Journal of
Cognition
and
Neuroethics

Interoceptive Inference and Emotion in Music: Integrating the Neurofunctional 'Quartet Theory of Emotion' with Predictive Processing in Music-Related Emotional Experience

Shannon Proksch
The University of Edinburgh

Biography
Shannon Proksch has recently completed her master's degree at the University of Edinburgh. Most of her work centers on music and language cognition, especially the increasing role of music in empirical and theoretical studies of embodied, enactive, and social cognition.

Acknowledgments
This paper was written in satisfaction of the MSc in Mind, Language and Embodied Cognition at the University of Edinburgh. It was subsequently presented at the Mind and Brain conference in Flint, Michigan. Thanks are due to Jami Anderson and the conference organizers. Many thanks as well to my supervisor, Dr Alistair Isaac for his guidance and support; to my music professor Dr Nikki Moran and the Edinburgh Music Psychology Research Reading group for fostering my early ideas; and to my friends and family who have provided detailed feedback and encouragement throughout this project.

Publication Details
Journal of Cognition and Neuroethics (ISSN: 2166-5087). February, 2018. Volume 5, Issue 1.

Citation
Proksch, Shannon. 2018. "Interoceptive Inference and Emotion in Music: Integrating the Neurofunctional 'Quartet Theory of Emotion' with Predictive Processing in Music-Related Emotional Experience." *Journal of Cognition and Neuroethics* 5 (1): 101–125.

Interoceptive Inference and Emotion in Music: Integrating the Neurofunctional 'Quartet Theory of Emotion' with Predictive Processing in Music-Related Emotional Experience

Shannon Proksch

Abstract
In this paper, I discuss how the predictive processing framework expands upon traditional, bottom-up, two-factor theories of emotion as passive physiological evidence-building and subsequent cognitive appraisal (James, 1884; Lange, 1885; Schachter & Singer 1962), by incorporating active inference based on predictive models of the causes of external and internal stimuli (Seth & Critchley 2013; Seth & Friston 2016). Accordingly, I posit that emotional content from music is evoked as a result of active exteroceptive inference related to the physical musical stimuli (based on statistical regularities of the current musical event and past experience) as well as active interoceptive inference regarding the listener's current autonomic, and physiological states. In addition, I propose that this general predictive processing framework is implemented through a 'quartet' of neurofunctional mechanisms (Koelsch et al 2015) which are dynamically implicated in the neural and physiological processes underlying general, and music-related emotional experience. Conceptualising emotion as active inference over both external and internal processes, implemented and maintained through a dynamically interacting subset of neural pathways as offered by the Quartet Theory of Emotion, provides a more detailed mechanism by which music evokes emotion and results in the subjective difference in the reported emotional experience of music between individuals.

Keywords
Music, Emotion, Predictive Processing, Active Inference, Quartet Theory of Emotion, Brainstem, Diencephalon, Hippocampus, Neurofunctional, Social Cognition

Part 1

Pre-Show

How and what emotional content is communicated or evoked by music constitutes a central question for music cognition. It is a question that is especially difficult to investigate, since the emotional response to a given piece of music can vary so widely in content and precision from person to person. One popular story in music cognition is that music-related emotions arise from the violation and confirmation of musical expectancies based on statistical regularities of *external* musical features from an agent's past experience (Meyer 1956; Narmour 1990; Huron 2006). This fits well with predictive processing models of emotion, which are also based on expectancies, or predictions, that are honed by past experience. In predictive processing (PP) models, emotion arises from the process of minimizing prediction error[1] from active top-down predictions of the causes of *internal* bodily states (Hohwy 2013; Seth & Critchley 2013; Clark 2016). However, each of these accounts of emotion alone are too coarse grained to fully explain the underlying process of emotion in music and in general.

The predictive processing framework expands upon traditional, bottom-up, two-factor theories of emotion as passive physiological evidence-building and subsequent cognitive appraisal (James, 1884; Lange, 1885; Schachter & Singer 1962), by incorporating active inference based on predictive models of the causes of external *and* internal stimuli (Seth & Critchley 2013; Seth & Friston 2016). In the same vein, thorough descriptions of music-related emotional experience need to appeal to the *active* processing of expectancies of both interoceptive (internal, bodily) information as well as exteroceptive (external, structural) information, an explanation that is offered by the minimization of prediction error achieved through *active inference* within predictive processing.

