
The Extension of the Indicator-Function of Feelings

A Commentary on Joëlle Proust

[Iuliia Pliushch](#)

In the following commentary I will first briefly review the target article, then voice some critical points, and last offer a positive proposal according to which tension in self-deception is a kind of a metacognitive feeling. Proust offers a novel, inspiring view that feelings possess an indexical (non-conceptual) format, are transparent (that is, they may be re-described in propositional terms, but not thereby changed), and acquire valence if the rate of change towards fulfilling the given affordance is greater or less than expected. In my critique I will first point to difficulties in disentangling feelings from emotions, then try to provide a more precise description of the formal object of feelings, along with some examples, and offer a definition of “directness” that is consistent with predictive coding—as well as argue that feelings might be influenced by concepts even if they themselves are non-conceptual. Last, I propose that tension in self-deception is a metacognitive feeling.

Keywords

Affective feelings | Appraisal | Metacognitive or noetic feelings | Predictive coding | Self-deception | Two-factor account

Commentator

[Iuliia Pliushch](#)

pliushi@students.uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Target Author

[Joëlle Proust](#)

joelle.proust@ehess.fr
Ecole Normale Supérieure
Paris, France

Editors

[Thomas Metzinger](#)

metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

[Jennifer M. Windt](#)

jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 The expressive mode of feelings

First, I would like to repeat, in short, the main claims of the target paper that will serve as a basis for my subsequent comments and extensions in the following sections. Joëlle Proust’s article is concerned with the functional and informational characterisation of feelings. She argues that the concept of “feeling” consists of the following components:

1. Reactive (associated with appraisal)
2. Subjective experience

3. With distinctive embodied phenomenal quality (somatic markers have the function of expressing intensity and valence of feelings, [Proust this collection](#), p. 8)
4. Possessing a formal object (not always, e.g., feeling depressed; absence of a formal object is typical of moods, footnote 5)

The formal object of feelings is argued to be affordance-sensing, a “non-conceptual and entirely subjective appraisal of the environment by the

agent” (Proust [this collection](#), p. 7) or a “*subjective, embodied relation* to an opportunity in an input from the environment” (p. 8). Assuming the non-referential indexicality of feelings, or that feelings signal a *relational* affordance (p. 7) that depends on the representational system (p. 8), Proust argues that feelings can misrepresent only if they are re-described in propositional terms. She argues that feelings are *transparent*, because of the experienced connection between their somatic markers and affordances, as well as because of the direct influence of their valence and intensity on an agent’s motivation (p. 12). Though subjects feel directly, in order to report their feelings they have to “translate one mode of representation into another, with no guarantee that this translation will not enrich or modify FS intentional content” (p. 12). Subjects might reinterpret and mis-describe their feelings, but they cannot thereby change the nature of those feelings (feelings being *cognitively impenetrable*; p. 19).

Feelings are argued to be a plausible candidate for a natural kind on the basis of the comparison between feelings and emotions—which she considers not to constitute a natural kind (Proust [this collection](#), p. 3). Two kinds of subjective appraisal might be part of an emotion: *primary feelings* on the one hand and *appraisals cum conative dispositions* on the other. While the first kind corresponds to an earlier time in our evolutionary development, is independent of concepts, induces specific responses, and possesses distinct somatic markers, the second kind is not and might be a blend of different instances of the first kind. Apart from primary affective feelings, somatic, agentive, and metacognitive feelings are argued to form a natural kind.

The function of feelings is to non-conceptually evaluate and signal the result of a comparison process between prediction and outcome through embodied experience (Proust [this collection](#), p. 4). Due to their non-conceptual monitoring nature, feelings do not convey, but merely approximate a causal relation between internal states, external states, and actions (p. 11). There are three kinds of functional relations between feelings and actions (pp. 4–5):

1. Determination of a kind of action in response: approach vs. avoidance
2. Specific orientation in time: predictive vs. retrospective
3. Level of arousal: elevation in energy vs. soothing effect

Feelings are argued to be the result of a *comparator* or control mechanism that is *automatic* and *encapsulated*. The latter requirements are imposed in order to explain the independence of feelings of beliefs and intentions (p. 5) such that, e.g., one could still feel the adrenalin rush even though the hypothesized venomous snake turned out to be a twig.

Metacognitive feelings (M-feelings) are held to express informational, instead of environmental affordances, arise in mental acts, and trigger similar actions of approach or avoidance. M-feelings involve appraisal of the properties of the informational processes underlying contents of thought, but not those content themselves. Against Schachter & Singer’s (1962) two-factor theory of emotions (interpreted as feelings possessing intrinsic arousal but extrinsic valence), Proust argues that feelings have *intrinsic intensity and valence*. Cues on which those feelings are based can be conveyed verbally though, and thus, the heuristics (implicitly or explicitly) might change in the long run. The main claim is thus that *context-dependency is not concept-dependency* (Proust [this collection](#), p. 17). Experience of tokens of feelings differs with respect to the kind of affordance they express (several affordances might be linked to the same type of feeling) and actions they trigger.

