http://doi.org/10.7213/1980-5934.31.054.DS13
ISSN 1980-5934

Emotion and the predictive mind: Emotions as (almost) drives

Emoção e a mente preditiva: Emoções (quase) como impulsos

JOSÉ M. ARAYA[1]

Abstract

Given its simplicity and enormous unifying and explanatory power, the predictive mind approach to mental architecture (predictive processing) is becoming an increasingly attractive way of carrying out theoretical and experimental research in cognitive science. According to this view, the mind is constantly attempting to minimize the discrepancy between its expectations (or sensory predictions) and its actual incoming sensory signals. In the *interoceptive inference view of emotion* (IIE), the principles of the predictive mind have been extended to account for emotion. IIE holds that, *in direct analogy to visual perception*, emotions arise from interoceptive predictions of the causes of current interoceptive afferents. In this paper, I argue that this view is problematic, as there are arguably no regularities pertaining to emotion in the physiological inner *milieu*, from which the relevant interoceptive expectations could be learned. Therefore, it is unlikely that our expectations relative to emotion involve interoceptive expectations in the way required by IIE. The latter view should then be amended. In this respect, I suggest that emotions might arise via *external interoceptive active inference*: by sampling and modifying the external environment in order to change a valenced feeling. Thus, if the predictive mind approach is on track, emotions are not to be understood in direct analogy to perception (e.g., vision). Rather, I suggest that the view of emotion that emerges from the predictive mind is in line with motivational approaches to emotion. In the suggested view, (almost) just as drives (or 'homeostatic motivations'), emotions are suited for the active regulation of the inner *milieu* by sampling the environment in order to finesse our emotion expectations. In this view, emotions are individuated, and differ from drives, in virtue

[1] Instituto de Filosofía y Ciencias de la Complejidad (IFICC), Santiago de Chile. PhD, email: jo.araya.g@gmail.com

of the distinctive sampling policies ('actions') characteristic of the high levels of the predictive hierarchy.

Keywords: Emotion. Interoception. Interoceptive inference. Perceptual theory of emotion. Motivational theory of emotion.

Resumo

Dada sua simplicidade e enorme poder explicativo e unificador, a abordagem da mente preditiva à arquitetura mental (processamento preditivo) vem se tornando uma forma cada vez mais atrativa para a realização de pesquisas teóricas e experimentais em ciências cognitivas. De acordo com essa visão, a mente está constantemente tentando minimizar a discrepância entre suas expectativas (ou previsões sensoriais) e seus reais sinais sensoriais de entrada. Na visão de inferência interoceptiva da emoção (IIE), os princípios da mente preditiva foram estendidos para explicar a emoção. A IIE sustenta que, em analogia direta à percepção visual, as emoções surgem de previsões interoceptivas das causas de atuais aferentes interoceptivos. Neste artigo, argumenta-se que essa é uma visão problemática, pois indiscutivelmente não há regularidades relativas à emoção no meio fisiológico interno a partir do qual as expectativas interoceptivas relevantes poderiam ser aprendidas. Portanto, é improvável que nossas expectativas relativas à emoção envolvam expectativas interoceptivas da forma exigida pela IIE. A última visão deve então ser alterada. A esse respeito, sugere-se que as emoções podem surgir via inferência interoceptiva externa ativa: amostrando e modificando o ambiente externo para alterar um sentimento valenciado. Assim, se a abordagem da mente preditiva estiver no caminho certo, as emoções não devem ser entendidas em analogia direta à percepção (por exemplo, a visão). Ao invés disso, sugere-se que a visão da emoção que emerge da mente preditiva é compatível com abordagens motivacionais à emoção. Na visão sugerida, (quase) como impulsos (ou "motivações homeostáticas"), as emoções são adequadas para a regulação ativa do meio interno, amostrando o ambiente para satisfazer nossas expectativas emocionais. Nessa visão, as emoções são individualizadas, e diferem dos impulsos em função das políticas de amostragem distintivas ("ações") características dos altos níveis da hierarquia preditiva.

***Palavras-chave**: Emoção. Interocepção. Inferência interoceptiva. Teoria perceptiva da emoção. Teoria motivacional da emoção.*

Introduction

Given its simplicity and enormous unifying and explanatory power, the predictive mind approach to mental architecture, a.k.a. predictive processing, is becoming an increasingly attractive way of carrying out theoretical and experimental research in cognitive science. Now, predictive processing (PP) is not just another compelling theoretical approach to some cognitive function. PP has the ambitions to constitute itself as an overarching paradigm shift in our understanding of the functioning of the mind. The ambition is high. The principles of PP promise to give us a unifying account of *all* the seemingly disparate variety of mental phenomena, ranging from perception to action (Clark, 2013, 2016; Hohwy, 2013).

PP is already doing explanatory work in a wide variety of psychological domains. However, PP was mainly conceived and developed as an account (and re-conceptualization) of *perceptual* processes. That is why its principles have been mainly applied in the explanation of mental phenomena that, in some way or another, can be readily understood as perceptual in nature – e.g., visual perception, binocular rivalry, illusions and delusions, etc. (for a review, see, e.g., Clark, 2013; Friston, 2005).

Now, according to the Jamesian view of emotion (James, 1884), a.k.a. perceptual theories, emotions can be understood as *perceptions* of distinct bodily, interoceptive changes. Considering that the Jamesian view that emotions can be understood as a perceptual process has recently seen a resurgence of interest in emotion research (e.g., Prinz, 2004), an obvious next step for PP's explanatory ambitions is to apply its principles in accounting for emotion.