However, relying heavily on the involvement of the insular cortex, with offhand mention of other emotional systems, these active inference accounts of emotion and interoception are importantly lacking the details of a distinct mechanism over which active inference is taking place. Such a mechanism can be described by appealing to neurofunctional models of emotional experience, (such as the Quartet Theory of Emotion (QTE) as outlined by Koelsch et al. 2015) which take into account how the brain and associated biological systems actually effect homeostatic maintenance and generate affective experience. I posit that incorporating QTE's multifaced neurofunctional mechanism can solve problems concerning the limitations within predictive processing accounts of emotion.

I propose that a network of interlinked neural systems, as offered by QTE, provides a set of low-level, high-level, and integrated neural mechanisms over which I claim PP occurs simultaneously, thus contextualizing the prediction error minimization within and between each system. Integrating this neurofunctional mechanism with predictive

1. Prediction Error Minimization (PEM) is basic principle underlying the predictive processing framework with the goal of "creat[ing] a closer fit between the predictions of sensory input based on an internal model of the world and the actual sensory input" (Hohwy 2013). This may be achieved in two ways: model-updating, or world-updating—as will be addressed further in this paper.

processing accounts of emotion in this way provides a more nuanced explanation of emotional experience than either theory offers alone—a claim that I will show is justified through an exploration of how QTE and PP, together, enhance our understanding of music-related emotions, shedding light on the answer to the question of how and what emotional content is communicated or evoked by music.

Set List

I will first provide a brief outline of current theories of music-related emotions and their proposed underlying mechanisms. Various mechanisms proposed by Juslin and Västfjäll (2008) will be shown to be linked by the further underlying component of expectancy. I will then explore predictive processing accounts of emotion, specifically active interoceptive inference, to address what theories of musical emotion and expectancy have right so far, as well as offer a further incorporation of interoceptive, physiological processes. Following this brief explanation of active interoceptive inference, a specific neurofunctional model as put forth by Koelsch et al (2015) will be reviewed: The Quartet Theory of Emotion (QTE). I will show that active interoceptive inference, in conjunction with the QTE, offers a mechanistically and functionally appealing account of emotional processes in general, and music-related emotions in particular. This project will end with a brief introduction to the explanatory value of combining interoceptive inference with QTE through reflecting on a series of vignettes of two fictional concert-goers with differing music-related emotional experiences.

Audience Introductions

> Music is the shorthand of emotion.
> —Leo Tolstoy

If this quote is accurate, then listening to music with someone might be a quick shortcut to sharing a similar emotional experience. We can listen to our country's national anthem at an Olympic games ceremony and share in the social emotions of pride and happiness or to a funeral hymn and share in the grief of our friends and family. However, it is easy to consider a case in which individuals do not share the same emotional experience when listening to the same music. Below, I introduce two fictional concert-goers, whose story will help us to evaluate the explanatory upshot of an integrated QTE + PP model in the case of music-related emotional experience.

Audience Spotlight 1

Let's reflect on the summer that Rylan, a college student in her late teens or early twenties, takes Henry, her grandfather, to a Rammstein concert. Henry wants to know what the young kids are listening to these days, and Rylan promised him that he would love their music. In fact, these industrial metal, hard rock performers make up Rylan's favorite band. Although she doesn't know German, their concerts always leave Rylan feeling energized and exuberant, as if she is right where she belongs.

Henry prefers country music oldies: Merle Haggard, Patsy Cline. But he loves his granddaughter, so he had happily tagged along for the experience. This concert, however, leaves Henry feeling agitated and uneasy, as if he distinctly doesn't belong.

Rylan and Henry are listening to the same music, with the same exteroceptive information stimulating their senses, and surrounded by the same physical context in a concert environment. However, they seem to be having substantially different emotional responses. There is something occurring in each concert-goers body and brain besides merely processing and reaction to external musical stimuli.

Part 2

Current Attempts at Understanding Music-Related Emotions

The Search for Underlying Mechanisms

Much has been said about the relationship between music and emotion. Music, as a language of emotion, is meant to evoke an emotional response in its listeners and those who partake in musical experience. Many theories have attempted to account for this pivotal role of emotion in music, or music on emotions, from appeals to the extra-musical associations such as the context of a sad moment of a play, or appeals to the "choreographing of expectation" (Meyer 1956). Rylan might enjoy Rammstein's music because it reminds her of her trip to Germany. Henry might dislike the concert because the music reminds him of German-language videos from his primary school lessons on the horrors of the Holocaust. However, it has been noted that, despite the plethora of accounts of music-related emotions and varying appeals to extra-musical associations, the current trajectory lacks a focus or explanation of the underlying neural and physiological mechanisms by which music effects emotional responses. Juslin and Västfjäll (2008, henceforth J&V) attempted to address this issue through identifying and explaining the underlying mechanisms behind music-related emotional experience.

