Replies to Langland-Hassan, Nagel, and Smith

JOËLLE PROUST
Institut Jean-Nicod

Feeling, predictive control, and self-awareness

I am grateful to Peter Langland-Hassan, Jennifer Nagel, and Barry Smith for their reactions to my book: whether critical or sympathetic, their observations were stimulating and immensely useful. Peter Langland-Hassan and Jennifer Nagel both invite me to attempt to pry apart terminological from empirical results in comparative metacognition. The first section of my comments will address issues related to this question. First of all, does self-evaluation, when performed by experimental animals, really require metacognition? In other terms: is procedural metacognition not merely "cognitive"? Is not the possible role of feelings in adaptive decisions to opt out from difficult tasks purely epiphenomenal? Second, why should cognitive control have anything to do "with an organism's awareness or understanding of mentality", as Langland-Hassan asks? This question is echoed by Smith's discussion of the role of recursion in human thought. Third, ask Smith and Nagel, what exactly is the relation between feelings and judgments in human procedural metacognition? Can feelings replace judgments, or do they need to be coupled with judgments to guide epistemic decision, which would make every kind of metacognition analytic? Fourth, Smith asks, is metacognition one kind of process, or just a diverse set of regulatory systems? How do our olfactory feelings of familiarity differ from our olfactory feelings of knowing? And what are these feelings actually indicating?

I. Three questions about the very existence and nature of procedural metacognition in nonhumans

Does animals' success in self-evaluating their perceptual or memorial abilities involve metacognition? A common reason for posing this question has been that the methodological soundness of the experimental apparatus used to collect evidence about animal metacognition has seemed questionable from the start. Trial-by-trial reward was initially provided to the subjects; their answers could thus depend directly on response feedback and conditioning to a type of stimulus, rather than on any endogenous cue such as relative uncertainty about what they perceive, as a step guiding an epistemic response. From 2006 on, however, feedback was delivered after a block of trials, and hence, could no longer be associated with a given stimulus. Furthermore, the ability to opt out was shown to immediately transfer to new stimuli and new tasks. Part of the community considered that the evidence collected thereafter in favor of a nonhuman ability to assess uncertainty was sound and valid. The debate moved to the best way of characterizing this unexpected animal capacity.

Part of the community, however, remained skeptical about the existence of such a capacity. Assessing task difficulty might depend on more subtle environmental cues, rather than on endogenously directed attention. If environmental cues turn out to be sufficient to decide when a task is difficult, then animals would not need to attend to their uncertainty to decide how to act.[1] This is the substance of "Perner's challenge", which Langland-Hassan

[1] Perner (2012).

endorses, and which also worries Smith. Another way of expressing the challenge, Perner claims, is that if animals are conditioned to use their own uncertainty state to select a response, they do not need to *know that* they are in that state. These are, however, two different challenges: The first is an empirical question that should be addressed. The second is framed from the viewpoint of the proponent of a different definition of what constitutes metacognition, and hence, is of no concern to the proceduralist.

I.I. First response to Perner's challenge: Causally indexical notions are not environmental

Are categorizations, such as "this room is hot" and "this person is attractive", driven by external properties of the environment? Philosophers of mind and epistemologists have offered a negative response: Temperature and attractiveness are properties that need to be assessed against an individual's own standards. The terms used to refer to them have an element of indexicality: this room is hot *for me*, this person is attractive *to me*. They instantiate what John Campbell has called "causally indexical" representations.[2] Indexicality is still more evident when one is representing a task as "difficult": this representation is causally indexical because, first, it emanates from an individual subject with given resources, needs and motivations (nothing is "intrinsically" difficult or easy), and second, because it has immediate implications for the subject's action. When assessed to be "too difficult", a task will be rejected. When found easy, it will be accepted.

Is there not, however, an objective counterpart of "difficult" that might explain the subject's behavior non-representationally? Causally indexical notions, Campbell observes, may indeed ultimately depend on non-indexical physical facts. "Too hot for me" may "relate to the causal impact of the heat upon the subject", a relation that can be grasped in a third-person way. "Too difficult for me" may similarly relate to the objective probability of success in a given cognitive task given my objective competence. Subjects who find a task difficult, however, rarely if ever have access to these objective facts: they are rather using their causal-indexical representations without having the remotest clue about any of the nonindexical notions on which their assessment of difficulty might be based, and without having a concept of self. To conclude: difficulty is not a property of a task that can be perceived. It has to be evaluated against a standard. When the difficulty has to do with epistemic decision, metacognition is needed to appreciate it.

I.II. Second response: Epistemic assessment vs. S-R conditioning

In a confidence assessment, the environmental cues available consist of the task stimuli and the benefits and losses incurred. How, then, are we to pry apart an epistemic assessment based on a causally indexical notion -- a metacognitive decision -- from a response prompted by Pavlovian conditioning? Here is how: If an action is selected on the basis of a subject's assessment, it should be independent of reward. Subjects should make different assessments of confidence for one and the same reward schedule. Such adaptive flexibility is indeed a permanent finding in the comparative literature.[3]

Another objection from Nagel, echoing Nate Kornell's remarks, is that an agent's wavering between two courses of action, or reacting longer to difficult stimuli, might predict failure to get a reward. Granting that these cues are non-cognitive, they cannot generate metacognition. Let us observe first that pronouncing an indicator as "non-cognitive" on the basis of its vehicle is a dubious assumption. On a procedural view, a cue is metacognitive if and only if it carries information about a first-order cognitive task allowing the subject to

