Synthese
https://doi.org/10.1007/s11229-021-03094-8

Epistemic feelings, metacognition, and the Lima problem

Nathaniel Greely[1]

Received: 12 May 2020 / Accepted: 19 February 2021

Abstract
Epistemic feelings like tip-of-the-tongue experiences, feelings of knowing, and feelings of confidence tell us when a memory can be recalled and when a judgment was correct. Thus, they appear to be a form of metacognition, but a curious one: they tell us about content we cannot access, and the information is supplied by a feeling. Evaluativism is the claim that epistemic feelings are components of a distinct, primitive metacognitive mechanism that operates on its own set of inputs. These inputs are heuristics that correlate with the presence of mental content that can't be accessed directly. I will argue that evaluativism is unmotivated, unsupported, and ill-conceived. I will critique the philosophical and empirical arguments for evaluativism and conclude that there is no reason to posit a distinct mechanism to explain epistemic feelings. I will conclude, however, that epistemic feelings may constitute a nonconceptual form of metacognition, which if true is a significant claim.

Keywords Epistemic feelings · Evaluativism · Feelings of confidence · Feelings of knowing · Introspection · Metacognition · Tip-of-the-tongue experiences

1 Introduction

Do you know the capital of Peru? You may have come up with the answer, "Lima," and felt confident that this was correct. You may have failed to come up with the answer immediately but felt that you *did* know it. You may have even felt that the answer was on the tip of your tongue.[1] Each of these is an experience that researchers have termed an 'epistemic feeling'. *Feelings of confidence* are retrospective, indicating that a previous response was correct. *Feelings of knowing* are prospective, indicating that a correct response could be given. *Tip-of-the-tongue experiences* (TOTs) indicate that the correct response is tantalizingly close. These feelings have

[1] This example is taken from Dokic (2012).

✉ Nathaniel Greely
ngreely@ucsd.edu

[1] Department of Philosophy, University of California San Diego, La Jolla, California, USA

Published online: 10 March 2021

been dubbed 'epistemic' because they assist us in the retrieval and management of mental states that admit of accuracy. They tell us when our judgments were correct, when our memories or other cognitive capacities have the potential for successful performance on some task, and how likely that success may be. If epistemic feelings are about our epistemic states in the intentional sense of 'about', then they are also metacognitive states—they take other mental states as their intentional objects. This has some phenomenological support. A TOT, for example, seems to be telling us that a memory is there to be retrieved. And empirical studies show that epistemic feelings are typically correct (for feelings of knowing, see Hart, 1965; Koriat, 1993, pp. 609–610; for TOTs, Brown, 1991; for confidence, Siedlecka et al., 2016), which suggests that epistemic feelings do function to relay information about the content of our own minds.

This poses a problem for some prominent accounts of metacognition. Epistemic feelings inform us about mental content that we cannot access, a puzzle for theorists who explain metacognition in terms of direct access to our minds (e.g., Locke, 1689/1975; Lycan, 1996) or redeployment of first-order content (e.g., Byrne, 2005). And the work is done by feelings, which are not traditional elements of a theory of mind (e.g., Carruthers, 2011; Gopnik, 1993). Call this set of challenges the Lima problem.

The Lima problem motivates an account of epistemic feelings known as 'evaluativism'. Evaluativists claim that epistemic feelings reveal a distinct metacognitive mechanism, which they associate with "System 1" of dual-process theory (Arango-Munoz, 2011; Koriat, Nussinson, Bless, & Shaked, 2008; Proust, 2013).[2] Among the proponents of evaluativism are Proust (2013), Dokic (2012), Arango-Munoz (2011, 2013, 2014a, 2014b, 2019), and Koriat (1993, 1997, 2000, 2006).[3] While these researchers' accounts of epistemic feelings differ in some respects, they all commit to a functional claim, which in turn supports their claims of a distinct mechanism. They claim that the information epistemic feelings provide about our mental states is not derived directly from those states but from heuristics. Call this claim *core evaluativism*. Core evaluativism is supported by a body of empirical research that purports to show that these heuristics are the causal factors to which epistemic feelings are sensitive (e.g., Koriat, 1993; Koriat & Levy-Sadot, 2001; Reder & Ritter, 1992). The most commonly cited heuristics include cue familiarity, recall of related information, and processing fluency.

I will argue that core evaluativism is unmotivated, unsupported, and ill-conceived. The structure of the article is as follows. In Sect. 2 I will introduce the phenomenon of epistemic feelings and discuss a range of views about their nature. I will then consider the Lima problem and argue that it is no problem at all, eliminating one source of motivation for evaluativism. But evaluativists also offer empirical arguments in

[2] Some evaluativists (e.g., Proust, 2013) reserve the term 'metacognition' for the System 1 version. This form of metacognition is described in detail in Sect. 2.3. I will use 'metacognition' to denote any intentional relation between mental states or processes. Thus, I will refer to theory-theoretic processes (Sect. 2.2) as 'metacognition', whereas evaluativists would use 'metarepresentation'.

[3] Dokic (2012, p. 312, note 16) stops short of committing to the existence of a distinct mechanism, but does commit to what I will call 'core evaluativism'.

support of core evaluativism. In Sect. 3 I will describe the three primary heuristics posited as the causal inputs to epistemic feelings– cue familiarity, related information, and fluency. In Sect. 4 I will evaluate the relevant empirical evidence and conclude that it is does not support core evaluativism. I will then re-examine the proposed heuristics and argue that, as conceived, they are not adequately distinguished from the content they are claimed to replace. One exception is fluency conceived as reaction time, but I will argue that this heuristic is flawed in a different way. In Sect. 5 I will conclude by arguing that it is not necessary to posit a distinct mechanism for epistemic feelings, and that instead emphasis should be placed on the possibilities for nonconceptual metacognition suggested by epistemic feelings.

2 Epistemic feelings, the lima problem, and evaluativism

2.1 Epistemic feelings

My argument will not require a strict definition of epistemic feelings. Nor have I encountered an adequate one in the literature.[4] I have described a few epistemic feelings – feelings of confidence, feelings of knowing, and tip-of-the-tongue experiences. Other examples include feelings of familiarity, feelings of ease of learning, perceptual confidence, and many more (see de Sousa, 2009; Dokic, 2012; Dellantonio & Pastore, 2019 for longer lists).

Beyond a degree of overlap in the extension of the term, there is little agreement on the characteristic properties of epistemic feelings – even their epistemic nature and their status as feelings. Epistemic feelings are "epistemic" because they are involved in our cognitive processes in a way that admits of accuracy, but the manner of their involvement is a matter of dispute. Some theorists, including evaluativists, characterize many epistemic feelings as metacognitive, in that they take epistemic states as their intentional objects. A TOT or a feeling of knowing (FOK), for example, can tell us that we *know* the capital of Peru. Others take their epistemic role to be primarily first-order, aiding us in decision-making when purely rational deliberation is too time-consuming or costly (DeSousa, 2009; Carruthers, 2017). DeSousa (2009), for example, describes fear as an epistemic feeling that estimates objective degrees of risk in the environment. Carruthers (2017) denies even that FOKs are metacognitive.