An especially interesting claim for me is that affective feeling in general, and metacognitive feelings in particular, have a meta-monitoring function of signalling “the rate of reduction of discrepancy toward a confidence threshold” (Proust [this collection](#), p. 21). If the rate of discrepancy reduction is above expected, the valence of a feeling is experienced as more positive, and, if below expected, as more negative. “Cold” feelings without valence are those for which the expectation has been correct. This claim is interesting for two reasons. On the one hand, to the reader familiar with the self-decep-

tion literature the key-concept “confidence threshold” will stand out. It plays an important role in accounts of self-deception that regard it as a kind of hypothesis testing (one prominent proponent of this view is [Mele 2012](#)). In short, according to this type of account, gathering of evidence in favour of a certain hypothesis is pursued up to a certain point: up until the amount of evidence has reached a confidence threshold that is enough to push an acceptance or rejection of the hypothesis (for more see [Pliushch & Metzinger 2015](#)). On the other hand, “prediction error”, or difference between prediction and sensory input, is the key-term in the model of mental representation that has lately gained a large amount of acceptance—predictive coding (for a short introduction to the free-energy principle of which predictive coding is a particular implementation see [Friston 2009](#); see also [Clark, Hohwy, Seth this collection](#)). Predictive coding provides a unifying explanation for perception, cognition, and action as a result of hierarchical Bayesian inference: at different levels, predictions are compared to propagated precision-weighted prediction error that, under different conditions, leads either to changes in the model of causes of sensory input or to action directed at testing the current model ([Clark 2013](#)).

The idea that feelings signal the rate of reduction of prediction error might be worth elaborating in the predictive coding framework, particularly given the recent study by [Furl et al. \(2010\)](#) who argue that facial expressions are represented as anticipated *trajectories* of the change of those expressions: pictures of neutral and fearful faces were morphed to different degrees such that participants got to see trajectories from a neutral to a fearful face and vice versa. After seeing such a sequences of pictures, participants had to rate another picture for fearfulness. The results indicated that predictable sequences in which the degree of being morphed rose or fell monotonously, thus forming a trajectory, biased perception ([Furl et al. 2010](#), p. 696). Combining Proust’s idea with the results of Furl et al.’s study: feelings might also be represented as anticipated trajectories of change, particularly given the possibly bi-direc-

tional causal influence between feelings and facial expressions (see section 2.2).

2 Critique: Affect and implicit heuristics in feelings

2.1 Use of the term “affect”

The aim of this section is threefold: 1) show difficulties in disentangling feelings from emotions; 2) attempt to give a more precise characterisation of the formal object of feelings, along with some examples; 3) criticize the use of the term “direct” and offer another definition that is consistent with predictive coding. The first problematic point that I see is Proust’s use of the term “affective”, which is ambiguous. She employs at least two different definitions of “affective”:

1. Feelings that possess valence (p. 20). Yet *all* kinds of feelings, according to Proust, possess affect *and* valence¹ (p. 1). Given her distinction between “hot” (emotional) feelings and those that have valence² (p. 21), emotional feelings might differ from mere feelings with valence due to the differently-experienced valence, maybe if emotional valence were a richer experience. Thus, the question is about the minimal requirements on valence and intensity in feelings.
2. Feelings that express emotions.
3. *Difference between feelings and emotions*: if agentive and metacognitive feelings can be affective, then the categorization of feelings into bodily, agentive, metacognitive, *and* affective (p. 5) might be better restricted to the first three, with the fourth being a dimension along which they vary. If affective in this categorization means emotional (p. 2), then there is an ambiguity of terms—affect-

1 The following quotations might help to elucidate the matter: “[f]eelings typically express affect and valence in sensation (25-26), all the feelings vary in affect in roughly the same way, because they all include valence in their informational structure” (p. 20).

2 In Proust’s words, the difference between “hot” feelings and feelings with valence, on the example of M-feelings, is that “although all M-feelings do not often have a definite ‘hot’ quality comparable with fear and love, they always have a valence, according to whether they predict the agent’s progress toward or away from her cognitive goal” (p. 21).

ive = having valence and affective = being part of an emotional experience—because the latter seems to be more complex.

4. *Difference between formal objects of feelings and emotions*: if “feelings are affective ingredients in emotional awareness” (p. 3), then there is a circularity in understanding affectivity here: feelings are affective in virtue of being part of an emotion, while at the same time they themselves are the affective component in the emotion of which they are part. The first part of this claim can be followed from that defended by Proust, namely that feelings that do not express emotions are not affective (p. 2). The second part of the claim follows from Proust’s claim that feelings are affective ingredients of emotions (p. 3). As elaborated in the previous section, emotions are said by Proust to contain one of two kinds of subjective appraisals: feelings or appraisals cum conative dispositions. Further, if feelings are components of emotions, but both can have a formal object, then those objects might diverge. The consequence is that an emotion and a feeling that is part of it might be directed at different objects. Thus, Proust on the one hand distinguishes feelings from emotions and yet on the other hand claims that not only emotional feelings, but also agentive and metacognitive feelings might be “feeling toward” experiences (p. 3, pp. 20–21). The latter claim that both feelings and emotions are directed at intentional objects has been used as an argument to identify both (see [de Sousa 2014](#) section 2 for a discussion of this question). Given Proust’s claim that there are somatic, affective, agentive, and metacognitive feelings, and given the claim that at least in metacognitive feelings the formal object is not the cognitive disposition itself but the rate of change of its execution above or below discrepancy, an interesting question focuses on the formal object of emotional feelings.³ For example, can it be

that while the formal object of the emotion of fear is some dangerous object, the object of a feeling is a rate of change in the assessment of the situation before and after the change of the formal object of an emotion? This might explain why, e.g., the first bite of a bar of chocolate makes one happier than the following bites.