---

2 Campbell (1994), p. 43 ff.

3 See Couchman et al. (2012), Kepecs and Mainen (2012).

predict likely success or failure. Following Fred Dretske, an indicator is a representation if it has the function of indicating.[4] Hence, nonconceptual indicators can obviously serve a "cognitive" role. Second, terminological issues are conflated with theoretical stances, which obscures the main issue. The point is not to know who will win the naming game; it is rather, to account for the evidence. For an attributivist, a cue only qualifies as metacognitive if it is based on a metarepresentation of the first-order state. For example, Kornell[5] claims that a metamemorial assessment qualifies as metacognitive only if it is based on access to the contents of one's memory. Now experimental work, as he himself recognizes, clearly indicates that ease of processing, rather than a metarepresentation, is the basis of metamemory assessment in humans. Should one consider that an inference, because it does not involve a metarepresentation, does not qualify as "meta", or, rather, accept that the meaning of metacognition should be extended to accommodate these findings, as many theorists have done?

The behavioral objection, in addition, is neutralized by neuroscientific evidence showing that onset, intensity and coherence of the neural activity currently triggered by the first-order cognitive task together provide a neural signature that, when compared to a threshold indicating sufficient confidence, reliably guides a cognitive decision.[6] Behavioral cues, when they exist, are thus anticipated by fully "private cues". A detailed computational analysis of rats' olfactory decision-making[7] also provides evidence that the animals, like humans, are basing their decision in part on their confidence level, and not only on the expected reward. Correct and erroneous choices for the same stimulus, again, are found to systematically vary in their confidence level. Thus the dynamical neural signature described above has epistemic and not only instrumental significance. This model is consistent with Asher Koriat's theory of memory-monitoring in humans: confidence in one's memory "is based on structural, contentless cues", i.e., on internal consistency within a collection of representations (consistency being here understood as holding among various subpersonal cues, when they converge on a single response).[8]

An essential ingredient in a metacognitive decision, as observed by Langland-Hassan, is for the reliability of the epistemic decision to be secured on the basis of past feedback. Crucially, the threshold for sufficient confidence needs to be continuously recalibrated for the self-evaluative mechanism to reliably optimize success in cognitive decision.[9] Does this look like an "almost ordinary" achievement, like a Monet when looked at closely (as Kornell suggests)? For most Monet viewers, looking closely at the details makes the overall effect still more impressive.

Ignoring the calibration "metaloop"[10], as many theorists in the comparative literature tend to do, leaves unanswered the question of how epistemic reliability is secured. This metaloop tracks past efficient decision thresholds, and uses them as a cut-off point in future decisions. The functional role of calibration, however, makes S-R conditioning an inadequate account of procedural metacognition.

I.III. Do feelings play any role in "procedural metacognition"?

---

4 Dretske (1988).
5 See Kornell (2014a) and (2014b).
6 Kepecs & Mainen, (2012), Kiani & Shadlen (2009).
7 Kepecs & Mainen, (2012).
8 Koriat (2012), 227.
9 Baranski & Petrusic (1994).
10 See Vickers & Lee (1998).

Many theorists of animal uncertainty balk at using the term "metacognitive" for the predictive mechanisms they have found, on the ground that *an assessment of uncertainty fails to be explicit*: the animals may not "know that they are uncertain".[11] Although animals do not have a linguistic representation of their uncertainty, as Smith emphasizes, it seems natural to speculate that they entertain conscious feelings of uncertainty - a speculation that affective neuroscience will hopefully turn into a testable hypothesis. Nagel is skeptical about this speculation, on the basis of the following analogy: Is not the comparator in charge of assessing uncertainty closely similar to the mechanism determining the weightings of each sensory channel in processing multimodal perceptual input? In the latter case, variance in each input is the information that influences the weight to be given to the corresponding channel. In this way, the final estimate has lower variance than either the visual or the haptic estimator. Feelings of uncertainty, should they exist, would in this case play no role. Why should they do so, then, when animals decide to opt-out?

Let us take intermodal integration as our target case. There are indeed interesting commonalities between the two kinds of processes. First, variance is used as a predictor in a comparator-based Bayesian system: the more noise in a channel, the more variance between neuronal responses, the less likely it is that the channel will perform a correct estimate and hence that a decision will be correct. Second, this process involves a dynamic cue combination rule: the relevant information belongs to the neural vehicle rather than to the content itself. Third, the thresholds are constantly recalibrated to optimize sensory integration or epistemic decision. Fourth, the weights that are finally attributed to each channel or possible decision result from the interaction of the neural populations for each modality: neural populations directly represent the probability distributions of target phenomena. These features indeed characterize a much wider class of functional phenomena called "prediction error minimization" (PEM).[12] From this theoretical viewpoint, almost all the brain functions as classically identified – perception, action, reasoning, planning, etc. – consist of top-down predictive assessments of the probability of errors and attempts to minimize them.