Some epistemic feelings seem indisputably felt. TOTs, in their folk understanding, have a strong phenomenal component. But confidence and feelings of knowing (despite the name) are less obviously phenomenal. They are typically operationalized as predictions or evaluations of one's own performance on an experimental task, and as such are indistinguishable from judgments lacking any phenomenal component. Some characterize epistemic feelings as emotions (DeSousa, 2009; Carruthers, 2017), but it is not at all clear that a subject's confidence rating on

[4] Arango-Munoz (2014a) offers this: "E-feelings are phenomenal experiences that point towards mental capacities, processes, and dispositions of the subject, such as knowledge, ignorance, or uncertainty" (p. 158). We will soon see that every part of this definition is controversial.

the twentieth forced-choice trial reporting the orientation of a Gabor grating is an emotional experience. Despite this general climate of disagreement and ambiguity, evaluativists agree on a significant and controversial claim about epistemic feelings, which I call core evaluativism. This claim is motivated in part by the Lima problem, to which I turn in the next section.

2.2 The lima problem

Dokic (2012) poses the Lima problem for transparency and direct access accounts of metacognition. Transparency theory (Byrne, 2005) holds that knowing whether I believe *p* is simply a matter of determining whether *p* is the case and then applying an "ascent routine." That is, when asked whether I *believe* that the cat is on the mat, I "look to the world" and employ whatever ability I would use to determine whether there *is* a cat on the mat. I then follow a simple rule: If there is a cat on the mat, conclude that I believe it. But if self-attributing knowledge that *p* is simply a matter of such an ascent routine, how can we explain the Lima example? In the Lima case I cannot report the first-order proposition, that Lima is the capital of Peru, and thus it would seem I cannot apply an ascent routine to draw any conclusions regarding my beliefs about it.

But transparency theory is more flexible than Dokic allows. The "world" to which transparency theorists look must certainly include questions asked of the subject. The real issue is how to account for the general accuracy of the resulting beliefs. The ascent routine could look something like: 'If asked *x* and *y* occurs, believe that you know the answer to *x*', where *y* is an FOK. Of course, if one conceives of an FOK as a mental state then the ascent routine fails, and it might seem obvious that an FOK is a mental state. It's a *feeling*, after all. But such tricky cases are the transparency theorist's bread and butter. All that is needed is a first-order characterization of y that picks out the same phenomenon. We can find a precedent in Byrne's ascent routine for thought—"If the inner voice speaks about *x*, believe that you are thinking about *x*" (2008, p. 117). Byrne characterizes the inner voice qualitatively, as "degraded" relative to outer voices (p.118). Including an epistemic feeling in the ascent routine is no different in principle. One might attempt a qualitative characterization of the FOK but general agreement on those qualities seems unlikely.[5] And a qualitative characterization of FOKs wouldn't *explain* the accuracy of *y*. Carruthers (2017) offers a first-order characterization of FOKs based on partial recall, and this could both serve in an ascent routine and potentially offer a non-evaluativist account of the accuracy of the resulting belief. A similar solution is suggested in the course of Dokic's argument against direct access solutions to the Lima Problem, to which I now turn.

[5] Dokic claims that epistemic feelings register "internal physiological conditions and events" (2012, p. 307). Perrin, Michaelian, and Sant'Anna (2020) offer a phenomenological description of the feeling of remembering as one of "pastness, self, causality, and singularity." In either case the properties picked out are arguably non-mental.

While transparency theory holds that knowledge of our mental states is obtained through knowledge about the world, other theories posit a direct informational channel to one's own mind. A classic example is inner sense. Inner sense accounts of metacognition hold that we become aware of our own first-order mental states through a quasi-perceptual mechanism within the mind (e.g. Locke, 1689/1975; Lycan, 1996). Lycan identifies the inner sense with attention and applies it to both propositional attitudes and sensory states. Carruthers (2011) opposes inner-sense views of propositional attitudes but does hold that we directly perceive mental imagery (e.g., subvocalization), which facilitates knowledge of propositional attitudes. Pitt (2004) holds that there is a proprietary phenomenology of propositional thought that makes it directly introspectable. The Lima problem here is that, whatever the mode of direct access, it's not clear how a direct informational channel could tell me *that* something is in my mind without telling me *what* it is. If I believe that I know the capital of Peru on the basis of directly perceiving that I have precisely *that* information, then I ought to be able to produce it.

But these considerations are not fatal for direct access. Dokic himself considers a solution but rejects it. It might be the case that, in the Lima example, the subject is directly aware of only part of the first-order belief, say 'The capital of Peru is ___.' Dokic argues that this strategy is incompatible with a common view of introspection.

> … introspection makes the subject aware of her own intentional mental states only by revealing their contents (see, e.g. Tye 2009). In other words, introspection is *fully transparent* with respect to the contents of the introspected states (whenever they have contents). The Direct Access Model denies that introspection is always transparent in this sense, since feelings of knowing are precisely introspective states about particular first-order memories, while their contents are only partially revealed to the subject. (p. 306)

The argument seems to be that, if the only input to inner awareness is first-order mental content, then the introspected content can't have gaps. But the consequent only follows if one equates transparency with exhaustivity. Consider outer perception. The view that perception is transparent with respect to its object is compatible with the view that perception is not exhaustive. I can perceive part of an object, or only some of its properties, such as its shape but not its color in dim light. Likewise, a direct informational channel might reveal only partial information about the content of my mind, even if it is transparent with respect to what it does reveal. A similar response might be available to the transparency theorist as well: If asked x and you produce a partial answer to x, then you know the answer to x.

This general type of solution to the Lima problem exists in the psychological literature as well. According to Brown and MacNeil (1966), TOTs occur when we access semantic information, the meaning of the word, in the absence of phonological information. Here again we access some first-order information directly, and that is enough to let us know that the information is there even if it doesn't allow us to report it. This type of account can easily be extended to FOKs as well. Nelson, Gerler, and Narens (1985) offer further possibilities compatible with direct access. It may be that "associative strength," a posited relation developed between the cue and the target, is the source of epistemic feelings (pp. 295–296). When the strength

of the association is above a certain threshold, recall occurs. When the strength is lower, recall fails but there is a FOK. When the strength is even lower, there is neither recall nor feeing of knowing.[6] A third possibility is that access is "multidimensional," in that various informational properties of the memory trace are accessed, but not the word form itself.[7][8] The Lima problem, then, does not motivate abandoning transparency theory or direct access for evaluativism.

Epistemic feelings pose a different problem for theory-theoretical accounts of metacognition. Theory-theory (Gopnick, 1993) is the claim that we come to know our own mental states much as we do the mental states of others – by observing our behavior and deriving metacognitive conclusions using a theory of mind. Even if I don't recall the capital of Peru, I might infer, based on my knowledge of myself, my history, and my abilities, that I do know it. But if, in the case of epistemic feelings, the vehicle of higher-order information is a feeling, this appears incompatible with its being an inference from a theory of mind (Dokic, 2012, p. 304). Much depends here on how one characterizes feelings, of course. If we mean emotions, and emotions are conceived as cognitive states, then there is no problem. And however we conceive these feelings, it is a commonplace that feelings can result from propositional knowledge. The knowledge that a loved one has died (or an inference to that effect) will produce a feeling of sadness, so it might be that an epistemic feeling is the result of (possibly unconscious) inferences from a theory of mind. Theory-theoretic explanations of the Lima problem are also threatened by the possibility that animals lacking a theory of mind nonetheless demonstrate metacognitive abilities facilitated by epistemic feelings (Proust, 2013, ch. 5). But the evidence for animal metacognition is controversial (see Carruthers, 2008 for a skeptical view), as is the claim that animals lack a theory of mind (see, e.g., Emery et al., 2004).

While the Lima problem is offered as a threat to some philosophical theories of metacognition, it is not fatal. Perhaps for these reasons, when evaluativists make such arguments, they invariably bolster them with empirical evidence for a more general claim about epistemic feelings. This claim is that epistemic feelings are components of a distinct metacognitive mechanism that takes heuristics as its inputs. If this claim could be supported, it would rule out transparency, direct access, and theory-theoretical explanations of the Lima problem and instead solve it in a novel way. In the next section I will describe this claim in detail.