5. *Bodily phenomenology of feelings as their formal object*: Proust argues that while somatic feelings are about bodily sensations (or, more consistently, about the rate of their change), in affective (emotional) and possibly metacognitive feelings “the bodily phenomenology tends to recede to the fringe of consciousness” ([this collection](#), p. 2). The example that Proust gives with respect to metacognitive feelings is that feelings of remembering are correlated with but not about facial muscle activity (p. 3). Proust acknowledges that there might be mixed cases (experience of bodily feeling + intentional content, pp. 2–3), but I want to argue that in some emotional feelings bodily phenomenology is, to borrow a metaphor, in the *foreground*. There might be emotional feelings whose objects are bodily sensations, e.g., the anxiety that arises during a panic attack: when I concentrate on my accelerated heartbeat, then if I come to associate the heartbeat with some threatening aspects of a situation, such an experience might lead to anxiety, and thus the initial anxiety leads to even more anxiety, leading to a vicious cycle of panic (for a discussion of heartbeat perception in panic disorder see [Ehlers & Breuer 1996](#)). This might be a case of an emotion whose formal object is the rate of change of bodily sensations, or maybe a meta-feeling (for a discussion of meta-emotions see [Mendonça 2013](#)).

In the given panic example it might have seemed as if I had embraced the analogy between feelings and perception that Proust

positive or negative (*ibid.*, p. 7). Positivity and negativity are dimensions along which valence changes, and valence has been characterised as the rate of change of discrepancy towards the (cognitive) goal. For more on why the latter characterisation is interesting see section 3.

³ To be more precise, the question is about the functional description of the formal object of feelings. Proust ([this collection](#)) says that “[f]eelings express [...] affordance as their focus (for formal object), along with its graded valence, ranging from very unpleasant to very pleasant, and with its intensity gradient, which ranges from small to large” (p. 8). Affordance is defined as “perceived utility”, and can be

denies, so I will explain why it may be more beneficial to use the term “direct” in another sense to that used by Proust. Proust makes a sharp distinction between feeling and perceiving: “[w]hile percepts allow recognition and identification of external objects and properties, feelings express specific affordances in a perceived, imagined, or remembered situation” (this collection, p. 10). Non-conceptual parts of perceptions are said to relate to “objective, external contrastive cues” (Proust this collection, p. 10), while in feelings they relate to evaluative states. Perception is said to involve “direct sensory access to the world” (p. 10), while the access of feelings to the world and the body is claimed to be indirect. Proust’s evidence for a disanalogy between feeling and perception is based on the neuroscientific research of Barrett & Bar, who say that absence of “internal affective context” impairs the categorization of objects (2009, p. 12).⁴ Their evidence for this hypothesis is based on reviewing the anatomic connections involved in affective processing and that of object perception. One critique of this might be that the time of activation of certain regions responsible for emotional processing and perception might justify the claim that emotional processing comes before perception, but not how direct such processing is. Moreover, in light of predictive coding, perception, emotion, and cognition might all be indirect (Hohwy 2014; for more technical elaboration Friston et al. 2014). In other words, predictive coding provides the term “direct” with a meaning other than that used by Proust. In predictive coding directness is an absence of the evidentiary boundary, where the evidentiary boundary is the inferential isolation between the model of the world and the hidden causes of sensory input (Hohwy 2014). This means that causes beyond the boundary have to be inferred on the basis of independent evidence (ibid., p. 6), or, in Hohwy’s words, “[t]he brain doing the inference is secluded at least in the sense that certain

kinds of doubt about the occurrence of the evidence are unanswerable without further, independent evidence” (p. 7). Relating this observation to Proust, on the premise of accepting predictive coding, there might not be a sharp distinction between feeling and perceiving such as Proust postulates, or at least not in the form presented in the target article. If interoception as perception involves inferences about circumstances *beyond* the (same) evidentiary boundary, as suggested by Hohwy (2014), then feeling and perceiving would both be indirect (to the same degree).⁵ If interoception does not go beyond the evidentiary boundary, feelings might be direct, even if perception is not.

2.2 Concept-based feelings?

In this part of the review I will point out the dangers of interpreting the relation between feelings and concepts too simplistically and argue that it is possible that at least some kinds of feelings are influenced by concepts, even if they themselves are non-conceptual. Proust argues that for metacognitive feelings to arise an important affordance, as well as an implicit heuristic, has to be present (this collection, p. 18). This heuristic is based on cues about the dynamics of information processing, but not its contents (p. 15). The dichotomies that Proust uses in the description—implicit–explicit, unconscious–conscious, evolutionarily-old–evolutionarily-new, associative–rule-based (pp. 3-4, p. 17)—have often been mapped onto two different kinds of processes in dual processing theory (e.g., Frankish & Evans 2009). Dual processing theory states that there are two kinds of processing that possess the dichotomous characteristics mentioned above. A minimal description provided by Evans (2009) for type 1 is “fast, automatic, high processing capacity, low effort”, and for type 2 “slow, controlled, limited capacity, high effort” (p. 33). Along these lines, “implicit”, “unconscious”, “evolutionarily old”, “associative” have been also used as descriptors for type 1 and “explicit”, “conscious”, “evolutionar-

⁴ Barrett & Bar (2009) define affect as an influence on bodily states that is either unconscious or, if conscious, experienced as pleasurable or unpleasurable to varying degrees (pp. 1327–1328). Barrett & Bar’s (2009) basic claim is that the orbitofrontal cortex (OFC) integrates into a unified multimodal representation sensory information from both world and body in a dynamic way.