These theories reviewed by J&V, as with most theories of emotion, are appraisal theories. They rest on the assumption that an emotion arises as a result of a cognitive appraisal, or a subjective evaluation informed by context on the personal level, relating to life goals or survival functions, and to which music-related emotions generally do not pertain (J&V 2008; Frijda 1993). Given the apparent lack of survival function in music-related emotion, the primary question J&V seek to answer is "how does music evoke emotions?" They claim the answer involves cognitive appraisal to some extent, as well as six key mechanisms that underlie emotional responses to music, briefly listed here to be explained in more detail in the next section: (a) brain stem reflexes; (b) evaluative conditioning; (c) emotional contagion; (d) visual imagery; (e) episodic memory; and, (f) musical expectancy.

J&V are correct in drawing attention to the lack of focus on underlying mechanisms of music-related emotions. However, they fail to notice that the six 'underlying' mechanisms they list are not the most 'basic' or 'fundamental' of mechanisms, and may be further grounded[2] in their relation to physiological response and homeostatic function,[3] which will be introduced later in this section, and further reviewed in part three.

What's Missing in the Search

Throughout the remainder of the paper, we will focus our attention on component (f) musical expectancy, and expand this notion to include physiological expectancy. In accord with Vuust & Frith (2008), I will claim that the combination of musical and physiological expectancy can be said to guide the first five of the proposed mechanisms. As will be elaborated throughout this paper, active interoceptive inference accounts of emotion, in conjunction with a neurofunctional model provided by the Quartet Theory of Emotion, provide a more nuanced explanatory account of both emotions in general and music-related emotional experience through an emphasis on the role of physiological

2. I take 'grounded,' in this sense, to refer to being further rooted in a more basic and fundamental mechanism.

3. Homeostasis refers to an organism's tendency to maintain its functioning with in a viable range, ensuring organism's survival. For example, the bodies of diabetic individuals fail to maintain homeostasis between levels of sugar and insulin without lifestyle intervention in the form of medicine or dietary changes.

homeostasis. Already, the six mechanisms provided by J&V can be linked through an underlying process of this sort, by cueing physiological and neural responses.

The Importance of Expectancy

In response to J&V, Vuust & Frith (2008) highlight the privileged role of expectancy in understanding music and its effects on emotion as an important component in a hierarchy of music-related emotional mechanisms. They claim that musical expectancy should be the most fundamental mechanism of the others in J&V's list. This is supported from a consensus among music theoreticians that musical experience and emotion is importantly conveyed by anticipation of local auditory events as well as of deeper musical and mental structures, such as overarching rhythmic and harmonic patterns, and contextualized by information about the piece and the memories of the listener. This view was originally explored in Meyer's 1956 work, "Emotion and meaning in music," and has been elaborated and extended to include physiological and neural processes of expectation in Huron's 2006 work "Sweet Anticipation." Accounts such as Meyer's and Huron's claim that expectation arises from culture dependent statistical learning, familiarity with a piece of music, short term memory for immediate musical history, and deliberate listening strategies (attention), and are consistent with the predictive processing framework. Vuust & Frith (2008) summarily claim that "the musical experience is dependent on the structures of the actual music, as well as on the expectations of the interpreting brain."

This response accurately points out that expectancy has a large and likely privileged role in the processing and perception of music-related emotions—however, it is incomplete to assume that so much emphasis should be placed on expectation related solely to the exteroceptive content in the auditory musical structure. Given the important role of physiological state in an emotional experience, expectancy theories of musical emotion should move beyond merely appealing to predictive processing of the structural features of the music itself. Instead, expectancies related to interoceptive/physiological states of the listener/performer themselves should be incorporated. As such, I will provide a more focused view of predictive processing of emotion in music below, emphasizing the role of active interoceptive inference. First, let's revisit our concert-goers:

Audience Spotlight 2

When we introduced Rylan and her grandfather, Henry, they were both attending the same concert and listening to the same music, however they had substantially different emotional responses. This means the music and context alone cannot serve as a one-to-one cue for a certain emotional state. Instead, maybe the music is cuing different extra-musical associations for our concert-goers which are themselves tied to different emotional responses. This remains an open possibility, however it does not help to fully explain the fundamental processes giving rise to their respective joy and unrest at the Rammstein concert. After a journey through some proposed underlying mechanisms, we paused on the important role of expectancy. Although Rylan and Henry are both listening to the same music in the same context and at the same time, they each have different musical expectations due to differing life and musical experiences. So, some of the difference in emotional response may be due to the mere exposure effect: Rylan has had more exposure to industrial metal music such as Rammstein and to the atmosphere of a Rammstein concert; thus, she has more precise expectations of the musical input and is more likely to enjoy it.