[6] Whether associative strength is a direct access account may depend on how one conceives the metaphysics of the relation.

[7] The traditional claim that memory consists in a "trace" of an original event stored in memory is contested by those who take memory to be an entirely reconstructive process (e.g., Michaelian, 2016). But the issue here is whether memory is the informational source of epistemic feelings, not whether that source is reconstructed or stored.

[8] Hart (1965) and Nelson and Narens (1990) are also cited by evaluativists as the sort of psychological direct-access accounts they oppose (Arango-Munoz, 2019; Proust, 2013). But Hart's claims about a memory "monitor" are too brief and general to characterize as direct access, and while Nelson and Narens sometimes seem to favor direct access (p. 150) at other times they sound like evaluativists (p. 158).

2.3 Evaluativism

The term 'evaluativism' is used by Proust (2013) to characterize her view of the function of epistemic feelings ('noetic' feelings in her parlance). Epistemic feelings, for Proust, are metacognitive in that they serve to *evaluate* a subject's own cognitive processes in terms of success or failure. Proust claims that epistemic feelings are the output of a phylogenetically primitive form of metacognition, which she calls 'procedural' metacognition. Procedural metacognition is distinct from forms of metacognition that employ propositional attitudes, instead operating in a nonconceptual format. This distinctive form of metacognition does not receive direct information from memory, but instead takes as its inputs simple cues ("heuristics") that merely tend to correlate with cognitive success or failure. A FOK, for Proust, tells me that recall is possible not because it receives direct information that the item is in memory, but because it receives information that the recall process is operating "fluently." Fluency will be discussed in detail in Sect. 3, but the basic notion is intuitive enough. A fluent process, be it speaking in a foreign language, swinging a bat, or retrieving an item from memory, is one that runs smoothly and without difficulty. This, for Proust, is what epistemic feelings report. And insofar as fluent mnemonic processes tend to be successful ones, epistemic feelings can be said to report on the potential for successful recall, and thus indirectly on the presence of the item in memory. Other evaluativists, we will see, favor other heuristics, but in each case the functional picture is similar. Call this claim *core evaluativism*—epistemic feelings are metacognitive but their direct inputs are not the cognitive states that they are "about." Rather, they take as inputs cues and heuristics that tend to correlate with the presence or absence of such states.

If the inputs to epistemic feelings are distinct from the mental content on which they report, then the Lima problem is solved. We know *that* we know without being to report *what* we know because the process doesn't involve access to the content in question. And if the inputs aren't propositional, this opens up an informational role for feelings. This solution to the problem also sits well with an influential current in philosophy and psychology. Dual process theories of cognition (Evans & Stanovich, 2013) hold that many of the mental abilities once thought to be governed by rational, epistemically grounded processes are often accomplished by more phylogenetically primitive systems using unconscious cues and heuristics that merely produce adaptive behaviors. These more primitive cognitive abilities are grouped together as "System 1" processes in contradistinction to the more sophisticated, rational "System 2" processes. Various heuristic biases that tend to trump logical and probabilistic reasoning in first-order decision-making have been demonstrated in humans and are taken as examples of System 1 processes (see Evans, 2003 for a review). Evaluativists identify procedural metacognition as a System 1 process, and if correct this is a novel and interesting extension of dual process theory into the realm of self-knowledge.

Core evaluativism, if true, would seem to rule out transparency, direct access, and theory-theoretic solutions to the Lima problem. It preempts any access to first order content as the basis of epistemic feelings and the metacognitive abilities they facilitate. This means that partial direct access or partial redeployment of first-order

content cannot be the basis of these abilities. It also rules out theory-theoretic explanations insofar as these are identified with System 2.

Proust is explicit about the functional claim:

> … the cues for cognitive success have nothing to do with the particular content of the words to learn, or with the intentional content of one's first-order thoughts. They are properties of processing, not properties of content. (Proust, 2013, p. 56)

Other researchers hold this view as well. Koriat developed much of the empirical support for evalutivism in his work from the 1970s to the present.

> Whereas information-based judgments entail deliberate, analytic inferences that rely on beliefs and memories, metacognitive feelings are mediated by the implicit application of nonanalytic heuristics. (Koriat, 2000, p. 128)

Similar expressions of the view can be found in Dokic (2012).

> The cues underlying noetic feelings are contingently but stably associated with epistemic states. This association holds in a normal (ecological) context, but it can be severed by psychologists, who can easily produce 'illusory' feelings of knowing … [epistemic] feelings have intentional contents beyond the body, but only in a derived way, through some kind of learning or association process. Such a process generates new heuristics, i.e. cognitive shortcuts that enable us to move spontaneously from our feelings to judgements concerning the task at hand. (Dokic, 2012, pp. 307-308)

Arango-Munoz also commits to core evaluativism.

> … for me (following the psychological tradition of metacognition research (e.g., Reder, 1987, 1996; Koriat, 1993, 2000), the main function of low-level metacognition is to elicit E-feelings and control mental action based on cues and heuristics…In other words, the monitoring mechanism does not actually scan mental states. (2014a, p. 150)

Evaluativists differ in their accounts of the representational format of epistemic feelings. Proust (2013) invokes a nonconceptual format borrowed from Strawson (1959) and Cussins (1992). Arango-Munoz (2014a) claims that, in addition to being nonconceptual, epistemic feelings are phenomenal states, whereas Proust claims they can be unconscious, which seems to preclude their being phenomenal. Koriat and Dokic are less explicit about the representational format, though Dokic does call them "experience-based" (2012, p. 304). But evaluativists converge on core evaluativism. The empirical basis of this claim comes largely from research on mnemonic (and some perceptual) epistemic feelings. Among these I have chosen FOKs, TOTs, and confidence, as they are most commonly cited, and I believe they offer the best case for evaluativism.[9] The most commonly cited heuristic inputs to these epistemic

[9] To take one example of the weaker evidence I will skip, Arango-Munoz (2019) cites work by Whittlesea and Williams (1998, 2001) in support of the claim that feelings of familiarity are causally sensitive to fluency. But Whittlesea and Williams operationalize feelings of familiarity as false alarms on a recognition task. False alarm rates are a first-order phenomenon, not a metacognitive measure.

feelings are cue familiarity, recall of related information, and fluency. In the next section I will explain these heuristics.

3 Heuristics

Core evaulativism is the claim that epistemic feelings are not directly informed by the first-order mental content on which they report, but instead are sensitive to "cues" or "heuristics." What are these cues and heuristics? Among the most frequently cited heuristics are related information, cue familiarity, and fluency (Arango-Munoz, 2019; Dokic, 2012; Koriat & Levy-Sadot, 2001; Proust, 2013).

Related information is typically invoked in the case of FOKs and TOTs. If the subject does not recall the word form sought in a memory task, but recalls some related details, this could be what triggers the epistemic feeling. So, while I might not recall the name of the capital of Peru, I might recall its location on the map, a dish I tasted on a recent visit, or its major exports, and on this basis feel that I know the answer. The cue familiarity heuristic takes the cue itself to be the source of epistemic feelings. If I were a subject in a memory study, I might be presented pairs of countries and their capitals: Canada: Ottawa, Ecuador: Quito, Peru: Lima. Later I might be asked to recall the capital based on the cue – the name of the country. Or I might be tested on general knowledge without prior training, and the cue might be "What is the capital of Peru?". In each case the claim is that a FOK would be based on whether the cue appears familiar to me, not on whether I have the answer in mind. What makes each of these factors heuristics is that they are thought to correlate with retrieval of the correct answer often enough that, for creatures like us, they provide a pretty good guide to whether the first-order content is in memory. They are fallible, of course, but a system employing these heuristics will likely be adaptive if direct access is time and energy consuming, or psychologically impossible.