⁵ One could also ask whether the *same* evidentiary boundaries would be involved in feeling and perceiving, since there could be many of them (Friston 2013).

ily new”, “rule-based” as descriptors for type 2. A belief bias (accepting more believable than unbelievable conclusions) might serve as an example for type 1 (*ibid.*, p. 41), and the conscious correction thereof for type 2. The worry I have is adding to those dichotomies another one: non-conceptual (meaning in this case non-propositional; Proust *this collection*, p. 7)–conceptual (propositional, belief-like). Proust holds that “cues (associative heuristics) dictate how an affordance is detected, assessed and exploited in a context, but these cues are not consciously available, and hence do not depend on a naïve theory of the task” (p. 17). This inference is not valid in the given form. I agree with Proust that “[a] cue-based, non-analytic heuristic is not inferential in the interpretive, first-person sense” (p. 17), but I hold that there is at least one step to consider in between non-conceptual⁶ affordances and consciously evaluated affordances. And this is automatic *concept-based* activation (the existence of automatic appraisal is acknowledged by Proust; *footnote 7*).

Evans (2009) distinguishes between different kinds of dual processing theories, among which are the sequential (first automatic processing, then controlled) and the parallel theory. Proust seems to embrace a sequential kind of dual processing theory, given the functional role she ascribes to metacognitive feelings (evaluation of mental actions before and after their execution; Proust 2013). Yet how far implicit heuristics are independent of concepts is in question. Proust (*this collection*) denies that “a concept-based interpretation will affect the experienced feeling itself” (p. 17). As mentioned in section 1, she also denies that feelings have a conceptual format. Thus, she seems to deny both that concepts play a causal role in the emergence of feelings and that feelings themselves possess a conceptual format. I will briefly demonstrate that the term “implicit heuristic” does not preclude automatic concept activation, if it implies the activation of knowledge or goal representations. Thompson (2009) argues that

⁶ Among those who agree with Proust that the content of epistemic feelings is non-conceptual and non-metarepresentational are, for example, Michaelian & Arango-Muñoz (2014). But the *content* of a metacognitive feeling being non-conceptual does not preclude that concepts play a causal role in its emergence.

heuristic processes are contaminated by background knowledge, as well as by beliefs and expectations (p. 172, p. 174). Frankish (2009) notes that “the concepts of belief and desire correspond to the psychologist’s concepts of knowledge (or memory) and goal structure” (p. 91). Hence, activation of knowledge that may provide the context for feelings could also be conceptual. Goal representations might also be activated in the course of context creation, provided that unconscious goal pursuit is flexible and context-sensitive (Aarts & Custers 2012). Further, unconsciously activated goals not only depend on context, but also *create* context by influencing the accessibility of knowledge, evaluations, and emotions (Fishbach & Ferguson 2007, p. 496). It follows that if goal representations are activated, then they might lead to the activation of conceptual knowledge. Another interesting point is that if there is a continuous interplay between goal representations and affordances (opportunities in the environment; Huang & Bargh 2014, p. 125) and if goal representations can change the experience of the world (*ibid.*, p. 124), then goal representations might change sensing of affordances and, hence, the feelings associated with it. Further, there has been a proposal to distinguish between associative and rule-based processes by the kind of architecture they operate upon: namely connectionist vs. classical computational (for a short discussion see Samuels 2009, pp. 141–142). Thus, implicit heuristics might be understood as certain connected representations in a network being activated by some cues, where the question is about the representational format of such knowledge, or a more precise description of the relational nature of the feeling affordance. Last, a general note about the similarity between feelings and other kinds of representations: if Bliss-Moreau & Williams (2014) are correct in defending the claim that all kinds of representations possess an affective component (valence + arousal in their definition), then affect is something that expressive and conceptual representations share.

Of course, Proust’s claim that in the case of feelings those cues relate to the dynamics, but not to the contents of processes, indicates a

more specific understanding of the kind of implicit heuristic in question. My point, though, is that if humans can “enrich their noetic feelings through concepts, and thereby revise their reliance on fluency where it is not justified” (Proust 2013, p. 144), then in humans implicit heuristics may also be influenced by concepts (in an automatic way) and in such a way influence feelings. Needless to say, the independent existence of such a schema (be it cognitive or emotional) is hard to prove (Eysenck & Keane 2010, p. 597). According to Koriat & Levy-Sadot (1999), as cited by Proust (this collection, p. 15), metacognitive feelings arise as a result of nonanalytic inferential processes (described as the implicit or unconscious application of heuristics), in distinction to the direct memory trace hypothesis, according to which feelings have direct access to memory traces (Koriat & Levy-Sadot 1999, p. 487). Koriat & Levy-Sadot (1999) argue that the presence of dissociations between knowing and the feeling of knowing speaks against the second hypothesis. Even if heuristics in feelings are non-conceptual, the fact that through feelings emotion gets its valence necessitates that we consider how concepts and memory traces influence feelings, given that they play a role in emotions. Lane et al. (forthcoming), for example, argue that psychotherapeutic change is made possible by updating prior emotional experiences, for which memory traces of those experiences have to be reactivated and reconsolidated. Thus, even if feelings are non-propositional (Proust this collection, p. 20), activation of concepts and their expression in propositional terms are to be distinguished. The point is not that metacognitive feelings themselves cannot have indexical formats,⁷ or that an agent could not possess expressive and conceptual representations at the

same time, but that in humans the generation of (at least) metacognitive and emotional feelings might be preceded by an automatic concept activation that influences them. If this is the case, then one could ask again whether feelings are transparent (see section 1).