However, we are focusing heavily on exteroceptive information in justifying this difference in emotional experience, and we are analyzing the differences on a personal level. We still lack focus or explanation of the underlying neural and physiological mechanisms by which music effects emotional responses.

In the following sections, predictive processing accounts of emotion, including the active interoceptive account, as well as the neurofunctional mechanism proposed by the 'quartet theory' of emotion will be laid out, followed by their role in the processing of music-related emotions.

Predictive Processing Account of Emotion

Incorporating Interoceptive Expectancy

The idea that there is an interoceptive component to emotion is not new. Two-factor theories such as those proposed by James (1884), Lange (1885), and Schachter & Singer (1962) all discuss emotion as an appraisal of physiological arousal. However, these theories focus on feed-forward (bottom-up) models, where the sensory and autonomic systems accumulate evidence in the form of various physiological and arousal states, before being cognitively appraised and thus labeled as an emotion state. The active interoceptive inference account of emotion, however, assumes that interoceptive processes are subject to the same reciprocal relation between top-down predictions and bottom-up stimuli as the exteroceptive and proprioceptive processes involved in perception and action under the predictive processing framework (Seth 2013). In predictive processing (PP), cascading predictions from top-down generative models are met with bottom-up prediction errors, which serve to either update the predictive models or to motivate action (Clark 2013).

PP models of exteroceptive signals enable inference regarding the states of affairs in the external world which are most likely to cause a set of sensory states through the process of prediction error minimization (PEM). PP models of interoceptive signals, however, serve control and regulation of physiological states with the goal of maintaining homeostasis (Seth 2013; Sel 2014; Seth & Friston 2016). Predictions in interoceptive inference, rather than generative models about an external state of affairs, are internal homeostatic 'set points,' toward which the activity of autonomic and affect systems is driven as a result of minimizing the difference (prediction error) between this set point and current physiological state (Seth & Friston 2016). This grounds interoceptive prediction, at the subpersonal level, in evolutionary goals toward maintaining homeostasis, which is essential for biological fitness, and fulfills the survival function of emotions emphasized by Frijda (1993). Importantly, active interoceptive inference alone does not account for every aspect of emotional content, but augments existing two-factor theories of emotion by contextualizing interoceptive predictions with concurrent active inference over proprioceptive and exteroceptive cues. This involves a 'cognitive' appraisal in terms of higher cognitive functioning (such as stimulus evaluation),[4] but this appraisal need not be conscious or linguistically achieved. In contrast, linguistic labeling of an emotion state involves a translation, reconfiguration, and ultimately an approximation of a much more specific and nuanced underlying process:[5] an underlying process which integrates multiple sources of predictive information. The components of a process of this

4. Stimulus Evaluation Checks (SEC) are outlined in the component process model of emotion (Scherer & Zentner 2001; Scherer 2009; 1999). SECs refer to a multi-level appraisal process consisting of sequences of appraisal checks of the appropriateness of a stimulus and associated emotion (akin to the level of prediction matching or prediction error) from lower to increasingly higher levels of perceptual and cognitive processes.

5. In addition to the lack of semantic precision of emotion terms to emotion events, another known reason for the ineffectiveness of studying emotion appraisal gained from linguistic self-report is that these reports "may contain appraisals that are part of the emotional response, rather than belonging to its causes" (Frijda 1993).

sort are outlined in the following table in regard to the predictive information available in music.

Music Specific Sources of Predictive Information	
Exteroceptive Inference: PEM over the causes of external physical stimuli • the music itself (acoustic stimuli) • observation of the performer(s) movement, appearance, etc • other's body movement (dancing, etc)	**Interoceptive Inference:** PEM over the causes of autonomic, physiological stimuli (cued by exteroceptive/proprioceptive information) • heartbeat, breathing rate, etc • skin conductance • arousal • hormone activity
The integration of which leads to music-related emotions.	