To date there is no fully developed PP account of emotion on offer. However, Seth (Seth, 2013, 2015; Seth et al, 2012; Seth and Critchley, 2013)—lo see also Hohwy (2013, pp. 242- 244)—have recently extended the principles of PP to emotion generation, and offered a first sketch of how such extension might go. Taking into account the fact that

PP mainly worksas an account of perception, and that perceptual, interoceptive views of emotion have been recently compellingly defended, these first sketches can be read as suggesting a PP version of the perceptual, interoceptive view of emotion. In this account, in direct analogy to visual perception (Seth, 2015), emotion is seen then as arising from *interoceptive inferences*.

In the *interoceptive inference view of emotion* (IIE), the principles of the predictive mind have been extended to account for emotion. According to this view, emotions arise by minimizing the difference between our interoceptive expectations and the actual incoming interoceptive signal. That is emotions arise by minimizing interoceptive prediction error (PE). More precisely, IIE holds that, *in direct analogy to visual perception* (Seth, 2015; Seth and Friston, 2016), emotions arise from interoceptive predictions of the causes of current interoceptive afferents, so that interoceptive PE is minimized. In this paper, I argue that this view is problematic, as there are arguably no regularities pertaining to emotion in the physiological inner *milieu*, from which the relevant interoceptive expectations could be learned. Therefore, it is unlikely that our expectations relative to emotion involve interoceptive expectations in the way required by IIE. The latter view should then be amended. Now, I agree with IIE's claim that emotions arise by minimizing interoceptive PE. However, I will propose that, contrary to IIE, emotions do *not* arise by minimizing interoceptive PE in direct analogy to vision—as in Jamesian views. I will propose instead that emotions arise by minimizing interoceptive PE in direct analogy to action. That is, I will suggest that emotions, instead of arising via interoceptive *perceptual inference*, arise via *external interoceptive active inference*. In other words, the view of emotion that emerges from the predictive mind is in line with motivational approaches to emotion. In the suggested view, (almost) just as drives (or 'homeostatic motivations'), emotions are suited for the active regulation of the inner *milieu* by sampling the environment in order to finesse our emotion expectations. In this view, emotions are individuated, and differ from drives, in virtue of the distinctive sampling policies ('actions') characteristic of the high levels of the predictive hierarchy.

In section 2, I present the basics of the PP framework. Then I present IIE (section 3). In section 4, I show that IIE is indeed committed to the assumption that there must be regularities pertaining to emotion in the physiological domain. In section 4.1, I argue that certain strands of evidence suggest that such an assumption is likely not to be the case. Then, in section 5, I put forward the view that emotions, instead of arising via interoceptive perceptual inference, arise via external interoceptive active inference, in line with motivational approaches to emotion. I conclude (section 6) with some comments.

The predictive mind: the very basics

The mental architecture posited by PP describes the rich, hierarchically organized interplay between higher-level sensory expectations (top-down driven knowledge) and lower-level sensory information (Clark, 2013, 2016; Hohwy, 2013). In a nutshell, in the PP framework, the mind/brain uses its learned knowledge about the regularities of the environment in order to generate, form the top-down, predictions about the incoming sensory signals that the environment constantly triggers in its sensory periphery. Contrary to the traditional approach to perception, in which sensory signals are aggregated in a bottom-up fashion until a percepts is finally formed, PP holds that to construct a meaningful percept of what the environment offers for the agent, the mind/brain must infer the most likely environmental (hidden) causes of its incoming sensory signals.

Importantly, the difference between such sensory predictions and the actual incoming sensory signal is known as *prediction error*. According to PP, all what the brain does, in all its functions, is to minimize its prediction error (PE). The latter is then a key construct in the predictive approach to the mind: PE is used as a bottom-up learning signal that improves our top-down expectations/predictions (so that our predictions can achieve to successfully minimize PE), and it also functions, as I will comment below, as a signal that drives action so that the environment can be modified in line with the agent's top-down expectations.

Now, in order to adaptively (flexibly) calibrate how much weight to assign to its top-down expectations relative to the bottom-up PE signal, the mind/brain needs to infer, in a context- sensitive manner, how reliable is the error signal relative to its expectations (think of vision in a foggy day: prediction errors are not much reliable nor informative). This sort of deeply context-sensitive metacognitive process is known as *precision-weighting*. In the PP framework, PE is always weighted in terms of its precision.

In the PP framework, perception is then understood as a form of inference, namely, *perceptual inference*. Percepts are formed in a top-down fashion, by predicting incoming lower-level sensory signals from higher-level hypotheses of the likely (hidden) causes of those signals in the environment. That is, perception requires finding a hypothesis of the world able to predictively fit incoming information—for example, and to put it blatantly colloquially, "if it is a glacier, instead of a rock wall, such and such signals are expected. Do these predicted signals fit incoming information?" If expected signals fit incoming information, PE is minimized and a percept of a glacier is formed. Interestingly, according to the predictive mind approach, action can also be seen as operating under the same imperative toward PE minimization. Action takes here the form of *active inference*. The latter consists in changing the environment so as to obtain sensory signals that fit considered predictions. Thus, while perception consists in changing hypotheses to fit incoming signals; minimizing PE via active inference requires instead maintaining the hypothesis about the world unchanged—which will trigger PE, as the hypothesis's predictions won't fit incoming signals—and modifying the environment to fit the incoming sensory information. In this manner, such 'self-triggered' PE is minimized.