Further, instead of describing cognitive processes as serial, their *dynamic* (continuous) nature might be more worthy of emphasis. In the target article, Proust mentions that “[i]ncreased activity in the smile muscle, the zygomaticus major, produces feelings with a positive valence” (this collection, p. 15). This suggests that facial expression influences emotions. She also argues for the transparency (impenetrable nature) of feelings and the against two-factor theory, thus against the possibility that appraisal influences the valence of feelings (see section 1). I want to offer for clarification purposes a short review of the recent literature on which factors are supposed to influence feelings and factors feelings influence themselves. Rogers et al. (2014) emphasize the dynamic nature of emotions insofar as they depend on the social appraisal of a situation. Brosch (2013) also emphasizes the dynamic nature of appraisal that plays a causal role in eliciting emotions. The definition of appraisal that Brosch (2013) accepts also encompasses low-level appraisal based on learned schemata (p. 370). Brosch (2013) argues that first an initial low-level appraisal affects the physiology (1), action tendency (2), expression (3), and feeling (4) of an emotional experience, and then those changes in turn affect an on-going (low- and high-level) appraisal, establishing an appraisal loop. Here, the direction of influence is still in question, e.g., whether feelings influence expressions or the other way around. Laird & Lacasse (2014) defend the James–Lange theory of emotion, namely that facial expressions (e.g., BOTOX patients being less responsive to mild positive emotional stimuli; for the reference see *ibid.*, p. 29), expressive behaviour (e.g., romantic attraction as a result of shared, mutual gaze; *ibid.*, p. 29), and visceral responses that are interpreted according to situational cues (e.g., misattribution of emotion) are *causes* of emotions (for a critique of their evidence see Reisenzein &

⁷ A better understanding of the indexical mode of feelings might be provided by the following quotation: “Feelings can be seen as pre-specified states of a comparator, which predict ultimate success or failure in the actions that they monitor. Given that the information they carry is immediately used in controlling and monitoring current effort, it is misleading to present them as ‘reporting’ the epistemic properties of a mental state or referring to it (even *de re*). They are, rather, signals in a control mechanism, which work somewhat as traffic lights do: allowing traffic, stopping it, rechanneling it; no report or reference need be involved” (Proust 2013, p. 76). In another place Proust (2013) notes that feelings “do not properly ‘refer’, because they do not engage propositional thinking” (p. 77).

Stephan 2014). As such, they may influence emotional feelings too, which Proust acknowledges by pointing out the causal connection between measures in facial muscles and affective character of feelings (this collection, p. 25). Yet the direction of influence may also go the other way around (from feelings to facial expressions). Thus, the nature of feelings may also be dynamic, as are the nature of the underlying cognitive processes. Interestingly, Thagard & Schröder (2014) argue for a neurocomputational theory of emotions as semantic pointers (term introduced by Chris Eliasmith). They argue that physiological, appraisal, social, and psychological components of emotions can be integrated into one unified account: emotion tokens can possess both shallow and deep meanings. The compressed (shallow) form of emotions is reportable, while at the same time pointing to the uncompressed deep form that binds together situational, physiological, and appraisal components.

In the preceding paragraph I considered literature supporting the claim that feelings are embedded in *continuous* cognitive processes. The purpose of this was to show that how appraisal might influence feelings in some form is complex and might even be circular. In this paragraph I offer some additional evidence against a discontinuous interpretation of the connection between feelings and propositional descriptions thereof. The existence of *affective blindsight* (ability to discern emotional stimuli despite inability to consciously perceive them; Eysenck & Keane 2010, p. 581) would stand in line with the assumption that emotional and cognitive processing is based on different kinds of information. This is because affective blindsight demonstrates the dissociation between two different kinds of processing and, thus, a dissociation between the information needed for the one kind and for the other. Further, Scott et al.'s (2014) experiment demonstrating *blind insight* (accurate metacognitive accuracy in the absence of discriminative accuracy) on the one hand supports Proust's hypothesis that metacognition and first-order cognition are not based on the same kind of information, yet on the other it speaks against a serial interpretation

according to which feelings arise out of automatic processes and are then re-described in propositional terms and used in first-person inferential reasoning. Liu & Wang (2014), for example, argue that motivational intensity influences the effect of positive affect on cognitive control: low-approach motivated positive affect enhances cognitive flexibility and distractibility, while high-approach motivated positive affect (associated with goal pursuit) enhances cognitive stability. Thus, the role of feelings might be broader than just the indicators that may or may not be used in conscious reasoning.

3 Proposals: Tension in self-deception is a kind of metacognitive feeling

Proust (this collection, as well as 2013) argues that mental actions are preceded and followed by metacognitive feelings indicating the appropriateness of the cognitive process in question. I want to argue that tension in self-deception fits the characterisation of a metacognitive feeling. Tension is described as a feeling of uneasiness and distress, and as such I think that it is precisely this tension that is said to indicate to the self-deceiver that her belief-forming process is faulty.

Self-deception (SD) is a motivated (1) kind of typically subpersonal hypothesis-testing (2) that results in an evidence-incompatible mental representation of reality (3) which fulfils a belief-like role (4) (Pliushch & Metzinger 2015). Self-deception is usually discussed in the context of biased belief-forming processes and it is argued that phenomenological tension arises as a result of the execution of such processes (e.g., Lynch 2012). Thus, the same function has been ascribed to tension in self-deception as the one ascribed by Proust to metacognitive feelings, namely a comparison of the cognitive process to certain criteria. In self-deception, rationality criteria are typically emphasised.