There are also important functional differences, however. Crossmodal integration is a highly modular phenomenon. In contrast, monitoring uncertainty about one's own cognitive performance extends to several types of inputs (perceptual, memorial, etc). Second, crossmodal integration involves a single perceptual system, whereas metacognition refers to processes "that make use of one or more … representations of a property of a cognitive process" in order to control it.[13] On this definition, defended in chapter 2, only some forms of cognitive control are metacognitive; for example, typing errors are detected by a mismatch in the forward model currently activated, not by another system. Still more importantly: metacognition, in contrast with crossmodal integration, issues in a decision to act cognitively (or to refrain from acting). This epistemic decision in turn engages a behavior that may directly affect survival. This embedded structure of metacognition might explain, then, why noetic feelings only occur in perception when agents are to act or could be in a position to act on their assessment (for example, to move in order to improve their percept). Why are feelings functionally relevant in this case? For two reasons: they can prompt swift action, while also being able to merge with other feelings for a more balanced action.

Nagel expresses doubts that what I call procedural metacognition might contribute to the fitness of nonhuman organisms. It might, however, respond to evolutionary pressures associated with competition for food: Carruthers & Ritchie (2012) hypothesize that a specific sensitivity to what one can discriminate or remember might be related to "the extent to which anxiety is created in foraging situations". Being highly competitive, semi-nomadic macaque

---

11 See for example, Kepecs & Mainen (2012), Kiani & Shadlen (2009), Kornell (2014a ; 2014b).

12 See Clark (2013), Friston (2009, Hohwy (2013).

13 For example, the variance of the firing rate of a population of neurons. Shea et al. (2014), p.2.

monkeys seem to be eager to predict what they will find. In contrast, arboreal capuchin monkeys are not able to opt out adaptively in the lab, presumably because they have sharing food habits and live locally in environments with ample amounts of a variety of food.

II. Cognitive control and awareness of mentality

Langland-Hassan holds the view that metacognition should involve metarepresentations, not only as a terminological position, but as a substantial theoretical claim, extending to procedural metacognition. He thus joins the theorists cited above in considering that when animals present the same sensitivity as humans to what they see or remember, they are blindly relying on subpersonal associations, whereas humans follow the same heuristics (like ease of processing) in a conscious way. Being aware of one's own uncertainty, he claims, involves explicit identification of the corresponding mental state. Metacognitive tasks are "highly relevant" to elicit a conceptual characterization of one's own mental states. A simple epistemic rule could be applied (by humans), in order to "easily and securely generate true beliefs about their own current mental states". They might, for example, use a rule such as: when trying to decide whether one believes *p* with uncertainty *u*, "look outward" at task difficulty. An advantage of this solution, inspired by Evans' notion of ascent routine, and by Alex Byrne's "epistemic rules",[14] is, for Langland-Hassan, that procedural and analytic metacognition would be integrated in human minds. Procedural assessment of difficulty (in Perner's sense) would be first-order, while analytic metacognition would construct and select the relevant rule of inference, such as "if *p*, believe that you believe that *p*". As I understand this proposal, it is only from the viewpoint of a mature human conceptual system, then, that one can reinterpret first-order judgments of difficulty as "relevantly" metacognitive.

The ascent routine element in the proposal is compatible with the view of human cognitive architecture defended in this book. On my view, humans, and possibly other vertebrates, have two different types of representational systems, one encompassing non-conceptual, emotional features with valence and intensity and direct motivations to act, the other comprising conceptual, inferential and planning capacities with indirect motivations to act. Among the diverse concept-based systems existing in phylogeny, analytic metacognition seems to be uniquely human; one of its functions is to refer to the outputs of the procedural system, in order to exploit them conceptually. Mature human minds, having the concepts of belief and other mental states, can thus re-describe in propositional terms their having a given noetic feeling. Byrne's epistemic rules, which Langland-Hassan is right to present as a sophisticated version of the ascent routine approach, are meant to characterize how agents endowed with mental concepts can gain self-knowledge (i.e. a second-order true justified belief) from merely perceiving *p*. By parity with the case of perception, an epistemic rule for "believing that one is confident in one's ability to *r*", adapted to my own framework, would have the form:

CONFIDENCE: if *A (q),* where *A* is an affordance for accepting a question *q*, believe that you confidently believe that you have the ability to respond to *q*.

The interest of having such a rule, as Langland-Hassan observes, is that it explains how thinkers endowed with the relevant conceptual repertoire can judge that they are currently confident about their ability to respond to *q*. What it fails to do, however, is to explain how first-order mental concepts, such as believing, and second order concepts, such as believing that one finds something uncertain, are *acquired*. Is mindreading or, rather, some specifically epistemic training, needed to offer access to these concepts? The proposal seems, rather, to aim to restrict the ascent routine to a purely shallow role, where no such concept is acquired,

[14] Byrne (2011, 2012).

based on the speculation that "self-awareness does not get much deeper". It is unclear how this proposal, granting the existence of a gap between procedural and analytic metacognition, can integrate them.

Another problem with this proposal, related to its purported capacity to integrate metacognition, is that a conceptual redescription of one's epistemic experiences does not exhaust what analytic metacognition does for us. Although we share with nonhumans a sensitivity to fluency, we also use norms that are inherently metarepresentational. For example, truth, accuracy, exhaustiveness, and plausibility are epistemic norms that require mastery of mental concepts, and their associated inferential abilities. A whole range of new mental actions opens up to systems sensitive to these analytic norms, including strategic choices between norms.

On the view defended in the book, then, the integration of procedural and analytic metacognition in humans is explained in the same way as the integration of the control structures in general, by a hierarchy of controls that have mainly a top-down effect, but also propagate error signals in a bottom-up way. Conceptual redescription certainly plays an important role in this integration, but it is only one of the relevant integrating factors.