This non- (perhaps pre-) linguistic cognitive appraisal, in the form of stimulus evaluation checks and prediction error minimization, occurs over many levels, and is achieved when prediction error is minimized at the lowest possible level, contextualized by predictive information from a multitude of sources.[6] When the predictive model fits (what level of homeostasis should be expected) then the emotional response stabilizes (Gerrans 2017). Thus, emotional systems, such as those which will be elaborated in the Quartet Thoery of Emotion (Koelsch et al 2015), coordinate other perceptual and inferential systems in the task of determining self-relevance or survival value of a stimulus with information from both low-level affect, and high-level metacognition. Activity in the insular cortex 'allows us to *feel* how things in the world matter to us, in the form of affect' (Gerrans 2017), with primary interoceptive representations in the posterior insular

6. For the purposes of this paper we will largely focus on the exteroceptive information available in acoustic music stimuli to a listener. Proprioceptive information is gained from movement to music, as well as the physical act of producing music itself, and may enhance any of the expectation effects of exteroceptive information as well as provide more direct cues for interoceptive process (for example, playing the drums will do more to increase blood flow and breathing rate than mere listening because of the direct physical action involved.)

cortex, and secondary integrated experience in the anterior insular cortex (Seth 2013; Seth & Critchley 2013).

What's Missing

Applying active interoceptive inference to emotion does extend existing two-factor theories of emotion by incorporating an integration of prediction and prediction error minimization over exteroceptive, proprioceptive, and interoceptive stimuli; however, we are still left with a two-factor framework. Current physiological condition is merely replaced with current homeostatic condition of the body, followed by cognitive appraisal in at least the minimal sense of stimulus evaluation, and perhaps action to maintain homeostasis. Even so, this consideration of physiological homeostasis may provide us a bit more insight into the experience of our concert-goers, Rylan and Henry:

Audience Spotlight 3

The last time we saw Rylan and her grandfather, Henry, we guessed that some of the difference in their emotional responses may be due to their individual musical histories. Rylan and Henry have different musical expectations due to different musical experiences. Henry's range of possible musical expectations is relegated to traditional country tunes, while Rylan's range of possible musical expectations includes industrial metal music, so she has more precise expectations of the music of Rammstein. While this mere exposure effect can to some degree explain our concert-goers' relative familiarity and positive or negative affect with respect to the Rammstein concert, we still don't know what is going on at the subpersonal level. Why is the exteroceptive information apparently cuing different emotional responses? For this, we turn to interoceptive inference accounts of emotion.

In addition to expectations of the external musical stimuli, Rylan and Henry each have expectations relating to their own internal bodily states. So, some of the difference in emotional response is due to differing histories of physiological responses related to music listening. Rylan has had more experience with the effects of Rammstein's music and the atmosphere of their concerts on her internal states. Thus, she has more precise expectations of her body's response to their music, and is more likely to enjoy it.

The internal bodily expectations of Rylan and Henry are not necessarily cognitive expectations (such as, 'I expect loud music and fast rhythms to energize me'), but are rather subpersonal, interoceptive expectations of physiological homeostasis. Rylan listens to heavy metal music quite frequently, so when the external music cues internal changes—rising heart rate, breathing rate, heightened levels of arousal etc.—this occurs relatively near the range of Rylan's homeostatic set point, but drastically out with the homeostatic expectations of Henry, causing him to feel agitated while Rylan instead feels energized.

We seem to have a decent explanation of the valence of similar arousal related emotions such as 'agitated' vs. 'energized.' However, incorporating active interoceptive inference still does not explain how the homeostatic range, or 'set point,' impacts nuanced emotions and *complex* social emotions, or how external musical stimuli can modify these physiological processes.

Active interoceptive inference accounts of emotion focus heavily on activity in the insular cortex. Within the insula is a viscerotopic map with general representations of interoceptive states, akin to the retinotopic map in your visual cortex. Relying on this viscerotopic map and ambiguously integrated experience within the insular and cingulate cortices—with offhand mention to other emotional systems—these interoceptive inference accounts lack a clear description of the active *mechanism* which is implemented in the maintenance of homeostasis occurring across multiple neural and bodily systems. I propose that this mechanism can be filled in to the active interoceptive inference accounts by incorporating a neurofunctional model of emotional experience, which more explicitly details how the brain and associated biological systems effect homeostatic maintenance and generates affective experience. In the following section, a plausible mechanism over which active interoceptive inference occurs will be explored in reference to the neurofunctional model put forth by the quartet theory of emotions. It is not yet immediately clear how much active interoceptive inference alone adds to theories of music-related emotions besides a more detailed description of emotions associated with arousal states. However, once we incorporate QTE as the active mechanism, we will revisit the differing emotional experience of our two concert-goers and the enhanced explanations of their exuberance versus unease, and differing feelings of belonging.