The third heuristic is fluency. Fluency is the most difficult heuristic to characterize, as it is often defined and operationalized in different ways. Theorists have defined fluency as "subjective ease" of processing (Duke et al., 2014), "speed and accuracy" of processing (Reber, Fazendeiro, & Winkleman, 2002), or as a cluster of attributes comprised of "degree of activation," "speed," and "effort" (Winkielman et al., 2003). It is most frequently applied to confidence, feelings of familiarity, and judgments of learning (JOLs). JOLs are judgments that memorization has been successful, such that a given item can be recalled in the future. The idea behind the fluency heuristic is that when a stimulus is perceived by the subject as easier to see, easier to read, or in general easier to process, the subject predicts better performance on the relevant task. If a word pair like 'up-down' is processed more fluently than 'hammer-fish', subjects are more likely to report judgments of learning for the former (see Winkielman et al., 2003 for a review). "Regular" non-words like 'hension', which resemble words and are thus presumably processed more fluently than nonwords like 'jufict' are claimed to produce feelings of familiarity (Whittlesea & Williams, 1998). And subjects are more likely to report higher levels of confidence in their answers when the cognitive processes involved are more fluent (Finn & Tauber, 2015).

The phenomenon is tricky to operationalize, and as a result there are diverse ways of doing so. Some studies treat fluency itself as an epistemic feeling and measure it by subjective report. The epistemic feeling of fluency is then argued to be a causal factor in various judgments. For example, perceptual fluency appears to affect favorability ratings for various marketing materials (Graf et al., 2017). These studies are less relevant to evaluativism for our purposes, as they study the downstream effects of epistemic feelings, not their causal inputs. Some metacognitive studies operationalize fluency by manipulating the difficulty of tasks to be performed by subjects, for example the length of a list to be recalled (Schwarz et al., 1991), the regularity of a word to be memorized (Whittlesea & Williams, 1998), or in perceptual studies the clarity of the stimulus (Feustel, Shiffrin, & Salasoo, 1983; Kelley & Jacoby, 1998). In some studies fluency is operationalized as reaction time (Benjamin et al., 1998), and often priming or other methods are used to manipulate reaction time (Whittlesea & Williams, 2001; Winkielman et al., 2003). Elsewhere priming alone is used to operationalize fluency, and it is not always clear whether this is meant as an implicit manipulation of reaction time (Jacoby, 1983). This is to say, in studies that invoke fluency it is often not clear what the heuristics are, what the epistemic feelings are, and whether the measures and manipulations employed are meant to be identified with heuristics, epistemic feelings, or causal antecedents or products of either.

To narrow the options, I will rely primarily on Proust's (2013) interpretation of fluency, since among evaluativists she places the most emphasis on it. Within her work fluency takes on many shades of meaning and I cannot do justice to that richness here. She variously describes fluency as an epistemic feeling (p. 58, 61, 102, 105), as an alternative epistemic norm to truth appropriate to nonconceptual content (p. 10, 125, 129–130, 137), as a heuristic on which other epistemic feelings are based (pp. 58–59, 62, 73, 105, 129), as the fundamental heuristic or epistemic feeling from which others develop in ontogeny or phylogeny (p. 73), as the output or operation of a mechanism known as an "adaptive accumulator" (p. 105, 129), and as a property of the neural assemblies that realize the adaptive accumulator (p. 130). All these various characterizations of fluency are not necessarily inconsistent, particularly if it is allowed that epistemic feelings can serve as heuristics for other epistemic feelings. While there is room for interpretation of her view, her argument focuses heavily on Vickers and Lee's (1998) adaptive accumulator model of decision making and on reaction time. These will be my focus in Sect. 5.

4 The empirical argument for evaluativism

Evaluativists invoke a body of empirical research that purports to show that epistemic feelings are not causally sensitive to the target content but to heuristics. Three heuristics that evaluativists most frequently cite are cue familiarity, retrieval of related information, and fluency (e.g. Arango-Munoz, 2019; Dokic, 2012; Koriat & Levy-Sadot, 2001; Proust, 2013). In this section I will consider the empirical evidence for each in turn and argue that in each case evaluativism is not supported. Other heuristics have been cited as potential inputs to epistemic feelings (e.g., study

time as the heuristic for feelings of learning [Koriat & Ackerman, 2010]), but space only permits discussion of the most prominent. In Sects. 4.1 and 4.2 I will critique the evidence for the cue familiarity and related information heuristics, respectively. In Sect. 4.3. I will argue that evaluativism is ill-conceived, in that it fails to make a clear distinction between these first two heuristics and first-order content. In Sect. 4.4. I will critique the fluency heuristic and argue that the relevant models are best interpreted as positing direct access to first-order content.

4.1 Cue familiarity

Reder and Ritter's (1992) study is cited ubiquitously by evaluativists in support of their view (e.g., Arango-Munoz, 2013, 2019; Dokic, 2012; Koriat & Levy-Sadot, 2001; Proust, 2013). Although the study is rather old and, as we shall see, flawed, it plays an outsized role in evaluativist arguments. It is the only study cited by Dokic (2012) in support of the cue familiarity heuristic and is lauded by Koriat and Levy-Sadot as providing "remarkable support" for cue familiarity (2001, p. 35). Arango-Munoz (2013) also describes the study in detail as part of his argument.

Reder and Ritter (1992) trained subjects on a series of arithmetic problems that are difficult to work out mentally (e.g., 23 × 27). In experimental trials these same problems were presented, and subjects were offered a choice of answering strategy. They could choose to retrieve the answer from memory or to calculate the answer. If they chose retrieval, they were given less time (less that 1 s) but more points for a correct answer produced within that time. If they chose to calculate, they were given more time (~ 15 s) but fewer points for a correct answer within that time. The authors took a subject's choosing the retrieval option as an indication that they had experienced a FOK. FOKs did correlate with successful retrieval. The authors hypothesized, however, that the choice of the retrieval option was based on the familiarity of the question parts rather than the availability of the answer in memory. This was tested by examining trials in which question parts were the same as parts of previously seen problems, but in which the answers differed (e.g., 23 × 16, having previously seen 23 × 27 and 16 × 27). The authors found that choice of retrieval strategy showed a stronger correlation with previous exposure to the question *parts* than with correct responses. They concluded that cue familiarity, and not access to first-order content, drives FOKs.

While many researchers accept the results of this study uncritically (e.g., Herzog et al., 2010; Hosey, Peynircioglu, & Rabinovitz, 2008; Paynter et al., 2009; Walsh & Anderson, 2009), Koriat himself (1993) notes early on that many researchers commit the error of conflating what he calls "subjective" and "objective" properties of memory. A FOK, in the context of a typical memory study, is not understood merely as a feeling that the subject knows the contents of her memory. It is a feeling that she knows that the contents of her memory match the problem given in training. This is an artifact of how we often conceptualize memory. Memory, like perception and knowledge, is treated in such studies as a factive state. In order to remember I must not only be able to access a mental representation, but that representation must also

match the original stimulus. But this means that there are two ways a subject can fail on these tasks, and only one of them is a metacognitive failure.