Langland-Hassan's question of how cognitive control and self-awareness are related, however, remains a very hard nut to crack, in part because it cannot be addressed by purely conceptual methods. Granted that explicit self-knowledge is based in part on the conceptual capacity to attribute mental states to oneself, is there not room for other forms of self-awareness, that are not so based? As Smith suggests, should not some nonhuman animals and human infants be recognized as having a form of sensitivity to their current mental states? Shouldn't they even be granted a form of nonconceptual self-awareness? Neuroscientist Victor Lamme has put forth interesting hypotheses concerning this question. His claim is that, while deploying a feedforward model for controlling one's actions or thoughts is done unconsciously, the recurrent processing feedback from forward control, i.e., the monitoring part of the process, generates the core of phenomenal consciousness.[15] On Lamme's view, reportability and introspective access are not conditions for phenomenal consciousness, but are, rather, downstream effects of phenomenal consciousness on other mental processes (such as attention, working memory and language).

Reliable monitoring is certainly not sufficient for self-knowledge, but evidence from schizophrenic delusions also suggests that it may be an ingredient in nonconceptual self-awareness. As discussed in Chapter 12, some forms of recurrent processing –in particular the sense that one's body is moving, for example – are preserved in schizophrenic delusions of control. Others, however, having to do with the recognition of one's own actions, are impaired: patients fail to feel that they performed the corresponding action, or sometimes they even feel authorship for the actions of others. Do these phenomenological states result merely from a mistaken use of a conceptual self-attributive rule? Maybe, at the end of the day, self-attributive concepts will be found to play a role. A more proximal explanation, however, is that misguided error predictions, due to impaired control of one's action and one's cognition, are propagated upward, and create first an experience, and then an attribution of foreign influence on one's own agency.

Smith focuses on the role of metacognition in constituting persons, which is investigated in Chapter 11. He is puzzled by the fact that my view might entail that nonhumans, having a form of metacognition, might then have "a potentially available developing self". He observes that this consequence might be blocked by their lack of a "sense of a continuing project". But there is not just that, I would add: they also lack the kind of analytic metacognition that would

---

15 Lamme (2006).

be needed to elaborate such a project. I fully concur with Smith, then, "that we are not characterizing the same thing when appealing to person-constituting metacognitive processes and to task-related metacognitive processes that enable monkeys to act appropriately under conditions of uncertainty". The former processes involve analytic metacognition, the latter involve procedural metacognition. Why is analytic metacognition so central to selfhood? In chapter 11, I defend the view, inspired by Frankfurt, that self-affection is a necessary ingredient in the constitution of the self. By "self-affection", I understand the ability to take new mental states (desires, beliefs, intentions) and new properties (e.g. becoming more attentive to others, being more careful in examining evidence) as goals of one's mental actions, and of pursuing these goals until completion. [16] Self-affection, then, crucially involves "self-monitoring for the benefit of self-control", i.e., (analytic) metacognition. The overlapping memory episodes generated by self-affection provide a continuous link between memories of one's own life that can be exploited for re-identification of oneself. The Lockean simple memory theory of selfhood, in contrast, fails to build on self-affection, and hence, does not offer the opportunity for a self representation to be constructed and reidentified. Monkeys are what Frankfurt calls "wantons": they do not care about what they think and do from the viewpoint of long-term dispositions or plans, and they fail to organize their preferences by using temporally distant constraints and goals. From mere procedural metacognition no self-representation as humans understand the term emerges, because neither monkeys nor human "wantons" represent themselves as having mental states, long-term goals and values.[17] They can represent themselves as having the same body, however, and as occupying the same hierarchical rank in their group (or a different one than before), for this kind of self-awareness may depend on affects rather than metarepresentations.

III. About the relation between feelings and judgments in procedural metacognition

An important issue is whether noetic feelings are merely affective states, or involve judgments. Nagel's view is that "the special 'knowing how' that is enabled by metacognitive feelings always depends directly on 'knowing that'". In other terms, feelings of fluency, at least in humans, she claims, are formed in association with a conceptual judgment about the task, a view that I will henceforth refer to as "the doxastic view" about noetic feelings.

A brief digression will clarify the debate. As Nagel observes, the term "fluency" has unwanted polysemy. Fluency is, primarily, an objective epistemic norm: processing is objectively easy or difficult, in the sense that a processing sequence can be, ceteris paribus, shorter or longer, resource-demanding or not. Subjects are sensitive to this norm, by feeling whether a given mental action is easy or difficult for them to perform. A number of mechanisms serve the function of selecting fluent processes. Hence, the term "fluency" is also applied to the cues that track the norm ("fluency heuristics"), and to the neural processes that implement them. Finally, fluency is applied, by extension, to the accessibility of conceptual contents ("conceptual fluency"). Sensitivity to fluency is a procedural norm that is first deployed at a subpersonal level. According to Koriat's "cross-over principle", subpersonal heuristics (mechanisms detecting comparative fluency) can generate conscious "noetic" feelings, which in turn directly motivate and guide epistemic decisions.[18]

Nagel's first argument in favor of a doxastic view of feeling-based epistemic decision involves observation of the contrasting roles of feelings in decision. Feelings can be treated by a cognitive agent as "diagnostic" (i.e., as having a predictive value for the cognitive outcome), or as "incidental" to it (i.e., having no predictive value). Schwarz et al.'s

---

16 Proust (2013), 233.

17 See Proust (2013), 233.

18 Koriat (2000).

'assertiveness' experiment is spectacular: when subjects had to retrieve either six or twelve examples of situations in which they behaved assertively, their own sense of assertiveness was *inversely* correlated with the number of examples of assertiveness they retrieved, as a consequence of the higher effort expanded for twelve examples as compared to six. In a second test, however, subjects completed the same task while listening to music, which was presented as hindering or aiding recall. Given these instructions, processing effort no longer influenced their self-evaluation as being or not being assertive. This is taken to show that "a theory of feelings" (or "a theory of the task") is engaged in the control of epistemic decision.