Part 3

<u>Neurofunctional 'Quartet Theory of Emotion' and Predictive Processing</u>

The 'Quartet Thoery of Emotion' is a neurobiological theory proposed by Koelsch et al. (2015) that links four classes of emotion to a quartet of neurobiological affect systems (brainstem-centered, diencephalon-centered, hippocampus-centered, and orbitofrontal-centered), which interact in a dynamical way with biological effector systems (peripheral arousal; action tendencies; motor expression; memory and attention). Reciprocal interactions between these affect systems, effector systems, and conscious appraisal systems (as well as reciprocal interactions within elements of each system) provide a neurological mechanism by which an 'emotion percept' arises and becomes consciously attended to. I posit that these reciprocal interactions consist of predictive processes as *model-updating* features of the system (via the neurological affect systems) and *world-updating* features (via the biological effector systems). In the following section, I will outline each component of this neurofunctional model and demonstrate the role of active inference in each, extending active inference's role from merely interoceptive prediction error minimization in the insular cortex to prediction error minimization across a series of neural and biological structures—the integration of which, in varying combinations and degrees, results in emotional experience. Following this in-depth theoretical outline of the merging of PP and QTE accounts of emotion, I will show how the merging of these two approaches sheds insight into the subpersonal processes underlying the emotional experience of music in part four, while revisiting our concert-goers Rylan and Henry.

Model Updating: QTE Neurobiological Affect Systems

Brain-stem centered In order for active interoceptive inference to work toward the goal of improving physiological homeostasis in various visceral and autonomic systems, these systems must be somehow linked and integrated. At lower levels of processing, these physiological processes are linked and integrated already at the level of the brainstem, as well as the insular and anterior cingulate cortices indicated in predictive processing accounts of emotion. The brainstem is structurally and functionally implicated in relation to the auditory, vestibular, visceral, autonomic and parabrachial nerves. The brainstem and hypothalamus are also in the relevant structural location to receive and incite activation from/to neural pathways corresponding to interoceptive systems—including the insular cortex—and to integrate information from each system to form a cohesive emotional feeling state. These regions generate, modulate, and integrate somatomotor, visceromotor, and neuroendocrine activity essential for survival. As an important center for feelings of arousal, the brainstem regulates homeostatic activity throughout the body, including cardiovascular and hormone activity. Building on the brainstem's integral role in QTE, I posit that the brainstem is the key location for 'model-updating' activity in predictive processing of interoceptive states through tracking and adjusting the parameters of homeostatic 'set-points' given context from different bodily and neural processes. The brainstem is also implicated in 'world-updating' through cuing the motor activity necessary to regulate these processes.

In music-related emotions, we can see brainstem-centered predictive processes at work at electronic dance music concerts, or raves. The DJ holds the crowd in ever building anticipation, adding one new level of instrumentation at a time, building tension through introducing increasing amounts of information (and building up prediction error), before ultimately reintroducing the bass beat and resolving the tension in the music and the prediction error in the brains and bodies of every member of the audience. At each level of increased musical tension, the arousal and anticipation felt by the audience is increased because the amount and rate of deviation from homeostatic arousal norms is increased concurrently with the increasing amount and rate of musical change. This homeostatic arousal state is reset, or updated, as soon as the beat drops and the music and body are brought back in line with the listener's expectations.

Diencephalon centered Much of the work on active interoceptive inference highlights the role of the anterior cingulate cortex and the anterior insular cortex in the integration of prediction and prediction error from specifically interoceptive processes. However, the brainstem, as expressed above, as well as the diencephalon centered systems both integrate information from the body and other brain systems themselves. The diencephalon centered system is associated with the dopaminergic reward system and contains the thalamus, which is associated with perceptual aspects of pain, and the hypothalamus, which is associated with behavioral, autonomic, and endocrine activity and perceptual aspects of pleasure and fun. Information from all of the senses passes through the thalamus, and, given contextual information from the orbitofrontal cortex, becomes associated with affective valence before conscious perception. The hypothalamus processes: (1) homeostatic needs and fulfillment, incentive stimuli (potential to fulfill needs); threatening stimuli, novel stimuli; as well as, (2) input from other affect systems (such as the brainstem, hippocampus, OFC as well as the amygdala, anterior cingulate cortex and anterior insular cortex). These processes within the hypothalamus are important for ensuring the appropriateness of an emotion given the external, environmental context as well as internal bodily context.