Suppose that I am a subject in a similar study, and rather than being presented arithmetic problems and their answers I am presented with pairs of countries and their capitals. And suppose that, being a little mixed up after this barrage of information, I form the false memory (or belief if you like) that Quito is the capital of Peru. I may report a FOK but produce the wrong answer. In the studies in question this would be counted as a case in which FOKs do not correlate with metacognitive access to my first-order states. But this is not what happened in the proposed scenario. I did access my memory and my response was based on that content—the content just happened to be false. Furthermore, my FOK will correlate with the cue ('Peru'), but not for any deep reason. It will be an artifact of my having a false belief involving that cue stored in memory.

A similar analysis holds for the Reder and Ritter study. A subject who chooses the retrieval option, but answers incorrectly, may well have an answer in mind, it's simply incorrect. Many of us have experienced misleading TOTs. When we finally retrieve the name that was on the tip of the tongue, we find out that it was incorrect. And I may choose the retrieval option more frequently when I have already viewed parts of the answer because I incorrectly encoded problems I had already seen.

Of course, the analogy between arithmetic problems and the Peru question limps, and so the evaluativist analysis of this study, as well as my counter-analysis, may feel awkward. Complex math problems are not the sort of thing we typically encode in memory for any length of time, and thus it is difficult to determine which explanation of the data is more plausible. But even if one dismisses my analysis, the most that can be concluded is that in situations in which we are required to memorize things we don't typically tend to memorize, like difficult arithmetic problems, we might rely on cue familiarity. This falls far short of full-throated evaluativism.

Other studies of the cue familiarity heuristic suffer from similar problems. Metcalfe et al. (1993) use a paradigm from interference theory in which the word pairs to be memorized are provided alongside other word pairs designed to interfere with encoding and reduce memorability (e.g., the cue paired with a synonym of the target, the cue paired with a different word, or a word pair containing neither the cue or the target). Metcalfe et al. predict that if cue familiarity is the source of FOKs, then FOKs should occur in proportion to the frequency of appearance of the cue, not to memorability. The results of their study, even on its own terms, are rather ambiguous. In recognition tasks, subjects fail to perform in the way the interference paradigm traditionally predicts, and as a result, FOKs end up tracking recognition performance nearly as well as cue frequency. But the fundamental problem with the study is the same as in Reder and Ritter – the researchers identify direct access to memory with accurate performance on the recognition task. As we have seen, these are different things.

Other measures of direct access have been employed, but these are also insufficient to distinguish direct access from accuracy. There is room only to survey a few to demonstrate the variety of measures employed. Liu et al. (2007) manipulate target retrievability by repeating instructions to remember the target. Metcalfe and Finn (2008) manipulate target retrievability by providing multiple cues for a single

target (high retrievability) or a single cue for multiple targets (low retrievability). While each of these manipulations might increase the likelihood of accurate recall, in each case the low retrievability option is consistent with incorrect encoding and direct access to that incorrectly encoded content. Indeed, Metcalfe and Finn's low retrievability condition plausibly encourages incorrect encoding by pairing the cue with multiple targets.

Hanczakowski et al. (2013) recognize some of these difficulties and rather than attempt to measure or manipulate memorability simply measure the effect of priming the cues (a manipulation of cue familiarity) on FOKs regardless of objective performance. They show a significant effect of priming on FOK magnitude. This has the advantage of not identifying direct access with accuracy, but without *some* independent measure of direct access we can't compare the *relative* effects of the two variables, nor can we rule out the possibility that the manipulation of cue familiarity increased FOKs by increasing the accessibility of the target (or of content mistaken for the target).

There are many more studies involving cue familiarity than I can cover here. More recent studies often take evaluativism as given and instead focus on the relative contribution of various forms of cue familiarity and related information to FOKs, largely ignoring direct access as a contender.[10] Cue familiarity and partial information have been posited as responsible for different stages (Koriat & Levy-Sadot, 2001), different types (Liu et al., 2007), and different aspects (Isingrini et al., 2016) of FOKs. I will address some of these studies in the next section and argue that they do not offer methodological improvements.

4.2 Related information

Koriat (1993) is aware of the potential for ambiguity in the study of epistemic feelings. He notes a distinction between the objective and subjective factors involved in memory and agrees that ignoring the latter can cause researchers to draw hasty conclusions about the nature of epistemic feelings. He attempts an improved methodology and employs it in studies that purport to show that the related information heuristic is the causal factor to which FOKs are sensitive. I will describe and critique Koriat's seminal (1993) study as well as some more recent ones.

Koriat (1993) presents subjects with nonsensical four-letter stimuli (e.g., JKSD). Subjects are then asked to recall the letters, and the number of letters they recall in each case is recorded, along with reports of FOKs. It was found that the number of letters subjects were able to recall, whether correct or incorrect, correlated with FOKs. Although he is explicitly sensitive to the worries I describe in the last section, Koriat's study is not an improvement. The correlation he finds is consistent with the possibility that subjects sometimes incorrectly encode the letters to be

[10] To take one example, Hertzog et al. (2010) examine varieties of the partial information heuristic, claiming that "All extant theories of FOKs reject the idea that individuals have direct access to information held in memory (p, 772)."

recalled, report a FOK on the basis of the stored (incorrect) letters, and then report those letters.

Later studies do not improve matters. Koriat and Levy-Sadot (2001) manipulate the amount of related information by asking questions about categories with many recallable members (Who *composed* Swan Lake?) versus categories with fewer such members (Who *choreographed* Swan Lake?). The researchers suggest that larger categories should make direct access of the target less likely, since there are more competing items in memory. If so, a direct access account would predict fewer FOKs in such conditions (p. 38), whereas the partial information heuristic would predict more FOKs, given the larger number of recallable items (e.g., composers). But of course, the reason that people know fewer choreographers is that they tend to be less famous, so it's equally plausible that a direct access account would predict more FOKs in those conditions. Subjects are likely to know *some* composer, and even if it's not the correct one it may trigger an FOK.

More recent studies have expanded the partial information heuristic, positing an influence on epistemic feelings from a wide variety of factors, including emotional content of items to be remembered (Schwartz, 2010) or other contextual aspects of the encoding process (Hertzog et al., 2014), but none improve measures of direct access. Indeed, some of these studies are explicitly pluralistic, allowing for a degree of direct access (Schwartz, 2010). Others ignore direct access as a possibility altogether (Hertzog et al., 2014).

4.3 Heuristics or direct access?

But even if a better measure of direct access could be devised, there is a fundamental ambiguity in the related information heuristic. Koriat makes the point himself, noting that "the cues for the FOK are to be found in the very information that is activated or accessed during the course of the search-and-retrieval process" (1993, p. 611).[11] When the information happens to be correct, we have normal recall. When the information is incorrect, it can still generate FOKs. But in either case the functional architecture is the same. The heuristic that evaluativists claim operates in lieu of access to the relevant first-order content turns out to be access to *fragments* of the relevant first-order content. Recall Koriat's (1993) study, in which the related information is simply a group of letters. If I am attempting to retrieve the capital of Peru from memory and retrieve 'Lim_', is this a heuristic or just three-fourths of the answer? I think we should say the latter, as this differs not at all from the inner-sense account that Dokic rejects. And if instead I retrieve "Qui_" and have a FOK, I have argued, and Koriat appears to agree, that to take this as evidence for evaluativism is to conflate the objective and subjective factors of memory.