As Clark (2013, 2016) remarks, perception and action influence each other in a constant, mutually constraining cycle. Thus, it is arguably sterile to attempt to account for each of them in isolation. Even more, perceptual inference and active inference are two ways of doing the same thing, namely, minimizing PE. However, this should not be taken to imply that perception and action amount to the same thing, i.e., that perception confounds with action in such a way that there is no fundamental distinction to be made between the two. Even though they operate under the same principle of PE minimization, active inference and perceptual inference differ in an obvious functional respect: they exhibit different direction of fit. While perceptual hypotheses have mind-to-world direction of fit, active inference has world-to-mind direction of fit.

Now, the predictive mind is inherently hierarchical. Low levels of the hierarchy (which are closer to the sensory periphery) encode regularities that operate at fast timescales—and which involve spatially narrower aspects of the visual field. These levels capture variant aspects of experience. On the other hand, high levels (which are relatively further from the sensory periphery) encode increasingly more complex regularities that operate at slow timescales—and which involve spatially wider aspects of the visual field. These levels capture relatively more invariant aspects of experience. For example, in the case of vision, low levels encode regularities such as the details of edges and the changing contours of objects as one moves (represented in V1), which have small receptive fields. While high levels encode relatively more invariant information (represented in the temporal lobes), which involves wider receptive fields, such as the enduring face and body of someone you know. In a word, while low levels model the more circumscribed, fast changing aspects of the world (e.g., the moving shades of leaves in a windy day); high levels model the increasingly less circumscribed, more abstract aspects of the world (e.g., a whole tree, a living entity, wind season, spring, and so on).

Crucially, via learning, sensory expectations manage to recapitulate the structure of the world. As Hohwy (2013) remarks, as the process of hypothesis selection and revision unfolds, and learning thus takes place, visual expectations manage to extract regularities of their proper domain, namely, light-reflecting objects. In other words, expectations are learned from experience (i.e., exposure and training), and over time they recapitulate the regularities that configure the hierarchically nested structure of the world. This is precisely what allows the system to issue successful predictions of the worldly causes of incoming signals, and thus minimize (precision-weighted) PE[1].

The interoceptive inference view of emotion

In the *interoceptive inference* approach to interoception (Seth, 2013, 2015), the principles of PP have been extended to account for interoception, i.e., the perception of the homeostatic, physiological condition of all tissues of the body (Craig, 2015). As an account of emotion *per se* (e.g., anger, pride, fear, joy, guilt, etc.), the interoceptive inference approach becomes, what might be called, the *interoceptive inference view of emotion* (IIE) (Seth, 2013, 2015; Hohwy, 2013). As an account of emotion *per se*, IIE is then in line with Jamesian views of emotion, according to which emotions arise from bodily, interceptive perception.

IIE holds that emotions arise *in direct analogy* to the way in which visual percepts are formed (Seth, 2015; Seth and Friston, 2016). However, interoception, rather than vision, is the relevant modality: just as visual percepts arise via visual *perceptual inference*, emotions must arise then via *interoceptive perceptual inference*. In this view, emotions amount then to bodily, interoceptive perceptions: Emotions arise from interoceptive predictions of the causes of current interoceptive afferents. For an emotion to arise, emotion hypotheses need to predict from the top-down incoming interoceptive signals, by finding an emotion-hypothesis that fits those signals, and thus minimize interoceptive PE. According to this view, emotions are then "reduced to basic interoceptive states" (Hohwy, 2013, p.243) and our perception of them: "emotion arises as a kind of perceptual inference on our own internal states." (Hohwy, 2013, p.243). Now, in IIE, the content of a certain high-level emotion hypotheses—e.g., 'the anger- hypothesis' or 'the fear-hypothesis'—determines the content of the interoceptive percept that is formed, and consequently, it determines the content of the bodily experience that ensues, which according to IIE, constitutes the experience of emotion. So, in this account, emotion hypotheses shape interoceptive percepts from the top-down during their formation. IIE holds that this solves, analogously to the case of vision, the underdetermination between emotion types and physiological input. In this sense, emotion differentiation can be explained in terms of the content of the high-level hypotheses (e.g., anger-hypothesis vs fear-hypothesis) that are brought to bear on the modulation of interoceptive perceptions (see Hohwy, 2013). For example, let's say that an individual entertains two interoceptive hypotheses with the same posterior probability about the emotion that might be causing certain interoceptive signals: the fear-hypothesis and the excitement-hypothesis. In this case, she could decide between the two by cognitively determining the nature of the context in which she finds herself: Does the context make more likely the fear-hypothesis or the excitement-hypothesis? Let's say that, in the case in question, the individual sees that a snake is approaching. Thus, the interoceptive hypothesis for fear acquires higher posterior probability: the interoceptive signals expected for the fear-hypothesis are generated from the top-down, let's say that interoceptive PE is successfully minimized by such data, so fear then "[…] arises as interoceptive prediction error is actually explained away" (Hohwy, 2013, p. 243).

This aspect of IIE makes it a particularly interesting account of emotion. Insofar as it incorporates high-level knowledge into interoceptive perception, and claims that the content of interoceptive experience is determined by the content of higher-level emotion hypotheses, IIE puts together key insights of both, Jamesian and two-factor, Schachterian views of emotion[2]. However, note that the claim that high-level emotion knowledge shapes interoceptive perception does not make IIE a strictly two-factor, Schachterian view. This is the case because, in the latter kind of view, 'cognition' has the function of merely categorizing (or 'labeling') current interoceptive experience. Consequently, *an interoceptive percept*, which in Schachterian views, and contrary to Jamesian views, is ambiguous concerning a specific emotion type, *has already been formed*. Thus, in strictly two-factor, Schachterian views, contextual knowledge only merely categorizes (or 'labels') an already formed percept, without shaping it or playing a modulatory role in the formation of such percept, as a PP perceptual view must claim—remember that, in this respect, as Hohwy (2013) emphasizes, PP is not a view about categorization, but rather a view on percept formation.