This general conclusion, appearances notwithstanding, is not justified by the evidence. Three types of cases must be distinguished. Case 1: Subjects, when confronted with a task, bring their conceptual knowledge to bear on the various dimensions of a situation. Having conceptually constructed a task does not prevent them from responding to it in a purely emotional way (experienced chess players have extensive knowledge about the game, but still often make moves based on their present noetic feelings, with no explicit justification). Their "knowing how", then, is independent of their "knowing that". Case 2: Subjects may, in a particular case, be explicitly instructed not to trust their feelings, as in the second part of the assertiveness experiment. Subjects then develop what Rolf Reber calls "critical thinking about feeling".[19] In other terms, they actively prevent their feelings from influencing their decision. This strategic ability may not be available to most nonhumans, or to human children. In adult humans, however, as a function of circumstances (urgency, division of attention), control can be switched (back and forth) between impulsive, emotion-based, and concept-based action.[20] Unkelbach's experiment[21], used by Nagel as an additional argument in favor of her doxastic view of noetic feelings, involves a third case. This time, participants are not explicitly instructed to discard felt fluency as diagnostic of truth: they implicitly learn over a succession of trials that the high fluency experienced in reading a statement predicts falsity. Note that this is a non-ecological experiment. Fluent processing, in our world, reliably predicts truth.[22] This case does not constitute an argument for the doxastic view, however. Granting that the influence of ease of processing is calibrated over time by task success in a given domain, then it should fall into disuse if it is no longer predictive of success in this domain. In this case, the control of decision is transferred to system 2 processing.

Agents, then, can momentarily resist their impulse to act emotionally, or be led to use conceptual control, when circumstances require it. In the default mode, however, noetic feelings exert direct, unopposed control over a subject's decisions. Evidence for this claim is provided by the following experiment.

Nussinson & Koriat[23] exposed participants to unsolved and solved anagrams. They asked them to rate the difficulty of these anagrams for naive participants with no prior access to the solution. The participants' ratings were influenced by the differential fluency that the anagrams had for themselves: the higher fluency of solved anagrams biased their attributions of difficulty. This is the default mode, in which feelings influence decision. After being informed of the contaminating effect of knowing the solutions, the participants were able to strategically control their feelings of fluency: they re-rated correctly the difficulty of the anagrams for uninformed participants. Here is the crucial test, however: if the feelings were cognitively penetrable, i.e., if participants were able to use them on the basis of what they have learned, the feelings should no longer influence decision in the same way. However, when the subjects were re-tested subsequently, under time pressure, they again presented the

---

19 Reber (2013).

20 Who, then, operates the switch? See Proust (forthcoming a).

21 Unkelbach (2007).

22 Reber & Unkelbach (2010).

23 Nussinson & Koriat (2008).

initial bias for known anagrams. This suggests that receiving verbal instruction could shift participants' controlled responses, but did not lead the participants to re-compute them, as should have been the case if the feelings were cognitively penetrable. Control of noetic feelings through a "theory of the task", then, does not typically result in their suppression or reinterpretation.

A second challenging argument by Nagel in favor of the doxastic view is based on the notion that fluency has an underlying unity "as a single factor guiding judgment" across a variety of contexts. Mechanisms of fluency have been shown to influence liking, truth judgments, predictive and retrospective perception and memory assessment, familiarity, and judgments of learning. This suggests that "the meaning of fluency is learned, rather than fixed… The guidance provided by such a feeling is not hardwired into the nature of the feeling, but depends on acquired beliefs."

Note that, for epistemic agents, a feeling of familiarity and a feeling of beauty are qualitatively different. The view that the feelings listed above have a generic core, which belongs to biology, and a specific quality that depends on interpretation, is a prima facie strange hypothesis. It is based on a two-factor theory of feelings, first applied to affects by Schacter & Singer in their famous adrenaline experiment, and then to noetic feelings, by Kelley & Jacoby: arousal is generic, whereas valence is specific and based on an interpretation of the generic feeling.

In spite of its widespread influence, I think that a two-factor theory fails to overcome several objections. 1) It does not account for the anagram case described above. 2) It is incompatible with monkeys' and young children's epistemic evaluations based on fluency. Claiming that monkeys and humans use their noetic feelings radically differently is an unlikely hypothesis, given what we know about the evolution of brain structures. 3) The evidence for a concept-based decision system overruling affective guidance, however, is overwhelming. The crucial role of concepts, in this case, consists in inhibiting the impulsive emotional response, and in providing additional constraints to guide action. 4) Postulating that the mechanisms of fluency that intervene in all the cases listed above are similar is an approximation. A crucial element for justifying the objection would consist in demonstrating that one and the same subpersonal heuristic is used across the cases listed. Clearly, experiences of predictive and retrospective fluency cannot rely on the very same heuristics, because they tap different sequences of the action.[24]

In summary: Engaging in a particular cognitive task (e.g., trying to remember, evaluating retrieval, assessing frequency) does not need, per se, to involve a naïve theory of the task. It only requires having an implicit heuristic for metacognitive predictions in that task. Critical thinking about feeling, however, enables humans to adopt a critical view of their own feeling-based dispositions to act. A theory of the task, then, becomes an indispensable means of inhibiting these dispositions.