To be clear, evaluativists do not deny that we access our memories. But they do claim that epistemic feelings are not informed by the target content. Whether the related information heuristic counts as a distinct source of information, then, will

[11] Schwartz and Metcalfe (2010) also point out that the partial information heuristic is "compatible" with direct access.

depend on one's theory of mental content. Consider the Lima example. Like Reder and Ritter's arithmetic problems, nonsense strings of letters are not entirely analogous to everyday cases of epistemic feelings. In typical cases of TOTs and FOKs, the partial information being recalled is not only letters in the name. I may not be able to recall the name of the capital of Peru, but I might be able to picture its location on a map, name some of its neighborhoods, and recall an NPR segment on its cuisine. If we compare these bits of recalled information with Brown and MacNeil's (1966) direct access account, on which subjects access an amodal "pure" meaning, there might appear to be a principled distinction between heuristic information and the target content. Perhaps the content proper is the meaning and everything else is heuristic information. This apparent foothold for evaluativists, however, becomes much more precarious under scrutiny, for related information may be *part* of the meaning. While Brown and MacNeil seem to imply a distinction between information in a phonological format and a "pure," amodal meaning, the actual studies evaluativists cite make no attempt to determine the format of the related information, nor do any evaluativists I know of define related information in terms of such format.[12] One could do so, but different empirical studies would be required to support that account.

And there are reasons evaluativists should want to avoid committing to such specifics. To pursue this sort of gambit the evaluativist must draw a clean line between the meaning of 'Lima' and even related *amodal* information. This amounts to an a priori rejection of any holistic account of mental content. The target content – the meaning of 'Lima' – must be considered distinct from content like 'is the capital of Peru'. Thus classical, definitional views of concepts must be rejected, inferential role accounts, theory-theoretical accounts, activation in a semantic network, and so on. Evaluativism becomes a branch of semantic atomism. It would no longer be a general account of the functional architecture of epistemic feelings. This, I take it, is not evaluativists' aim.

And this analysis can be applied to the cue familiarity heuristic. Cue familiarity is plausibly interpreted as a special case of related information since the cue is a bit of related information available to the subject. It is part of the definition, semantic network, or whatever one takes the content of 'Lima' to be, on a holistic view of content, that it is the capital of Peru. The Reder and Ritter study takes this to an extreme. The math problems presented are not integrated into a semantic network in the same way as knowledge of countries and their capitals, or even the basic multiplication tables. In this case the numbers used as the cues are the *only* relevant associated information available, at least for the non-mathematician. Other studies often use similarly artificial pairings of cues and answers (e.g., Liu et al., 2007). The cue familiarity hypothesis, then, is consistent with the claim that related information is the source of epistemic feelings, it's just that sometimes there is relatively little related information available—just the cue. Recent research points in this direction.

[12] Carruthers (2011) makes such a distinction and claims that access to one's own propositional attitudes is indirect, but he is no evaluativist. He denies that epistemic feelings have metacognitive content and suggests a direct causal relationship between memory and action for feelings of knowing (2017).

Thomas, (2012) found that when participants were asked to focus on semantic properties of a cue during encoding, as opposed to "shallow" properties like the color of the type, FOK judgments were more accurate. Koriat and Bjork (2005) find that JOLs are inflated when there is a strong but unexpected semantic association between cue and answer (e.g., 'find-seek' as opposed to 'find-lose').

If cue familiarity is just related information, and related information is just first-order content, then the fact that these so-called "heuristics" give us information about the content of our minds does not support evaluativism. Unless one adopts a very specific semantic theory, evaluativism collapses into a direct access account. There are influential atomic accounts of content (Dretske, 1981; Fodor, 1998; Millikan, 1987), but the heyday of such views appears to be in the past, and to yoke evaluativism to them is to sacrifice its status as a general account of epistemic feelings. It seems, then, that core evaluativism is not only unsupported by the empirical evidence, there is a fundamental problem in its very conception.

4.4 Fluency

While Dokic and Koriat favor cue familiarity and related information as the source of epistemic feelings, Proust has a different account. She holds that epistemic feelings reflect varying degrees of "fluency." In Sect. 3 I noted the variety of interpretations of fluency in the empirical literature. Proust does an impressive job attempting to interpret and unify this unruly concept, but as I also noted in Sect. 3, questions remain for her account. Here I will consider two of the most promising interpretations of Proust's fluency heuristic. The first conceives fluency as the operation or output of an adaptive accumulator (Vickers & Lee, 1998). I will argue that the adaptive accumulator model does not in fact support evaluativism, nor do more recent models of the same phenomena. Elsewhere Proust seems to describe fluency entirely in terms of a monitoring of reaction time (e.g., pp. 129, 136). Therefore, it is also worth considering reaction time itself as the relevant heuristic. I will argue that the empirical evidence does not support evaluativism on either interpretation of the fluency heuristic.

4.4.1 Fluency as operation of an adaptive accumulator

Proust clearly considers fluency, as the heuristic for feelings of confidence, to be intimately related to the operation of an "adaptive accumulator" as proposed by Vickers and Lee (1998). In this section I will consider whether this model supports an evaluativist account of metacognitive confidence.

Vickers and Lee's (1998) double accumulator model is grounded in signal detection theory (SDT)—one of a class of models that enriches the traditional SDT framework to account for dynamic aspects of reaction time and the metacognitive phenomenon of confidence. Space does not permit a detailed description of either traditional SDT models or their dynamic successors, but the fundamental point of relevance is that such models posit a noisy internal signal that serves as evidence about the state of the world and is the basis of our judgments about the

world. Because the signal is noisy, there is no perfect correlation between first-order judgments and the world itself. Judgments are based on a variable criterion. If the signal strength, whether due to noise or the relevant state of the world, reaches that criterion a judgment is made. Vickers and Lee's adaptive accumulator also models the collection of evidence over time and confidence in first-order judgments. Their model of confidence is a "balance of evidence" account, on which the evidence for each of two options (SDT models are typically models of binary choice) is compared and confidence is based on the relative strength of the evidence in favor of the chosen option. Note that on this model confidence appears to be responsive to the same evidence as the first-order judgment—a different operation is simply performed on that evidence. Vickers and Lee do not explicitly interpret their model in term of mental states and their intentional objects, but the most obvious interpretation appears to cut against evaluativism. Evaluativism claims that confidence and other epistemic feelings take as their inputs dynamic or other properties of the first-order cognitive *process*. But the model takes these inputs to be the same as inputs to the first-order process. Proust largely avoids this issue by concentrating on a further function of the model – the calibration of confidence ratings based on past accuracy (e.g., p. 99). But if confidence itself is a metacognitive epistemic feeling, then this calibration function of the accumulator is a *third*-order process, distinct from any heuristic input to the feeling of confidence itself. If Proust identifies fluency with the operation of Vickers and Lee's adaptive accumulator, the obvious interpretation of that model cuts against evaluativism.

More recent models appear to do a better job than Vickers and Lee's of predicting various phenomena related to confidence, but they also seem to defy an evaluativist interpretation. Instead of a balance of evidence model of confidence, Pleskac and colleagues (Pleskac & Busemeyer, 2010; Yu et al., 2015) posit that evidence continues to collect after the first-order judgment is made. Confidence will reflect the amount of evidence accumulated in favor of the first-order judgment in the interval between that judgment and the confidence rating. This improved model of confidence is no more amenable to evaluativist interpretation, as again confidence ratings are posited as responsive to the same evidence as first-order performance. Ratcliff and Starns (2013) offer a model of confidence on which "decision processes transform the strength of the match between a test item and memory to a confidence judgment (p. 5)." The model explains confidence as the result of direct access to items in memory ("matching" of the memory and the cue). Grimaldi et al. (2015) model confidence judgments in Bayesian terms as the width of the posterior probability distribution of evidence for a given decision (p. 13). In each case first-order and metacognitive responses take the same evidence as inputs, they simply assess different properties or time-slices of that evidence. None of these models posit properties of the first-order process itself as the source of metacognitive responses.

If fluency is the operation of an adaptive accumulator, core evaluativism is ill-conceived, as the heuristic is not distinguished, even in principle, from direct access to first-order content.