IIE is problematic

As it stands, IIE exhibits a problematic key assumption. Remember that in the PP framework, top-down expectations, via learning, manage to extract the regularities that configure the hierarchical structure of the world. In other words, expectations are learned from experience (via hypothesis selection and revision in light of precision-weighted PE), and over time they manage to recapitulate the causal regularities in the world at its different time-scales (Hohwy, 2013).

According to IIE, emotions result from *interoceptive* predictions. Where do interoceptive expectations come from? From the causal regularities that obtain in the inner physiological world, as patterns of changes in the physiological landscape are the kind of thing that causes interoceptive incoming signals—this is analogous to the platitude that visual expectations come from regularities involving light-reflecting objects. Thus, IIE is committed to the assumption that there must be causal regularities pertaining to the different emotion types in the physiological domain, from which the relevant interoceptive expectations could be learned.

However, certain strands of evidence point to the claim that there are no regularities pertaining to emotion in the physiological domain.

Against some versions of the 'natural kinds' view of emotion, L.F. Barrett and her colleagues have compellingly made the case for the claim that the physiological landscape does *not* exhibit "distinctive sets of correlated properties" (Barrett, 2006, p.33) that could configure anger, fear, joy, etc. That is, there are no physiological response patterns that instantiate regularities pertaining to emotion types in the inner *milieu* (see, e.g., Barrett, 2006; Quigley and Barrett, 2014). This is mainly the case since statistical analyses of meta-analytical studies on emotion evince that there is no robust specificity in autonomic activity measures across emotion studies. This is not the place to unfold this one-hundred-years-controversy in any detail, so I refer the reader to Barret's work on the matter. However, it is worth mentioning that, within philosophy, her arguments to the effect that there are no regularities pertaining to emotion in the physiological domain have been widely taken to support this conclusion (e.g., Carruthers, 2011; Ritchie and Carruthers, 2015; though see Colombetti, 2014, pp.35- 36). Even influential sympathizers of the view that emotions are biological natural kinds, which have autonomic 'signatures', such as Scarantino (2009), have recognized that Barrett (2006b) indeed achieves to show that the conclusion in question is the case. Scarantino (2009) recognizes that Barrett has shown that there are no physiological response patterns that instantiate regularities relative to the kind of mental states that we take emotions to be (anger, fear, joy, etc.)—i.e., the kind of mental states that typically constitute the *explandum* of an emotion theory[3].

There seem to be then no distinct bodily, physiological regularities relative anger, fear, joy, sadness, etc. In other words, evidence strongly suggests that there is no significant causal regularity connecting emotion types and patterns of physiological changes, so that a certain emotion type could predict physiological patterns—emotion types and patterns of physiological changes are statistically independent phenomena. To put it this way, if, from the third-person point of view, we go and take a look at the physiological landscape, we would see that there are no distinct emotion types there to be found. Without emotions configured in the physiological domain, it is hard to see how interoceptive expectations relative to emotion types could be built in the first place.

Considering that there are no emotions configured in the inner *milieu*, it is unlikely that the mind/brain stores expectations about which *interoceptive* signals to expect given a certain emotion hypothesis. Therefore, it is unlikely that emotion hypotheses get to encode interoceptive expectations in the way required by IIE. The experience of emotion does not arise then by minimizing interoceptive PE by generating *interoceptive* signals from emotion-hypotheses. The latter do not seem then to be playing the role of shaping interoceptive percepts, so that the latter could constitute the experience of a certain emotion type. Emotion models must encode then, primarily, expectations about *other* sort of information.

This argument is analogous to the following, more familiar argument. Considering that there are no regularities pertaining to *cloud-types* in the auditive domain (or in the 'detectable vibrations domain')—i.e., there are no auditive regularities pertaining to *cirrus*, nor to *cumulus*, *stratus*, etc.—it is unlikely that cloud-hypotheses encode auditive expectations. Without clouds in the auditive domain, it is hard to see how cloud-hypotheses could get to build expectations about which auditive signals to expect given a certain cloud-hypothesis. Or to put it this way, given that there are no clouds in the auditive domain, the experience of clouds does not arise, primarily, by minimizing auditive PE by predicting auditive signals from cloud-hypotheses. If this argument works in this case, it should also work in the emotion case.

Emotion as external interoceptive active inference

As there are no emotions to be found in the physiological landscape, perceiving/feeling our physiology cannot be then the *whole* story about emotion *per se*. Contrary to IIE's claim, predicting interoceptive signals during perceptual inference cannot be what is primary in emotion generation. Thus, IIE lacks a thoroughly compelling way to account for emotion *per se*.

This might sound rather puzzling. On the one hand, forming an interoceptive percept by predicting interoceptive signals cannot be what is primary in emotion generation. On the other hand, common sense (and also experimental research) tells us that every time we experience an emotion, however, this experience is accompanied by interoceptive, bodily feelings. Of course, this is no real puzzle. This simply suggests that, even though having an emotion does not consist in perceiving interoceptive changes, having an emotion does involve some type of process that must be intertwined with interoception. It is at this juncture that the view proposed in this paper comes forward. Assuming that the PP framework is on track, emotions must arise then by minimizing interoceptive PE. However, emotions, as I argued, do *not* arise by minimizing interoceptive PE in the *specific* way proposed by IIE. The latter needs to be amended.