IV. Generality of metacognition as an explanatory requirement. The case of olfaction.

Barry Smith raises the question of "whether the notion of metacognition being called on to do explanatory work in all these cases [metaperception, metamemory, judgments of learning, etc.] is just one thing or many" (p. 3). Is this alternative grounded? Mental functions, such as perception or belief fixation, are usually seen as being unified in spite of being subserved by multiple processes operating at various time scales of information uptake and processing. It is quite acceptable, however, to claim that these various mental mechanisms collaborate toward one and the same global functional role. I have proposed to

[24] See, for example, the various time heuristics used in problem solving in Ackerman (2014).

limit metacognition to those processes whose function is to evaluate, predictively or in retrospect, the validity of informational decisions, i.e. to regulate mental agency. This definition excludes regulatory systems controlling and monitoring bodily actions, which others have included in metacognition..[25] It is adopted for two reasons. First, action regulation does not generally aim to satisfy conditions of informational validity, but rather instrumental conditions of efficiency and goal completion. Second, regulation of informational dispositions (i.e. mental action) is far more recent, in phylogeny, than the regulation of bodily action.

Granting that positing a metacognitive function along these lines is justified, then, the question still arises of whether the function so defined is fulfilled by one central mechanism, or by distinct mechanisms that are distributed across the various informational channels that are being monitored. Furthermore, granting that several mechanisms are used to regulate mental agency depending on the control domain, a second question is whether the metacognitive experience gained in one domain is transferable to others. Concerning the first question, there is no evidence, at present, for a global metacognitive ability. Subjects tested on four different metacognitive tasks – ease of learning judgments, feeling of knowing judgments, judgments of learning, and text comprehension – turn out to present no intra-individual stable performance level of metacognitive accuracy across these tasks.[26] One possible explanation for this negative result, however, would be that a high number of trials would be necessary to obtain stable measurements, as is the case in Psychophysics. An alternative, more plausible account is that, granting the multidimensionality of metacognitive tasks, there is no common "metacognitive ability" elicited by these various tasks. What the neuroscience of animal metacognition suggests is that different brain systems subserve prospective and retrospective evaluations, namely the medial and the lateral prefrontal cortices.[27] Furthermore, the left intraparietal cortex is activated in evaluating visual decision in monkeys, and the orbitofrontal cortex is used by rodents to evaluate olfactory decision. The brain correlates of human metacognition are not known in detail, but we know that they differ according to the metacognitive assignment and its first-order input.[28]

The lack of evidence for a global metacognitive capacity, then, supports the "multiple mechanisms" view. However, this does not threaten the functional character of metacognition. A perceiver can have good vision and poor audition. Similarly, an epistemic agent might be good in assessing her visual object recognition, and poor in assessing her ability to detect visual changes. Granting, in addition, that human metacognition can rely both on feelings and on concept-based judgments, variation in human responses is predictable.

It is not surprising that, in contrast with monkeys, there is at present no evidence of transfer in metacognitive tasks in the human case. Humans, however, have never been tested, so far as I know, on a perception-to-memory transfer of metacognitive ability. Should they ever be, an additional problem is that, in contrast with experimentally trained animals, whose environment and task are tightly controlled, humans have been exposed to a variety of environments and assignments, some of which included metaperceptual and metamemorial evaluations. Hence, transfer cannot be tested in adult humans (nor easily in children, given the training involved).

In summary, my answer to Smith's "generality of account" objection is that metacognition has not yet been found to be a global capacity associated with a stable ability in self-evaluating one's own cognitive performance, in whatever task and situation, and might

---

25 Metcalfe & Greene (2007).

26 For these results, along with a discussion of the methodological problems involved in these measures, see Kelemen et al. (2000).

27 Fleming & Dolan (2012).

28 The only common intra-individual feature across tasks that has been found in humans is overconfidence or underconfidence, which might reflect stable personal traits. See Kelemen et al. (2000).

never be. This does not constitute, however, an objection to the claim that metacognition constitutes a mental function, allowing experimental monkeys to win their daily food, wild monkeys to reduce their uncertainty about foraging opportunities, and humans to be sensitive to many epistemic norms vital for thinking, conversing, and cooperating.

Do humans have olfactory metacognition? Procedural (experience-based) metacognition has developed in areas that are evolutionarily relevant in a given species: for monkeys, visual perception, for rodents, olfactory perception, for both, memory for food locations. As emphasized by Smith, humans are endowed with language, which (among other architectural changes) "analytically" extends the scope of their metacognitive evaluations. They are able to evaluate their first-order epistemic decisions on the basis of conscious recognition that they have certain mental states and goals. Evaluations now track not only fluency, but norms such as truth, evidentiality, coherence, plausibility, consensus that have been culturally acquired, over the centuries, through collective activities such as legal, medical, political, philosophical and scientific reasoning. Sensitivity to these additional norms has considerably augmented the more primitive forms of social reasoning, by predicting, for example, what others will do, given their partial ignorance of the relevant facts, or the prevalent consensus relative to them. In addition, analytic metacognition, in contrast with procedural metacognition, is domain-general, insofar as concepts can, in principle, apply to any first-order task that is accessible to consciousness. Analytic metacognition, then, constitutes the epistemic underpinning of human reason, which would not have arisen without a communicational propositional language.