I want to argue that metacognitive feelings apply to self-deception, insofar as they might also monitor *unconscious* cognitive processes and arise not only before or after a cognitive process, but also *during* it. In case of self-deception these cognitive processes are belief-forming

processes. Proust (this collection, 2013) considers conscious mental actions: her argument is that unconscious comparison processes that give rise to metacognitive feelings precede and follow conscious mental actions. She argues that the “attentional-supervisory system” emerges from “distributed metacognitive abilities” (Proust 2013, p. 263). Ignorance of epistemic norms such as relevance, coherence, fluency, and informativeness lead to (pathological) errors in belief acquisition (Proust 2013, pp. 260–261). My argument in favour of the extension of metacognitive feelings to monitor unconscious cognitive processes is of a phenomenological nature. I agree with Proust (this collection) that the term “inference” has been used loosely in the literature and does not always indicate a first-person inference (p. 21). Yet the more basic problem might be that there is no sequential first-person inference as such in the first place. If the shift between mind wandering (task-unrelated cognitive activity) and task-directed cognitive activity goes unnoticed (Metzinger 2013), then there might be other shifts that we do not notice, e.g., the shift from unconscious to conscious cognitive processes, or some changes in the given process. Thus, the phenomenology of a cognitive process might be more complicated than a unified sequence with a starting point and an end. Further, given, for example, mood-state dependent cognition (Eysenck & Keane 2010, pp. 584), I doubt the plausibility of the assumption that only in breaks between conscious cognitive processes do subjects experience affective feelings.

In the previous paragraphs I argued that the functional role of metacognitive feelings fits that of tension in self-deception, and that metacognitive feelings arise not only before and after mental actions, but also before, after, and during unconscious (possibly self-deceptive) cognitive processes. In this paragraph I want to link Proust’s idea that feelings possess valence only if the rate of change of progress is unexpected to predictive coding, in order to provide a functional description of metacognitive feelings. Proust (this collection) argues that the affective quality of feelings arises only if the cognitive process violates expectations: if it progresses

quicker towards the goal, positive feelings arise, if slower, negative feelings arise⁸ (p. 21). Given that the terms “expectation” and “prediction error” have gained popularity in virtue of being key terms in predictive coding, which is a modelling strategy explaining perception, cognition, and action (Clark 2013), I will shortly discuss Proust’s claim about affect in metacognitive feelings in the context of predictive coding. According to predictive coding, prediction errors (deviation between expectation and outcome) are precision-weighted. Precision is the property of prediction errors (errors between the top-down prediction and the bottom-up signal one receives) that can be described as the weight of a prediction error that plays the role of selection: the more precise the prediction error, the more it will change the hypothesis about causes of input. Switching between perception and action depends on the precision of prediction errors: precise prediction errors change hypotheses, while imprecise ones lead to action (Brown et al. 2013). Precision⁹ is also argued to play a dual biasing role: biasing perception toward goal states and enhancing confidence in action choices (Friston et al. 2013). Low precision of prediction errors has been argued to cause anxiety (Mathys et al. 2011, p. 17).¹⁰ I argue that Proust’s proposal that violations of expectations of “a given rate of reduction of the discrepancies toward her [agent’s] cognitive goal” (this collection, p. 26) produce affective feelings might be described in predictive coding terms as violations of *transition probabilities* of reaching the goal state:¹¹ if a state conducive to the goal state or a goal state itself has been reached, despite a low probability of changing

⁸ Note the analogy to the “dark room problem” in predictive coding: if an agent wants to minimize surprise or prediction error, then she should stay in a dark room, given that there will be no surprise in it (e.g., Clark 2013). If there were no prediction error, this would cause uncertainty (e.g., Friston et al. 2012). Proust’s argument is similar: if there were no violations of expectations, then metacognitive feelings would not have any valence, because they only have valence if the rate of change is quicker or slower than expected.

⁹ Attention is precision optimization according to predictive coding (Hohwy 2013).

¹⁰ Mathys et al. (2011) are also interesting for the given topic insofar as Proust argues that the heuristics upon which metacognitive feelings are based might be changed via associative learning; Mathys et al. (2011) provide a predictive coding model of reinforcement learning.

¹¹ For a predictive-coding model of a goal-directed action see Friston et al. (2013).

into that state from the current state, then positive affective feelings might arise.¹²

The first step in the categorisation of tension as a metacognitive feeling has been an extension of the application of metacognitive feelings to unconscious belief-forming processes. The second is to clarify the representational content of tension. To do the latter, it might be beneficial to consider which other kinds of metacognitive feelings arise out of belief-forming processes. Those are intuitivity, counter-intuitivity, and anxiety, if one classifies them according to the phenomenology and not according to the norm that they control. Intuitivity indicates the appropriateness of a given belief-forming process.¹³ The reason for the ascription of the given functional role to intuitivity is that intuitivity signals 1) a good fit with respect to the network of our explicit background beliefs and 2) a good fit with respect to our conscious and unconscious model of reality (Metzinger & Windt 2014). An appropriate belief-forming process provides a good fit with respect to 1) and highly likely also with respect to 2). I further argue that counter-intuitivity represents that a certain cognitive process violates the chosen criterion of appropriateness, but is neutral with respect to the system's goal representations, while tension or anxiety represents that the cognitive process violates at least some important goal representations. The reason for this distinction is to account for the effect of motivation on belief-forming processes.

Thus, if feelings accompany our belief-forming processes, then readers might have experienced some while reading this commentary: hence the title. To conclude, I think that Proust has offered interesting ideas on the nature of feelings that will greatly contribute to the clari-

fication of the matter: the indexical (affordance-sensing and non-conceptual) format of feelings, their transparency, the taxonomy of feelings into sensory, emotional, agentive, and epistemic, the predictive and retrospective function of feelings signalling the appropriateness of the cognitive process they monitor, and the degree of change of expectation as the origin of valence of feelings. In this review I have tried to extend Proust's account. To do this, I attempted to provide some conceptual clarifications on the distinction between feelings and emotions, the formal object of feelings, and the conceptual influences to which they might be subject. Last, I argued that tension in self-deception is a kind of metacognitive feeling.