Now, remember that according to the predictive mind, all what the brain does is to minimize PE, and that there are two ways of doing this: via perceptual and active inference. If emotions do not simply arise via *perceptual* inference, as I argued above, we are left then with *active* inference. Emotions must arise then via *interoceptive active inference*, instead of via interoceptive perceptual inference. I want to explore this proposal. In other words, emotions do not consist in bodily, interoceptive perception. Rather, emotions are strategies for changing an interoceptive percept that has already been formed (via interoceptive perceptual inference). That is, emotions are specific strategies for regulating affective valence (more on this below). Now, interoceptive perceptions (i.e., valence) inform about our homeostatic condition (Craig, 2015). Then, emotions are better seen as specific strategies for regulating homeostasis.

Some motivations for the action-oriented approach to emotion

The claim that emotions are fundamentally action strategies (for reducing interoceptive PE) is not arbitrary. Emotions have motivational force. Emotions are motivational states that urge us to act in different ways. I take this to be rather uncontroversial. Many common-sense phenomena point towards the centrality of motivated action in emotion. Let me mention just a couple of such phenomena. In the first place, the very existence of virtues speaks of the quintessentially motivational character of emotion. Virtues such as self-discipline, resilience, prudence, and temperance, amount precisely to the ability to control the motivational force of emotions. These character traits would not be *virtues* in the first place, if emotions lack motivational power in their very constitution. In the second place, we commonly appeal to the motivational force of emotions for the sake of explanation: “Christine rapidly hid her bottle of gin because she was scared of the police”, “Michelle made loud noises in the middle of the night because she was secretly angry at her husband John”. People have the urgency to retaliate in bursts of anger, to kiss out of love, to repair damage out of guilt, and to stay on the couch out of shame. All these sorts of action exhibit a sense of urge, a ‘motivational oomph’, which is accompanied by the expectation that such an urge will vanish after action completion. Thus, interoceptive active inference looks as a more than promising place at which to look in order to better understand emotion. It is worth then exploring whether this action-oriented aspect of emotion might be the core of emotion in the predictive mind.

Internal and external policies during homeostasis maintenance

Interestingly, interoception takes place as the organism attempts to regulate homeostasis. The latter basically consist in maintaining an optimal overall physiological regulatory balance able to keep the entire organism's body within its limits of viability. The interoceptive system, which tracks whole-body physiological changes, evolved for maintaining homeostasis (Craig, 2015). Interoceptive percepts inform then the organism about its current homeostatic condition.

Now, if there is a discrepancy between the prediction relative to the high-level expectation of homeostasis—e.g., expecting certain levels of hydration—and what the current lower- level interceptive perception informs—e.g., the experience of thirst—interoceptive PE is triggered (Seth, 2013). In order to minimize this interoceptive PE (i.e., homeostatic imbalance), agents have two kinds of actions available. They might be called *internal actions* and *external actions*. The former consists in automatically executing physiological 'policies' (or sets of 'actions') by making use of resources that are already available within the organism—e.g. releasing vasopressin in the case of thirst. However, given the fact that inner physiological 'policies'—e.g., releasing vasopressin in the case of thirst—can rarely rectify homeostatic imbalances by triggering inner physiological resources alone (i.e., we simply lack the physiological resources to re-hydrate ourselves by producing water), the interoceptive system engages actions in the external environment in order to rectify homeostatic imbalances (e.g., looking for some water). Motivating action is part of what the interceptive system does, to put it this way (see Craig, 2015; Devinsky et al., 1995). This are *external interoceptive actions* (allostatic actions). In the predictive mind, the latter take place via, what might be called, *external interoceptive active inference*.

Emotion, external interoceptive active inference, and knowledge of sensorimotor contingencies

The claim suggested in this paper is that emotions arise by minimizing interoceptive PE via *external interoceptive active inference*. Here the task consists in minimizing the discrepancy between an already formed interoceptive percepts and the hard-wired expectation (or 'goal') of stable homeostasis.

Now, as Seth (2015) shows (independently of his work on emotion), active inference requires knowledge of 'sensorimotor contingencies'. That is, representations of the counterfactual relations that obtain between (possible) actions and its prospective sensory consequences (colloquially put: "if I act in this manner, sensory signals should evolve in such-and-such way"). Insofar as external interoceptive active inference is a form of active inference, it requires *interoceptive* knowledge of sensorimotor contingencies: "if I act in this manner, interoceptive signals should evolve in such-and-such way".

The view I propose is that a certain emotion type amounts to a specific strategy for minimizing interoceptive PE by way of an emotion-specific set of representations of 'sensorimotor contingencies'. That is, by way of stored knowledge of the counterfactual relations that obtain between (possible) actions and its prospective interoceptive consequences—"if I act in this manner, interoceptive signals should evolve in such-and-such way". An emotion arises when such knowledge is applied in order to regulate *affective valence*.

Very roughly, in the predictive mind, affective valence can be understood as arising from perceptual expectations of the rate of change of (interoceptive) PE (Jofilly and Coricelli 2013; Van de Cruys, 2017). As agents are *fundamentally* in the business of minimizing *interoceptive* PE, (Seth, 2015)—given that interoception tracks the physiological states that agents need to maintain within viability limits (homeostasis)— and valence inform us about the things that matter to us (e.g., Prinz, 2010), the idea is that affective valence tracks the dynamics of PE reduction. Roughly, if an agent infers that she is doing well reducing interoceptive PE, positive valence takes place (and negative valence, if the other way around). In this sense, affective valence can be taken to inform about how well the agent is doing in maintaining homeostasis.

