., Fiala, B., Goldberg, R. F., & Nichols, S. (2011). The Folk Psychology of Consciousness. *Mind & Language, 26*(3), 327-352.

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

Brown, A.L. (1987). Metacognition, executive control, self-regulation, and other mysterious mechanisms. In F. E. Weinert and R. H. Kluwe (Eds) Metacognition, Motivation, and Understanding (Hillsdale, NJ: Lawrence Erlbaum), 65–116.

Carruthers, P. (2009). Mindreading underlies metacognition. *Behavioral and brain sciences*, *32*(2), 164-182.

Carruthers, P. (2017). Are epistemic emotions metacognitive?. *Philosophical Psychology*, *30*(1-2), 58-78.

Castel, A.D., McCabe, D.P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14*(1), 107-111.

Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, *72*(3), 237-267.

Damasio, A. R. (1994). Descartes' error: Emotion, rationality and the human brain.

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

Dennett, D.C. (1981). Three Kinds of Intentional Psychology. *In* R. Healy (ed.) *Reduction, Time, & Reality.* Cambridge: Cambridge University Press.

Dennett, D.C. (1987). *The Intentional Stance.* Cambridge, MA: MIT Press.

DeSousa, R. (2008). Really, what else is there? Emotions, value and morality. *Critical Quarterly*, *50*(4), 12-23.

Dijksterhuis, A. (2004). Think different: the merits of unconscious thought in preference development and decision making. *Journal of personality and social psychology*, *87*(5), 586.

Dunlosky, J., & Metcalfe, J. (2009). Metacognition: A textbook of cognition, educational, life span, and applied psychology. *Thousand Oaks, CA: Sage.*

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist, 34*(10), 906.

Fletcher, L., & Carruthers, P. (2012). Metacognition and reasoning. *Phil. Trans. R. Soc. B*, *367*(1594), 1366-1378.

Goldie, P. (2002). *The emotions: A philosophical exploration.* Oxford University Press.

Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. In *Psychology of learning and motivation* (Vol. 41, pp. 293-320). Academic Press.

van Kesteren, M. T., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in neurosciences*, *35*(4), 211-219.

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge.

Koriat, A., & Nussinson, R. (2009). Attributing study effort to data-driven and goal-driven effects: Implications for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1338.

Mandler, J. M. (1984). Scripts, stories and scenes: Aspects of schema theory. Hillsdale, New Jersey: Lawrence Erlbaum.

Meylan, A. (2014). Epistemic emotions: a natural kind?. *Philosophical inquiries, 2*(1), 173-190.

Michaelian, K., & Arango-Muñoz, S. (2014). Epistemic feelings, epistemic emotions: Review and introduction to the focus section. *Philosophical inquiries, 2*(1), 97-122.

Minsky, M. (1977). Frame-system theory. *Thinking: Readings in cognitive science,* 355-376.

Morton, A. (2010) Epistemic Emotions. In P. Goldie (Ed.) *The Oxford Handbook of Philosophy of Emotion.*

Nagel, J. (2007). Epistemic intuitions. *Philosophy Compass, 2*(6), 792-819.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in cognitive sciences, 12*(6), 237-241.

Proust, J. (2012). Metacognition and mindreading: one or two functions. *Foundations of metacognition,* 234-251.

Proust, J. (2015). *The representational structure of feelings.* Open MIND. Frankfurt am Main: MIND Group.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive psychology, 19*(1), 90-138.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, memory, and cognition, 18*(3), 435.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory.

Robbins, P. & Jack, A.I. (2006). The Phenomenal Stance. *Philosophical Studies, 127,* 59-85.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory In RC Anderson, RJ Spiro, and WE Montague (Eds.), *Schooling and the acquisition of knowledge.*

Rumelhart, D. E. (1980). Schemata: the building blocks of cognition. In R. J.Spiro, B.C. Bruce, and W.F. Brewer (Eds.). *Theoretical Issues in Reading Comprehension.* Hillsdale, New Jersey: Lawrence Erlbaum.

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition, 39*(5), 737-749.

Shank, R. C., & Abelson, R. (1977). Plans, scripts, goals and understanding. Hillsdale, New Jersey: Lawrence Erlbaum.

Schwartz B, 2002, Tip-of-the-Tongue States, Phenomenology, mechanisms and lexical retrieval, London, Lawrence Erlbaum Associates, Publishers.

Stepper S, and Strack F, 1993, "Proprioceptive determinants of affective and nonaffective feelings", Journal of Personality and Social Psychology, 64, 211- 220.

Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron, 59*(2), 336-347.

Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2012). An analysis of the determinants of the feeling of knowing. *Consciousness and cognition, 21*(4), 1681-1694.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., ... & Morris, R. G. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science, 333*(6044), 891-895.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology, 5*(2), 207-232.