It would be simplistic, however, to speculate that higher forms of metacognition are exclusively based on concepts. Even higher levels of problem solving, like mathematical reasoning and philosophical or scientific thinking, have to use noetic feelings to estimate the relevance of arguments. On my view, a noetic feeling is a feeling combining sensitivity to fluency and to informativeness.[29] This form of representational thought, as discussed in chapter 6,[30] is non-objective, non-propositional, gradient-based and nonconceptual. Similarly, as Smith notes, procedural metacognition contributes to affectively simulating the uncertainty and noetic experiences of others, which plays a role in reading the minds of others. The hypothesis that it might also allow us to notice the affective impairments of persons with autism deserves scientific attention. In fact, much more remains to be discovered about the relations between the two types of evaluation, (conceptual and emotional) within higher forms of thought and communication

These remarks are relevant to the case of olfactory metacognition mentioned by Smith. Do humans have a procedural capacity for evaluating their uncertainty about olfaction, and if they do, are the metacognitive processes engaged similar to those at work in monitoring and controlling cognitive tasks? Humans differ greatly in their olfactory discriminative abilities, and also, probably, in the corresponding metacognitive abilities, depending on their profession, sociocultural background and other factors. As Smith observes, we routinely experience feelings of olfactory familiarity, when we smell an odor that we have frequently encountered, but have not yet recognized what its source is, for lack of visual access or other environmental cues. Such a sense of familiarity is an error signal: the information it carries is, roughly, that although the stimulus is not a new one, no identification of its source is presently available, but might be if more effort was expended. A priori, two accounts of olfactory familiarity are possible. On one, our feeling of familiarity indicates that the odor categorization is more fluent than expected (and hence, that the odor is not new), where what is categorized is an olfactory pattern, not the object causing it. If, alternatively, our feeling of familiarity is based on a heuristic associating visual and olfactory processing cues, then what is fluently categorized is an object with a given smell property. To know which account is

---

[29] Proust (submitted).

[30] See also Proust (forthcoming a; forthcoming b)

correct, let us look at our olfactory capacity to identify the smell-producing object: humans are not, in a majority of cases, able to identify the object from the odor alone. Research suggests that it is not the link between object and name that is hard to retrieve; it is, rather, the link between an odor and an object. Hence, familiarity, i.e. recognition of a smell as already encountered, seems to be based on the olfactory pattern alone.[31]

An intense tip of the nose experience (TON) (a rare event) does not predict, as does the common tip of the tongue experience (TOT), that the corresponding word will imminently be available. Notice that this nice expression of TON is nevertheless a misnomer: in contrast with the tongue, the nose does not carry information through its muscular activity. The supposed "tip of the nose experience" should rather be called "a feeling of being able to name the source of a smell". This feeling, like the feeling of being able to *perceptually* identify its source, is, in humans, utterly unreliable.[32] Even smell detection, in humans, is rather poor: blind smell, i.e. failure to detect a smell that is subpersonally processed, seems to be a common phenomenon.[33]

Smith asks 1) whether the same system is responsible for recognizing a smell as familiar and for assessing one's likelihood of identifying it, and 2) whether or not the kind of metacognitive processes at work in olfaction are similar to those used in cognitive tasks. In reply to the first question, a feeling of familiarity differs from a feeling of knowing (FOK) what the source is, insofar as a FOK is jointly based on cue familiarity (necessary to activate search) and on cue accessibility, namely the amount of information activated in association with the target.[34] In the olfactory cases, subjects' feelings are less predictive than in other perceptual cases. They have higher FOKs, however, when they are subsequently able to identify their source than when they are not, which suggests that their feelings are modulated by the amount of information available.[35] These two kinds of feelings, then, the feeling of familiarity and the FOK, do not differ in the systems generating them. They are, rather, two successive steps in the recall process.

Coming to the second question: olfactory evaluation aims at predicting one's own dispositions to recognize, discriminate and identify the source of a smell. It thus belongs to metacognition. Noticing a new smell, a metaperceptual achievement, may be as important as identifying its source, a metamemory achievement: the cues and heuristics involved are the same as those used in predicting epistemic decisions in other domains of perception and memory. Analytic metacognition is similarly to be blamed for the wrong convictions people have about being good at naming odors and being excellent change detectors. One limitation of the experimental results on matters of olfaction, however, consists in the fact that professional noses (wine tasters, perfume creators, cooks) have not yet been tested on their metamemorial performances, nor have people from other cultures where olfaction matters in their daily life.

Acknowledgements

I am grateful to Brie Gertler and to Dick Carter for their linguistic revision. This research has been supported by ERC Senior Grant 269616, by ANR-11-LABX-0087 IEC and ANR-11-IDEX-0001-02 PSL.