We are now in a better position to unfold the proposed view. Emotions arise via *external interoceptive active inference*: by sampling and modifying the external environment in order to change an already formed interoceptive percept. This percept constitutes valence, and informs about homeostatic imbalances. Thus, emotions are specific strategies for regulating affective valence, and consequently, homeostasis. A certain emotion type is generated when its characteristic sensorimotor contingencies ("if I act in this manner, interoceptive signals should evolve in such-and-such way") are used to fulfil the expectation of interoceptive, physiological balance. Emotions are specific strategies for regulating affect by way of specific forms of action-oriented knowledge.

The idea is that, initially, a certain event triggers physiological changes in the organism. This event is typically triggered by an exteroceptively perceived external event. For example, a letter stating that your landlord needs to take back the property. The physiological changes that have been triggered by some external event are interoceptively perceived as positive or negative, given homeostatic expectations and the dynamics of the relative dynamics of PE reduction (to put it colloquially: "how am I faring in reducing interoceptive PE?"). Now, remember that, as I commented above, the discrepancy between an already formed interoceptive percept that informs about current homeostatic condition and the hard-wired expectation (or 'goal') of stable homeostasis constitutes *high-level interoceptive PE* (think of the experience of thirst). In other words, negative valence reflects states which are incompatible with the high-level expectation (or 'goal') of maintaining homeostasis. (Note that, in a certain sense, *positive* valence also reflects states which are incompatible with the high-level expectation of homeostasis maintenance. This is the case since positive physiological changes amount to changes which are *approaching* the 'goal' set by homeostatic standards. That is, such physiological changes are not yet quite in line with the standard in question. The phenomenon of *allostasis* shows that this is the case (Cabanac, 1971). Pleasure typically takes place as a homeostatic imbalance begins to be rectified. However, pleasure stops as such an imbalance has already being rectified (Cabanac, 1971, 1979). Think of the 'homeostatic motivation' of hunger, and its corresponding process of satiation. When an organism is hungry and eats something nutritious, the pleasure obtained from that stimulus is significant. However, as the organism in question already begins to be satiated, the hedonic value of food decreases, to the point that, as the organism is already satiated, food tends to become aversive (Cabanac, 1979). In this sense, pleasure is a form of 'ongoing relief'.)

Now, once such high-level interoceptive PE is triggered, the main task of the interoceptive system is *not* now forming a percept, but rather bringing physiological variables to their expected state by minimizing high-level interoceptive PE. In the PP framework, this means that active inference needs to be engaged: actions must be brought forth to fulfil homeostatic expectations. Taking into account the fact that the organism cannot minimize high-level interoceptive PE via internal interoceptive actions, external interoceptive actions are motivated. Insofar as external interoceptive actions are a form of active inference, they require representations of 'sensorimotor contingencies'. Thus, the view I am suggesting is that when high-level interoceptive PE is triggered by any sort of event, and it is minimized via the set of 'sensorimotor contingencies' characteristic of the, let' say, 'anger-hypothesis', the emotion of anger arises and it is experienced.

Emotions as (almost) 'homeostatic motivations'

Note that this view sees emotion in analogy with 'homeostatic motivations' or drives, such as for example, hunger. Hunger amounts to a mental state that is constituted by both, the negatively valenced state of an empty stomach, plus the motivation to act in the world in such a way so as to change such a bad feeling. Analogously, emotions, if the suggested view is on track, are also constituted by both, a valenced state, plus the motivation to act in such a way as to change such state. In this view, the difference between emotions and homeostatic motivations or drives lies in the nature of the content of the sensorimotor contingencies used for regulating homeostasis. While the drive of hunger involves sensorimotor contingencies characterized by the expectation (or 'goal') of, let's say, finding food; the emotion of anger involves sensorimotor contingencies characterized by the expectation (or 'goal') of, let's say, eliminating the origin of a demeaning offense.

Individuating emotions

As we saw in section 2, the architecture posited by the PP framework is inherently hierarchical. Thus, knowledge of 'sensorimotor contingencies' must also be found across all levels of the cortical hierarchy: representations of 'sensorimotor contingencies' can be low- level or high-level depending on how variant or invariant are the regularities that they encode, respectively. For example, low-level, fast-changing actions include movements such as microsaccades. Slower time-scale actions include arm movements, or walking. Even more 'abstract', slower-timescale actions can include actions such as waiting for the night to fall, doing a Postdoc, or working as a Lecturer.

In the PP framework, high-level expectations of action constrain and modulate lower-level predictions. If the system has the high-level expectation (or 'goal') of eating, this can be achieved, depending on context, by several different cascades of lower-level precision- predictions. For example, and depending on context, the system can achieve the expectation of eating by extending the arm, walking to the fridge, cycling to the supermarket, etc. In turn, these lower-level predictions (or 'sub-goals') can be fulfilled in several different ways depending on context. In fact, the lower in the hierarchy, the more the context-dependent variability of the precision-weighted predictions in question: the relatively low-level expectation (or 'goal') of grasping your mug, can be fulfilled via very distinct predictions about shoulder and wrist micro-movements, depending on what the context affords—e.g., your initial position, room temperature, metabolic resources, etc. High-levels constrain and modulate lower-levels. High-levels encode expectations of action which are coarse-grained, while lower levels encode expectations of action which are fine-grained—the latter seem to be rather automatic, while he former seem to be more intentional.