4.4.2 Fluency as reaction time

But there may be another reading of Proust's account of fluency. She sometimes describes fluency as reaction time:

> Comparative fluency is the property, for a stimulus, of being processed more or less quickly and adequately, with respect to what is expected, in a kind of task (or in a control loop). (p. 127)
> A temporal lag presents the subject with an error-signal: as compared with a normal behaviour, present activity is impaired. Here is the essential point: although the delay is a natural consequence of task difficulty, it becomes in addition a natural signal carrying information about a need to know what the affordance is. A plausible hypothesis therefore is that a temporal comparison between expected time for completion of the task and observed time, occurring as part of a given controlled activity (i.e. including a comparator), offers a key to making an affordance salient to the animal. (p. 136)

Reaction time (RT) has been posited as a heuristic behind several epistemic feelings, including JOLs (Koriat & Ma'ayan, 2005) and confidence (Proust, 2013; Ratcliff, 1978; Volkman, 1934). It is less obvious how reaction time could inform TOTs and FOKs, since in these cases an answer is not yet retrieved. Unlike the other heuristics we have considered, however, reaction time does not appear to be subject to the criticism that it reduces to a direct access view. That is, if one's feeling of confidence in an answer is based on how quickly one retrieved it, this would be distinct from the first-order content. Instead it is a property of the cognitive process itself, and this sounds like evaluativism proper. The question is whether it's true that epistemic feelings are caused by reaction time. Here I will focus on confidence, since this is the subject of Proust's discussion.

Simple RT-based models of confidence exist which resemble Proust's description (Ratcliff, 1978; Volkman, 1934). On these accounts a longer RT means the process is not going well and should induce low confidence. The negative correlation does hold when there are no time limits on the response. But it has long been known that when responses are demanded quickly, this correlation no longer holds (Irwin et al., 1956). It seems unlikely, then, that RT is the only input to feelings of confidence. More recent models account for this by basing confidence ratings on the quality of the evidence, represented by "drift rate" – the speed with which the evidence approaches the decision criterion. When time is not limited, a slow response time is the result of a slow drift rate toward the first-order decision criterion. A model like Pleskac and Busemeyer's (2010) predicts low confidence in this case, as evidence will continue to collect slowly after the first-order decision. But when response time is set by the researcher, it is no longer a function of drift rate and the negative correlation with confidence no longer holds.[13]

[13] The model does not posit that subjects access drift rate directly, which would implicitly involve a measure of RT. It posits different response criteria for different confidence levels. Higher quality evidence will tend to hit a higher confidence criterion in a given time interval.

Other researchers claim that reaction time does have some causal influence on confidence, as studies suggest that manipulation of reaction time affects confidence when other variables are held constant (Kiani et al., 2014). These researchers conclude that reaction time may be a relevant factor *in addition to* first-order evidence in the production of feelings of confidence. But this is not support for core evaluativism, which denies direct access to first order content altogether.[14]

5 Conclusion

Evaluativists posit a distinct, primitive metacognitive process that produces epistemic feelings based on heuristic inputs rather than first-order content. I have argued that this is unmotivated, as there are viable transparency, direct-access, and theory-theoretical explanations of phenomena like the Lima problem. I then reviewed the empirical evidence that epistemic feelings take heuristics as their inputs. I concluded that the measures used in these studies do not distinguish heuristics from first-order content. I then argued that, as conceived, the heuristics *just are* first-order content. I conclude that the evaluativist account of epistemic feelings is unmotivated, unsupported, and ill-conceived. My argument thus far, then, has been a negative one. In what follows I will briefly discuss one aspect of Proust and Arango-Munoz's versions of evalautivism that remains viable, significant, and could form the basis of further research.

Part of the fascination of epistemic feelings, which gets lost when we focus on their functional architecture, is the notion that metacognition can occur in a non-propositional format. It seems to refute the familiar "thought about thought" definition of the phenomenon. But there is an implicit assumption in evaluativism that because nonconceptual content is understood as phylogenetically primitive it must also be informationally impoverished. Proust conceives the content as expressing something like "poor (excellent) *A*-ing affordance" where *A* is some cognitive ability (2013, p. 121). This conception of the content of epistemic feelings might be necessary if we insist that they operate in a system that can only process simple heuristics. But nonconceptual content, as standardly conceived, carries *more* information than conceptual content (e.g., Dretske, 1981). Cussins' (1992) account of nonconceptual content, on which Proust's is largely based, posits a nested series of behavioral dispositions that allows for a great deal of informational complexity and versatility (see Grush, 2000 for a similar account of spatial content). And crucially, Cussins holds that nonconceptual content exists on a continuum with conceptual content and that the relationship between the two is fluid, concepts frequently becoming unstable, devolving back into a nonconceptual format and refashioning into new concepts. If we abandon the dual-mechanism account, epistemic feelings might be conceived as one manifestation of a kind of nonconceptual metacognition that is deeply intertwined with and supports propositional metacognition. There may be other such

[14] I thank an anonymous reviewer for pointing this out.

manifestations. I propose that our ability to distinguish visual perception from visual imagery is one (Greely, 2021).

Arango-Munoz (2011) explicitly considers the possibility that conceptual metacognition is grounded in nonconceptual metacognition and that the two coexist in a single mechanism (p. 78). He dismisses this account partially on the basis of the functional claims I dispute, adding the point that a one-mechanism account predicts "parallelism between judgments concerning the self and others" (p. 78). The argument is telegraphic, but the latter point appears to rely on the assumption that the one-mechanism account would be a form of simulation theory (Goldman, 2006). The Cussins-style account I have sketched is not, and it may help Arango-Munoz with another problem. Evidence that epistemic feelings are susceptible to conceptual priming drives him to claim that some epistemic feelings are conceptual (2014), sacrificing that distinction between epistemic feelings and other forms of metacognition. If the conceptual and nonconceptual are continuous and part of the same system, then one needn't posit two kinds of epistemic feelings to explain the data. That epistemic feelings are a form of nonconceptual metacognition constitutes a significant claim, even if there is no evidence that they reveal the existence of a distinct mechanism. Evidence for nonconceptual metacognition could come from studies of animal cognition, but as I have briefly noted, the methods are controversial, and more work must be done. The situation is trickier in human subjects if we assume that conceptual and nonconceptual metacognition often intermingle. I have argued that the measures employed in studies of epistemic feelings are flawed. A different theoretical perspective may improve our methods.

Acknowledgements Thanks to Matthew Fulkerson, Eric Schwitzgebel, David Barner, Jonathan Cohen, Rick Grush, audiences at the 2020 meeting of the Southern Society for Philosophy and Psychology, and several anonymous referees for their comments on earlier versions of this article.

References

Arango-Munoz, S. (2011). Two levels of metacognition. *Philosophia, 39,* 71–82.

Arango-Munoz, S. (2013). Scaffolded memory and metacognitive feelings. *Review of Philosophy and Psychology, 4,* 135–152.

Arango-Munoz, S. (2014b). The nature of epistemic feelings. *Philosophical Psychology, 27*(2), 193–211.

Arango-Munoz, S. (2014a). Metacognitive feelings, self-ascriptions and mental actions. *Philisophical Inquiries, 2*(1), 145–160.

Arango-Munoz, S. (2019). Cognitive phenomenology and metacognitive feelings. *Mind and Language, 34,* 247–262.

Benjamin, A. S., Bjork, R. A., & Schwarz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55–68.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*(2), 204–223.

Brown, R., & McNeill, D. (1966). The"tip of the tongue"phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337.