I posit that the appropriateness of an emotion in the QTE—or at least of a certain physiological state—is determined by both the amount of prediction error (the total amount of deviation from homeostatic norms), as well as the expected rate of prediction error (the rate at which this deviation occurs, as well as resets). At our rave, the peak emotion state and ultimate release of tension at the drop, or reintroduction of the beat, not only resets homeostatic arousal states but also results in activation of the dopaminergic reward system within the diencephalon-centered affect system as the world and the model, the music and internal bodily state, now match the listener's expectations.

Hippocampus centered Thus far, we have examined how active interoceptive inference occurs at lower-level perceptual processing in maintaining physiological homeostasis. Informed by interaction from the brainstem and diencephalon centered systems, active inference is also occurring at 'higher' cognitive levels of processing. These higher levels of processing extend toward processes such as memory and social behavior, yet are rooted in interoceptive homeostatic aims associated with satiating emotions, or evolutionarily beneficial non-satiating emotions such as attachment-related emotions—both associated with hippocampal activity which will be discussed in more detail shortly.

The hippocampus has dense reciprocal connections with other structures that regulate behaviors essential to survival. These dense reciprocal connections enhance the hippocampus' structural relation to a complex network of emotion systems, and lend structural support to the interaction between memory and emotion. Hippocampal activity is less directly associated with fulfilling immediate homeostatic needs and is more associated with long term attachment related affects, which are implicated in social interaction, sense of belonging, and social cohesion.

I posit that *repeated* social interactions, such as group music making or listening, which *individually* enhance the maintenance of homeostatic needs (as well as positive reward and arousal from the dopaminergic system and brainstem respectively) builds up a predictive model associating higher-level social activity with cascading prediction error minimization down through lower-level physiological homeostatic maintenance. The attachment-related affects which result from this socially cued cascade of PEM results in the positive feelings of belonging. Consider that the rave we've been discussing is actually one of a weekly series of concerts. A group of people have been attending and enjoying these raves every Saturday night for some number of months. Although they don't otherwise know each other, these repeated social interactions, which have individually resulted in various low-level positive affect, have now come to develop in each of the rave-goers a deep sense of attachment and belonging.

Orbitofrontal cortex centered This system most clearly corresponds to the concept of cognitive appraisal; however, the OFC is not a language area. The OFC is responsible for forming concepts and norms, which are 'propositionally not available' or unconscious. The OFC evaluates external and internal stimuli, as well as information from other affect systems for reward/punishment potential and response 'by indicating vegetative, neuroendocrine, behavioral and cognitive programs according to social requirements and social norms' which are learned early in life (Koelsch et al 2015).

Predictive models of physiological homeostasis at lower levels are both contextualized by and contextualize higher level cognitive systems centered on the hippocampus and orbitofrontal cortex. For example, long-term attachment related affects generated by the hippocampus will be associated with predicted levels (or a certain homeostatic range) of dopaminergic and arousal levels in the diencephalon and brainstem centered systems, respectively, which are themselves associated with a specific homeostatic range of interoceptive bodily states. The predictive models maintained by the hippocampus and the OFC will be models developed early in life, and continually developed throughout one's lifespan.

In the arena of music-related emotions, I posit that 'lower'-level predictive models more directly associated with physiological homeostasis (such as the brainstem-centered and diencephalon centered systems) will be more immediately impacted by current external musical cues. However, the habitual expectations associated with these 'higher'-level models will be less susceptible to the immediate effects of musical cues and will be most impacted due to repeated experience with music listening and group music making. Consider our rave-goers on one Saturday when the music of the particular DJ was lackluster and unfulfilling. Although the music that night did not result in the same positive low-level affect, in the form of either homeostatic maintenance or reward activation, the rave goers still maintained a sense of attachment and belonging developed by months of rave attendance. In addition, the rave goers shared a sense of disappointment relating to their music listening experience that they could not quite articulate, as a result of the deviation from the OFC's higher level (non-propositional) concept of a 'good' rave vs. the lackluster event on this night.