It follows from the above paragraphs that the expectations of action that each emotion hypothesis encodes to minimize high-level interoceptive PE must also be seen as represented at different time-scales or levels of abstraction. That is, the actions specified by the sensorimotor contingencies encoded by emotion hypotheses exhibit different degrees of granularity. There are expectations of action relative to emotion which are very abstract. For example, and to keep the example of anger above, the expectation (or 'goal') of eliminating the origin of a demeaning offense. There are also expectations of action which are relatively low-level. The latter amount to context-sensitive ways of fulfilling the high-level expectation (or 'goal') in question. In this case, the abstract prediction in question can be fulfilled by several distinct low-level expectations (or 'sub-goals'). For example, attacking, making a phone call, making an ironic joke, sighing, etc. In turn, these lower-level expectations can be fulfilled by an even richer array of relatively lower-level predictions. For example, attacking can be fulfilled by running towards the offender, or by slowly walking towards the offender while expanding the chest, etc. In turn, these latter expectations of action can be fulfilled by several lower-level predictions, and so on and so forth. In a word, knowledge of sensorimotor contingencies exhibits different degrees of granularity.

This distinction between levels of abstraction relative to the expectations (or 'goals') that emotion models/concepts encode is thoroughly compatible, *mutatis mutandis*, with the distinction between *relational goals* and *situated goals* made by Scarantino (2014):

> "[…] *relational goals* are *abstract goals* that need to be situated in a *concrete context* in order to guide bodily changes. This is typical of most goal-oriented processes, including non-emotional intentional actions. When we decide to get to school by 10am in order to attend a talk, the *overarching action goal* of getting to school by 10am can be achieved through a variety of *situated goals* (e.g., taking a bus at 9:20am, taking the subway at 9:30am) (cf. Pacherie 2008). Each of these situated goals can in turn be achieved by a variety of *motor goals* that directly guide bodily changes. For simplicity of reference, I will distinguish between the *relational goal* of an emotion and its *relational sub-goals*, understood as the collection of *situated* and *motoric* goals by which the relational goal can be achieved." (Scarantino, 2014, p. 169)

The individuation claim suggested in this paper is that when high-level interoceptive PE is minimized via the set of sensorimotor contingencies that corresponds to stored knowledge about emotion *E*, the emotion *E* is generated. The claim is that the kind of knowledge in question is high-level. That is, it specifies action expectations which are abstract (i.e., 'relational goals'). For example, in the case of anger, the expectation of eliminating the origin of a demeaning offense. In this view, emotions are individuated by those emotion-specific abstract expectations.

Closely following Frijda (1986, 2010), the view here suggested avoids the classical problem that emotions cannot be individuated by sets of instrumental behaviours—as different sets of instrumental behaviours are involved in the same emotion type and vice-versa—by holding that the expectations of action which individuate emotion types are encoded at high-levels of the cortical hierarchy. At these levels, hypotheses encode slow time-scale regularities, which exhibit a rather abstract level of granularity. That is, these levels do *not* encode specific sorts of instrumental behaviour and motor 'policies'. The latter are situationally driven, as I commented above.

The emotion-specific knowledge of 'sensorimotor contingencies' that individuate emotions encodes expectations ('goals') relative to the types of problem with which a certain emotion type consistently needs to deal. For example, and following Frijda (1986, 2010), the emotion- specific knowledge of 'sensorimotor contingencies' that individuate *anger* can be taken to consist in expectations relative to the task of *regaining control of action to remove obstruction* (see Frijda, 1986, p. 88). In the case of *fear*, the emotion-specific knowledge of 'sensorimotor contingencies' that individuate it can be taken to consist in expectations relative to the task of *making oneself inaccessible to the relevant stimulus so as to avoid it*. These actions are engaged since the system predicts that, given the kind of situation which is considered to be taking place, the actions in question will achieve to trigger interoceptive signals compatible with the expectation of homeostatic balance.

Conclusion

IIE holds that, *in direct analogy to visual perception*, emotions arise via interoceptive perceptual inference. In this paper, I argued that this view is problematic, as there are arguably no regularities pertaining to emotion in the physiological inner *milieu*, from which the relevant interoceptive expectations could be learned. I proposed then a manner by which IIE can be amended. I suggested that emotions might arise via *external interoceptive active inference*: by sampling and modifying the external environment in order to change a valenced feeling. Thus, if the predictive mind approach is on track, emotions are not to be understood in direct analogy to perception (e.g., vision). Rather, in the suggested view, (almost) just as drives (or 'homeostatic motivations'), emotions are suited for the active regulation of the inner *milieu* by sampling the environment. In this view, emotions are individuated, and differ from drives, in virtue of the distinctive sampling policies ('actions') characteristic of the high levels of the predictive hierarchy.

Interestingly, the agential view of emotion here suggested straightforwardly accounts for those aspects which are left unexplained by the current philosophically more developed agential theory of emotion, namely, the motivational theory of emotion (MTE) (Scarantino, 2014). Roughly, MTE is the view that "An emotion is a prioritizing action control system […] with the function of achieving a certain relational goal while correlating with a certain core relational theme." (Scarantino, 2014, p. 178). In the first place, MTE leaves unexplained an aspect that any agential theory must explain, insofar as it gives to action a primary role in the generation of emotion episodes. Agential theories, insofar as they are action theories, should say something about why emotions have the motivational force that they have, by appealing to the resources that the proposed theory itself provides. MTE is silent in this respect. The agential theory that, as I argued, emerges out of the PP framework has a straightforward answer. Emotions have the motivational force that they have, because they, just as drives, are grounded in the interoceptive system. The latter, as we saw, motivates action so as to maintain the organism within viability limits. Anger is much closer to hunger than one may think.