Byrne, A. (2005). Introspection. *Philosophical Topics, 33*(1), 79–104.

Byrne, A. (2008). Knowing that I am thinking. In A. E. Hatzimoysis (Ed.), *Self-Knowledge* (pp. 105–124). Oxford University Press.

Carruthers, P. (2011). *The opacity of mind*. Oxford University Press.

Carruthers, P. (2017). Are epistemic emotions metacognitive? *Philosophical Psychology, 30,* 58–78.
Cussins, A. (1992). Content, embodiment and objectivity: The theory of cognitive trails. *Mind, 101,* 651–688.
De Sousa, R. (2009). Epistemic feelings. *Mind and matter, 7*(2), 139–161.
Dellantonio, S., & Pastore, L. (2019). How can you be sure? Epistemic feelings as a monitoring system for cognitive contents. In A. Nepomuceno Fernández, L. Magnani, F. J. Salguero-Lamillar, C. Barés-Gómez, & M. Fontaine (Eds.), *Model-based reasoning in science and technology* (pp. 407–426). New York: Springer.
Dokic, J. (2012). Seeds of self-knowledge: Noetic feelings and metacognition. In M. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *The foundations of metacognition* (pp. 302–321). Oxford University Press.
Dretske, F. (1981). *Knowledge and the flow of information*. MIT Press.
Duke, D., Fiacconi, C. M., & Kohler, S. (2014). Parallel effects of processing fluency and positive affect on familiarity-based recognition decisions for faces. *Frontiers in Psychology, 5*(328), 1–11.
Emery, N., Dally, J., & Clayton, S. (2004). Western scrub-jays (Aphelocoma californica) use cognitive strategies to protect their caches from thieving conspecifics. *Animal Cognition, 7,* 37–43.
Evans, J. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454–459.
Evans, J., & Stanovich, K. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241.Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112,* 309–346.
Finn, B., & Tauber, S. K. (2015). When confidence is not a signal of knowing: How students' experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educational Psychology Review, 27,* 567–586.
Fodor, J. (1998). *Concepts: Where cognitive science went wrong*. Oxford University Press.
Goldman, A. (2006). *Simulating minds*. Oxford University Press.
Gopnik, A. (1993). How we know our own minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16,* 1–14.
Graf, L., Mayer, S., & Landwehr, J. (2017). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology, 28*(3), 393–411.
Greely, N. (2021). *Nonconceptual metacognition.* Manuscript in preparation.
Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience and Biobehavioral Reviews, 55,* 88–97.
Grush, R. (2000). Self, world and space: The meaning and mechanisms of ego- and allocentric spatial representation. *Brain and Mind, 1,* 59–92.
Hanczakowski, M., Pasek, T., Zawadzka, K., & Mazzoni, G. (2013). Cue familiarity and 'don't know' responding in episodic memory tasks. *Journal of Memory and Language, 69,* 368–383.
Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*(4), 208–216.
Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition, 38,* 771–784.
Hertzog, C., Fulton, E., Sinclair, S., & Dunlosky, J. (2014). Recalled aspects of original encoding strategies influence episodic feelings of knowing. *Memory and Cognition, 42,* 126–140.
Hosey, L. A., Peynircioglu, Z. F., & Rabinovitz, B. E. (2009). Feeling of knowing for names in response to faces. *Acta Psychologica, 130*(3), 214–224.
Irwin, F. W., Smith, W. A. S., & Mayfield, J. F. (1956). Tests of two theories of decision in an "expanded judgment" situation. *Journal of Experimental Psychology, 51,* 261–268.
Isingrini, M., Sacher, M., Perrotin, A., Taconnat, L., Souchay, C., Stoehr, H., & Bouazzaoui, B. (2016). Episodic feeling-of-knowing relies on noncriterial recollection and familiarity. *Consciousness and Cognition, 41,* 31–40.
Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of a stimulus. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(1), 21–38.
Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica, 98,* 127–140.
Kiani, R., Corthell, L., & Shadlen, M. (2014). Choice certainty is informed by both evidence and decision time. *Neuron, 84*(6), 1329–1342.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*(4), 609–639.
Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349–370.
Koriat, A. (2000). The feeling of knowing: some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9,* 149–171.
Koriat, A. (2006). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *Cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press.
Koriat, A., & Ackerman, R. (2010). Metacognition and mindreading: Judgments of learning for self and other during self-paced study. *Consciousness and Cognition, 19,* 251–264.
Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187–194.
Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 34–53.
Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (p. 117–135). Hove: Psychology Press.
KoriatMa'ayan, A. H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52*(4), 478–492.
Liu, Y., Su, Y., Xu, G., & Chan, R. (2007). Two dissociable aspects of feeling-of-knowing: Knowing that you know and knowing that you do not know. *The Quarterly Journal of Experimental Psychology, 60*(5), 672–680.
Locke, J. (1689). *An essay concerning human understanding*. UK: Oxford University Press.
Lycan, W. (1996). *Consciousness and experience*. MIT Press.
Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1084–1097.
Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Philosophy: Learning, Memory, and Cognition, 19*(4), 851–861.
Michaelian, K. (2016). *Mental time travel*. MIT Press.
Millikan, R. (1987). *Language, thought, and other biological categories*. MIT Press.
Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation, 26,* 125–173.
Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments fore predicting perceptual identification and learning. *Journal of Experimental Psychology: General, 113*(2), 282–300.
PaynterRederKieffaber, C. A. L. M. P. D. (2009). Knowing before we know: ERP correlates of initial feeling-of-knowing. *Neuropsychologica, 47*(3), 796–803.
Perrin, D., Michaelian, K., & Sant'Anna, A. (2020). The phenomenology of remembering is an epistemic feeling. *Frontiers in Psychology, 11*(1531), 1–14.
Pitt, D. (2004). The phenomenology of cognition or what is it like to think that p? *Philosophy and Phenomenological Research, 69*(1), 1–36.
Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117*(3), 864–901.
Proust, J. (2013). *The philosophy of metacognition*. Oxford University Press.
Ratcliff, R. (1978). Theory of memory retrieval. *Psychological Review, 85,* 59–108.
Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review, 120*(3), 697–719.
Reber, R., Fazendeiro, T. A., & Winkielman, P. (2002). Processing fluency as the source of experiences at the fringe of consciousness. *Psyche, 8*(10), 175–188.
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,* 435–451.
Schwartz, B. L. (2010). The effects of emotion on tip-of-the-tongue states. *Psychonomic Bulletin & Review, 17,* 82–87.
Schwartz, B. L., & Metcalfe, J. (2010). Tip-of-the-tongue (TOT) states: Retrieval, experience, and behavior. *Memory and Cognition, 39*(5), 737–749.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality & Social Psychology, 61,* 195–202.

Siedlecka, M., Paulewicz, B., & Wierzchon, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology, 7*(218), 1–8.

Strawson, P. F. (1959). *Individuals*. Methuen.

Thomas, A., & K., Bulevich, J. B., & Dubois, S. J. . (2012). An analysis of the determinants of feelings of knowing. *Consciousness and Cognition, 21,* 1681–1694.

Vickers, D., & Lee, M. D. (1998). Dynamic models of simple judgments I Properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology, and Life Sciences, 2*(3), 169–193.

Volkman, J. (1934). The relation of the time of judgment to the certainty of judgment. *Psychological Bulletin, 31,* 672–673.

Walsh, M. M., & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive Psychology, 58*(3), 416–440.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica, 98,* 141–165.

Whittlesea, B. W. A., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: The heuristic basis of feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 3–33.

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Earlbaum.

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). *Journal of Experimental Psychology, 144*(2), 489–510.