Acknowledgements

I am very grateful to the Barbara-Wengeler foundation for generously supporting my PhD project and all the reviewers of this commentary for their helpful improvements.

¹² Emotional valence has been also argued to be modelled as the rate of change of free energy: Instead of estimating volatility or "slow and continuous changes in states of the world" the rate of change of free energy is argued to take that role of estimating (known) uncertainty (Joffily & Coricelli 2013, p. 1). Here Joffily & Coricelli (2013) accept Yu & Dayan's (2005) distinction between expected and unexpected uncertainty: Expected uncertainty is the one about known unreliability of predicting relationships *within* a context and unexpected uncertainty is the one about the appropriateness of the context itself such that when unexpected uncertainty is high, it is a signal that a *context switch* should be made.

¹³ For an elaboration on the phenomenal signature of knowing in intuitions of certainty, see Metzinger & Windt (2014).

References

- Aarts, H. & Custers, R. (2012). Unconscious goal pursuit: Nonconscious goal regulation and motivation. In R. M. Ryan (Ed.) *The Oxford handbook of human motivation* (pp. 232-247). Oxford, UK: Oxford University Press.
- Barrett, L. F. & Bar, M. (2009). See it with feeling: Affective predictions during object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1325-1334.
[10.1098/rstb.2008.0312](https://doi.org/10.1098/rstb.2008.0312)
- Bliss-Moreau, E. & Williams, L. A. (2014). Tag, you're it: Affect tagging promotes goal formation and selection. *Behavioral and Brain Sciences*, 37 (2), 138-139.
[10.1017/S0140525X13001969](https://doi.org/10.1017/S0140525X13001969)
- Brosch, T. (2013). Comment: On the role of appraisal processes in the construction of emotion. *Emotion Review*, 5 (4), 369-373. [10.1177/1754073913489752](https://doi.org/10.1177/1754073913489752)
- Brown, H., Adams, R. A., Parees, I., Edwards, M. & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14 (4), 411-427.
[10.1007/s10339-013-0571-3](https://doi.org/10.1007/s10339-013-0571-3)
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36 (3), 181-204.
[10.1017/S0140525X12000477](https://doi.org/10.1017/S0140525X12000477)
- Clark A. (2015). Embodied prediction. In T. Metzinger and J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- de Sousa, R. (2014). Emotion. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy (Spring 2014 Edition)*.
<http://plato.stanford.edu/archives/spr2014/entries/emotion/>
- Ehlers, A. & Breuer, P. (1996). How good are patients with panic disorder at perceiving their heartbeats? *Biological Psychology*, 42 (1-2), 165-182.
[10.1016/0301-0511\(95\)05153-8](https://doi.org/10.1016/0301-0511(95)05153-8)
- Evans, J. St. B. T. (2009). How many dual-process theories do we need? One, two, or many? In K. Frankish and J. St. B. T. Evans (Eds.) *In two minds: Dual processes and beyond* (pp. 33-54). Oxford, UK: Oxford University Press.
- Eysenck, M. & Keane, M. T. (2010). *Cognitive psychology: A student's handbook*. Hove, UK: Psychology Press.
- Fishbach, A. & Ferguson, M. (2007). The goal construct in social psychology. In A. W. Kruglanski and E. T. Higgins (Eds.) *Social psychology: Handbook of basic principles* (pp. 490-515). New York, NY: Guilford Press.
- Frankish, K. (2009). Systems and levels: Dual-system theories and the personal-subpersonal distinction. In K. Frankish and J. St. B. T. Evans (Eds.) *In two minds: Dual processes and beyond* (pp. 89-107). Oxford, UK: Oxford University Press.
- Frankish, K. & Evans, J. St. B. T. (2009). The duality of mind: An historical perspective. In K. Frankish and J. St. B. T. Evans (Eds.) *In two minds: Dual processes and beyond* (pp. 1-29). Oxford, UK: Oxford University Press.
- Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, 13 (7), 293-301. [10.1016/j.tics.2009.04.005](https://doi.org/10.1016/j.tics.2009.04.005)
- (2013). Life as we know it. *Journal of The Royal Society Interface*, 10 (86), 20130475.
[10.1098/rsif.2013.0475](https://doi.org/10.1098/rsif.2013.0475)
- Friston, K., Adams, R. A., Perrinet, L. & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, 3 (151).
[10.3389/fpsyg.2012.00151](https://doi.org/10.3389/fpsyg.2012.00151)
- Friston, K., Schwartenbeck, P., Fitz-Gerald, T., Moutoussis, M., Behrens, T. & Dolan, R. J. (2013). The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience*, 7 (598).
[10.3389/fnhum.2013.00598](https://doi.org/10.3389/fnhum.2013.00598)
- Friston, K., Sengupta, B. & Auletta, G. (2014). Cognitive dynamics: From attractors to active inference. *Proceedings of the IEEE*, 102 (4), 427-445.
[10.1109/JPROC.2014.2306251](https://doi.org/10.1109/JPROC.2014.2306251)
- Furl, N., van Rijsbergen, N.J., Kiebel, S. J., Friston, K. J., Treves, A. & Dolan, R. J. (2010). Modulation of perception and brain activity by predictable trajectories of facial expressions. *Cerebral Cortex*, 20 (3), 694-703.
[10.1093/cercor/bhp140](https://doi.org/10.1093/cercor/bhp140)
- Hohwy, J. (2013). *The predictive mind*. Oxford, UK: Oxford University Press.
- (2014). The self-evidencing brain. *Noûs*.
[10.1111/nous.12062](https://doi.org/10.1111/nous.12062)
- Hohwy, J. (2015). The neural organ explains the mind. In T. Metzinger and J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Huang, J. Y. & Bargh, J. A. (2014). The selfish goal: Autonomously operating motivational structures as the proximate cause of human judgment and behavior. *Behavioral and Brain Sciences*, 37 (2), 121-135.
[10.1017/S0140525X13000290](https://doi.org/10.1017/S0140525X13000290)
- Joffily, M. & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Computational Biology*, 9 (6), e1003094. [10.1371/journal.pcbi.1003094](https://doi.org/10.1371/journal.pcbi.1003094)