References

Barrett, L. 2006. Are emotions natural kinds? *Perspectives in Psychological Science*, *1*, 28– 58.

Cabanac, M. 1971. Physiological role of pleasure. *Science*, *173*(2), 1103–1107.

Cabanac, M. 1979. Sensory Pleasure. *The Quarterly Review of Biology*, *54* (1), 1-29.

Carruthers, P. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181– 204.

Clark, A. 2016. *Surfing Uncertainty*. Oxford: Oxford University Press.

Colombetti, G. 2014. *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge, MA: MIT Press.

Craig, A.D. 2015. *How do you feel? an interoceptive moment with your neurobiological self*. Princeton University Press.

Devinsky, O., Morrell, M.J., and Vogt, B.A. 1995. Contributions of anterior cingulate cortex to behaviour. *Brain*, *118*(1), 279-306.

Frijda, N. H. 1986. *The Emotions*. Cambridge: Cambridge University Press.

Frijda, N. H. 2010. Impulsive action and motivation. *Biological Psychology*, *84*, 570–9.

Friston, K. 2005. A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *360*, 815– 36.

Hohwy, J. 2013. *The predictive mind*. Oxford: Oxford University Press.

James, W. 1884. What is an emotion? *Mind*, *9*, 188-205.

Joffily, M., and Coricelli, G. 2013. Emotional valence and the free-energy principle. *PLoS Computational Biology*, *9*(6), e1003094.

Prinz, J. 2004. *Gut Reactions: A perceptual theory of emotion*. Oxford: Oxford University Press.

Prinz, J. 2010. For valence. *Emotion Review*, *2*, 5–13.

Quigley, K. S., and Barrett, L. F. 2014. Is there consistency and specificity of autonomic changes during emotional episodes? Guidance from the Conceptual Act Theory and psychophysiology. *Biological Psychology*, *98*, 82-94.

Ritchie, J.B., and Carruthers, P. 2015. The bodily senses. In M. Matthen (ed.), *The Oxford Handbook of the Philosophy of Perception* (pp. 353-371). Oxford University Press.

Scarantino, A. 2009. Core affect and natural affective kinds. *Philosophy of Science*, *76*, 940–957.

Scarantino, A. 2014. The motivational theory of emotions. In D. Jacobson and J. D'Arms (eds.), *Moral Psychology and Human Agency* (pp. 156–185). Oxford: Oxford University Press.

Seth, A. 2013. Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Science*, *17*(11), 565-573.

Seth, A. 2015. The cybernetic Bayesian brain: from interoceptive inference to sensorimotor contingencies. In T. Metzinger and J. Windt (eds.), *Open MIND* (pp. 9–24). Frankfurt, Germany: MIND group.

Seth, A. and Critchley, H. 2013. Extending predictive processing to the body: Emotion as interoceptive inference. *Behavioral and Brain Sciences*, *36*, 227–228.

Seth, A., Suzuki, K., and Critchley, H. 2012. An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *2*, 395.

Seth, A. and Friston, K. 2016. Active interoceptive inference and the emotional brain.

Philosophical Transactions of the Royal Society B: Biological Sciences, *371* (1708), 1–10.

Van de Cruys, S. 2017. Affective Value in the Predictive Mind. MIND Group; Frankfurt am Main. Retrieved from https://lirias.kuleuven.be/retrieve/443336

Notes

1. I am assuming a representationalist reading of the predictive mind (Hohwy, 2013). I am aware that in some corners of the cognitive science community it is held that sensory processing, and cognition more generally, does not harbour representations (Varela et al., 1991; Chemero, 2009). Enactivists emphasize that it is unlikely that the mind's job is to recover a mind-independent world in the way that a mirror captures the things that get to be in front of it. Minds evolved to act within its own ecology, to put it that way. What an organism is able to do specifies what she perceives, and vice-versa. As long as sensory representations are taken to be mirrors of an agent-neutral world, representations should certainly be looked with suspicion. I think these insights are on the right track. However, they do not speak against representations. There is no need to take representations as mirrors of an agent-neutral world. In fact, representations are arguably *action-oriented* (Clark,

1997; Millikan, 1996). That is, they jointly encode aspects of the world and specify relevant actions. In line with the insights of enactivism, the aspects of the world that they encode are better seen as capturing the task-relevant, ecologically salient aspects of the niche that the agent contributes to specify.

2. Roughly, Jamesian views of emotion holds that emotions amount to bodily perceptions, while Schachterian views hold that emotions amount to cognitive interpretations of current bodily experience ('arousal' for Schachter). Thus, in Schachterian views, emotions require one more 'factor' than Jamesian views. Schachterian views exhibit two-factors: bodily perception plus 'cognition'.

3. Scarantino goes on to propose, however, that emotion research should change the explanatory target of emotion theories. Emotion research should not have as explananda the mental states that we take emotions to be. Instead, emotion research should find another explanandum, though similar to the mental states that we take emotions to be. However, I think this is not a satisfactory move, as it attempts to offer poor substitutes of the mental states that we want to understand in the first place (see Dennett, 2009). It simply changes the subject. Anyway, this controversy cannot be resolved here.

RECEIVED: 09/02/2019
APPROVED: 09/28/2019

RECEBIDO: 02/09/2019
APROVADO: 28/09/2019