References

31 Jönsson (2005).
32 Jönsson (2005).
33 Sobel et (1999).
34 Koriat & Levy-Sadot (2001).
35 Jönsson et al. (2005)

Ackerman, R. 2014. The diminishing criterion model for time investment in problem solving. *Journal of Experimental Psychology: General.* 143(3): 1349-1368.
Baranski, J.V. & Petrusic, W.M. 1994. The calibration and resolution of confidence in perceptual judgments. *Perception and Psychophysics,* 55: 412-28.
Byrne, A. 2011. Knowing that I am thinking. in A. Hatzimoysis (ed.). *Self-Knowledge*, Oxford, Oxford University Press, 105-124.
Byrne, A. 2012. Knowing what I see. In D. Smithies and D. Soljar (eds.). *Introspection and Consciousness*, Oxford, Oxford University Press, 183-210.
Campbell, J. 1994. *Past, Space, and Self*. Cambridge, Mass: MIT Press.
Carruthers, P. & Ritchie, J.B. 2012. The emergence of metacognition: affect and uncertainty in animals. In: Beran, M.J., Brandl, J. Perner, J., Proust, J. (eds), *The Foundations of Metacognition,* Oxford: Oxford University Press, 76-93.
Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181-204.
Couchman, J.J., Beran, M.J., Coutinho, M.V.C., Boomer, J. & Smith, J.D. (2012). Evidence for animal metaminds. In Beran, M.J., Brandl, J., Perner, J. & Proust, J. (eds.), *The Foundations of Metacognition*, Oxford: Oxford University Press, 21-35.
Dretske, F. 1988. *Explaining Behavior, Reasons in a World of Causes,* Cambridge, MA: MIT Press.
Evans, G. 1982. *The Varieties of Reference,* Oxford: Clarendon Press.
Fleming, S. M., & Dolan, R. J. 2012. The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1338-1349.
Friston, K. 2009. The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences*, *13*(7), 293-301.
Hohwy, J. 2013. *The Predictive Mind.* Oxford, Oxford University Press.
Jönnson, F. U. 2005. Olfactory metacognition: A metamemory perspective on odor naming. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 6. Publications.uu.se.
Jönsson, F. U., Tchekhova, A., Lönner, P., & Olsson, M. J. 2005. A metamemory perspective on odor naming and identification. *Chemical Senses*, *30* (4), 353-365.
Kelemen, W.L., Frost, P. J. & Weaver, C.A. 2000. Individual differences in metacognition: evidence against a general metacognitive ability. *Memory & Cognition*, 28 (1), 92-107.
Kelley, C. M., & Jacoby, L. L. 1998. Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica*, *98*(2), 127-140.
Kepecs, A., & Mainen, Z. F. 2012. A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1322-1337.
Kiani, R. & Shadlen, M.N. 2009. Representation of Confidence associated with a decision by neurons in the parietal cortex. *Science:* 324 (5928, 759-764.
Koriat, A. & Levy-Sadot, R. 1999. Processes underlying Metacognitive Judgments: Information-based and experience-based monitoring of one's own Knowledge. In: S. Chaïken & Y. Trope, (Eds.), *Dual-Process Theories in Social Psychology*. London: The Guilford Press, 483-502.
Koriat, A. 2000. The Feeling of Knowing: some metatheoretical Implications for Consciousness and Control. *Consciousness and Cognition*, 9, 149-171.
Koriat, A. 2012. The subjective confidence in one's knowledge and judgments: some metatheoretical considerations, In: Beran, M.J., Brandl, J. Perner, J., Proust, J. (eds), *The Foundations of Metacognition,* Oxford: Oxford University Press, 213-233.
Kornell, N. 2014a. Where is the "meta" in animal metacognition? *Journal of Comparative*

Psychology, 128, 2: 143-149.
Kornell, N. 2014b. Author's Response. *Journal of Comparative Psychology*, 128, 2: 160-162.
Lamme, V. A. 2006. Towards a true neural stance on consciousness. *Trends in cognitive sciences*, *10*(11), 494-501.
Metcalfe, J., & Greene, M. J. 2007. Metacognition of agency. *Journal of Experimental Psychology: General*, *136*(2), 184-199.
Nussinson, R., & Koriat, A. (2008). Correcting experience-based judgments: the perseverance of subjective experience in the face of the correction of judgment. *Metacognition and Learning*, *3*(2), 159-174.
Perner, J. 2012. Minimeta: in search of minimal criteria for metacognition, in: Beran, M.J., Brandl, J., Perner, J., Proust, J. (eds.), *The Foundations of Metacognition,* Oxford: Oxford University Press, 94-116.
Proust, J. 2013. *The Philosophy of Metacognition*. Oxford: Oxford University Press.
Proust, J. Forthcoming a. Time and action: impulsivity, habit, strategy. *Review of Philosophy and Psychology*.
Proust, J. Forthcoming b. The representational structure of Feelings. On-line in *MIND Open*.
Proust, J. Submitted. Expressive representation and metacognition in primate communication.
Reber, R. 2013. Critical feeling. In C. Unkelbach & R. Greifeneder (Eds.), *The Experience of Thinking.* London & New York: Psychology Press: 173-189.
Reber, R., & Unkelbach, C. 2010. The epistemic status of processing fluency as source for judgments of truth. *Review of philosophy and psychology*, *1*(4), 563-581.
Schachter, S., & Singer, J. 1962. Cognitive, social, and physiological determinants of emotional state. *Psychological review*, *69*(5), 379.
Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in cognitive sciences*, *18*(4), 186-193.
Sobel, N., Prabhakaran, V., Hartley, C. A., Desmond, J. E., Glover, G. H., Sullivan, E. V., & Gabrieli, J. D. (1999). Blind smell: brain activation induced by an undetected air-borne chemical. *Brain*, *122*(2), 209-217.
Unkelbach, C. 2007. Reversing the truth effect: learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33* (1), 219.
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, 169-194.