- Koriat, A. & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken and Y. Trope (Eds.) *Dual-process theories in social psychology* (pp. 483-502). London, UK: The Guilford Press.
- Laird, J. D. & Lacasse, K. (2014). Bodily influences on emotional feelings: Accumulating evidence and extensions of William James's theory of emotion. *Emotion Review*, 6 (1), 27-34. [10.1177/1754073913494899](https://doi.org/10.1177/1754073913494899)
- Lane, R. D., Ryan, L., Nadel, L. & Greenberg, L. (forthcoming). Memory reconsolidation, emotional arousal and the process of change in psychotherapy: New insights from brain science. *Behavioral and Brain Sciences*. [10.1017/S0140525X14000041](https://doi.org/10.1017/S0140525X14000041)
- Liu, Y. & Wang, Z. (2014). Positive affect and cognitive control: Approach-motivation intensity influences the balance between cognitive flexibility and stability. *Psychological Science*, 25 (5), 1116-1123. [10.1177/0956797614525213](https://doi.org/10.1177/0956797614525213)
- Lynch, K. (2012). On the "tension" inherent in self-deception. *Philosophical Psychology*, 25 (3), 433-450. [10.1080/09515089.2011.622364](https://doi.org/10.1080/09515089.2011.622364)
- Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, 5 (39). [10.3389/fnhum.2011.00039](https://doi.org/10.3389/fnhum.2011.00039)
- Mele, A. R. (2012). When are we self-deceived? *Human-a.Mente - Journal of Philosophical Studies*, 20, 1-15.
- Mendonça, D. (2013). Emotions about emotions. *Emotion Review*, 5 (4), 390-396. [10.1177/1754073913484373](https://doi.org/10.1177/1754073913484373)
- Metzinger, T. (2013). The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers in Psychology*, 4 (931). [10.3389/fpsyg.2013.00931](https://doi.org/10.3389/fpsyg.2013.00931)
- Metzinger, T. & Windt, J. (2014). Die Phänomenale Signatur des Wissens: Experimentelle Philosophie des Geistes mit oder ohne Intuitionen? In T. Grundmann, J. Horvath and J. Kipper (Eds.) *Die experimentelle Philosophie in der Diskussion* (pp. 279-321). Berlin, GER: Suhrkamp.
- Michaelian, K. & Arango-Muñoz, S. (2014). Epistemic feelings, epistemic emotions: Review and introduction to the focus section. *Philosophical Inquiries*, 2 (1), 97-122.
- Pliushch, I. & Metzinger, T. (2015). Self-deception and the dolphin model of cognition. In R. Gennaro (Ed.) *Disturbed consciousness*. Cambridge, MA: MIT Press.
- Proust, J. (2013). *Philosophy of metacognition: Mental agency and self-awareness*. Oxford, UK: Oxford University Press.
- Proust J. (2015). The representational structure of feelings. In T. Metzinger and J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Reisenzein, R. & Stephan, A. (2014). More on James and the physical basis of emotion. *Emotion Review*, 6 (1), 35-46. [10.1177/1754073913501395](https://doi.org/10.1177/1754073913501395)
- Rogers, K. B., Schröder, T. & Scheve, C. v. (2014). Dissecting the sociality of emotion: A multilevel approach. *Emotion Review*, 6 (2), 124-133. [10.1177/1754073913503383](https://doi.org/10.1177/1754073913503383)
- Samuels, R. (2009). The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds. In K. Frankish and J. St. B. T. Evans (Eds.) *In two minds* (pp. 129-146). Oxford, UK: Oxford University Press.
- Schachter, S. & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69 (5), 379-399. [10.1037/h0046234](https://doi.org/10.1037/h0046234)
- Scott, R., Dienes, Z., Barrett, A. B., Bor, D. & Seth, A. K. (2014). Blind insight: Metacognitive discrimination despite chance task performance. *Psychological Science*, 25 (12), 2199-2208. [10.1177/0956797614553944](https://doi.org/10.1177/0956797614553944)
- Seth A. (2015). The cybernetic bayesian brain. In T. Metzinger and J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Thagard, P. & Schröder, T. (2014). Emotions as semantic pointers: Constructive neural mechanisms. In L. F. Barrett and J. A. Russell (Eds.) *The psychological construction of emotion* (pp. 144-167). New York, NY: Guilford.
- Thompson, V. A. (2009). Dual-process theories: A metacognitive perspective. In K. Frankish and J. St. B. T. Evans (Eds.) *In two minds: Dual processes and beyond* (pp. 171-195). Oxford, UK: Oxford University Press.
- Yu, A. J. & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46 (4), 681-692. [10.1016/j.neuron.2005.04.026](https://doi.org/10.1016/j.neuron.2005.04.026)