World Updating: QTE Neurobiological Effector Systems

The affect systems above concerned localized neurological systems which mediate, interpret, or control other bodily processes and generally constitute 'model-updating' processes. The effector systems are those bodily processes which can perform action in the world; they fulfill the 'active,' 'world-updating' component of active inference, and can act to bring the body and world in line with homeostatic expectations generated by the activity of the neurological affect systems. The four emotional effector systems are: motor systems, peripheral physiological arousal systems, attention systems, and memory systems. Information and action from all four of these systems contextualizes information and action from the other effector systems, and are variously integrated within the neurological affect systems described above. Motor systems govern action tendencies (skeletal and muscular activity related to behavior) as well as the expression of emotion (through facial expression or vocalization). Peripheral physiological arousal systems modulate endocrine activity, vegetative systems (changes in sympathetic/parasympathetic activity), as well as motor and non-motor activity of all organ systems (motor activity including heart activity, breathing, vasoconstriction/dilation; non-motor including immune function, wound healing, energy metabolism). Attention systems can include motor activity such as head turning and eye gaze, as well as non-motor activity concerned with cognitive/psychological attention. And finally, memory systems monitor the selection of information for long and short-term storage, as well as access to that information.

Each of these effector systems perform the actions necessary to minimize prediction error within the neurobiological affect systems. For instance, in our rave example, attention and motor systems are engaged as the listeners attentively wait on input from the DJ, increasing their movement in conjunction with the increased activity of the music until finally jumping in sync with the reintroduction of the bass. The movement of their body then corresponds with the movement of the music, actively minimizing prediction error between their body and the environment. In part four, we will investigate this phenomenon in more detail, as well as how our interlude to the world of electronic music have actually helped us to explain the emotional experience of our heavy metal concert-goers.

Part 4

Enhanced Explanation of Emotional Experience in Music

Audience Spotlight 4

Since we've met Rylan and her grandfather, Henry, we have attributed some of the difference in their emotional responses to the Rammstein concert to differences in exteroceptive expectations related to their individual musical histories, as well as differences in interoceptive expectations related to their experiences of music-associated physiological responses. Henry feels *agitated* both because of his unfamiliarity with Rammstein's music, as well the unfamiliarity with how their music makes him feel. In contrast, Rylan feels *energized* because she has expectations that encompass both Rammstein's music, and how she feels when listening to their music.

Their more nuanced emotional experience of unease or exuberance is impacted by the amount and rate at which any unfamiliarity or expectation changes *over time*. As Rylan's expectations are continually being met for both the musical and internal stimuli, her energized feeling turns to *exuberance*, especially at peak musical moments. As Henry's deviation from physiological homeostasis is continually increasing, his agitation morphs to unease.

As a bonus, Rylan's repeated Rammstein concert attendance reinforces these energized, exuberant feelings and contributes to a sense of belonging within the Rammstein crowd. Henry's repeated Merle Haggard concert attendance has the same effect in the context of country music fans, but he has no such built up, positive, contextual associations for Rammstein concerts. Thus, Rylan is more likely to feel *a sense of belonging* when listening to Rammstein's music and attending their concerts.

I have established that integrating the neurofunctional Quartet Theory of Emotion with the predictive processing accounts of emotion outlined by active interoceptive inference gives us a more clear and nuanced picture of the subpersonal activities underlying emotional experience than any one account alone. Below, we will consider the enhanced explanations provided by integrating these accounts in the case of emotional experience associated with: (a) basic arousal; (b) nuanced emotional states; and, (c) complex social emotions, expanding on the short vignettes of Rylan and Henry.

Basic Arousal

Active interoceptive accounts of emotion rely heavily on the viscerotopic map within the insular cortex, but as I have demonstrated in part three, more attention must be paid to other neural systems implicated in interoceptive and emotional processing. Music, as an exteroceptive cue for interoceptive activity, already affects physiological arousal at the level of the brainstem (Koelsch 2014). The location of the reticular formation, a part of the brainstem which is implicated in maintaining arousal and homeostatic functioning, is at a structurally advantageous location to mediate and integrate information from the cochlear and vestibular nerves. Music, as an auditory event, activates both of these nerves through acoustic signals and affects the movement of fluid within the vestibular system. In fact, this vestibular fluid movement partially explains a drive to move your head along to the beat of a particularly rhythm of bass heavy song (Phillips-Silver 2009). I posit that this movement along to a beat corresponds to the world-changing, active inference of matching physiological vestibular state by activating motor effector systems to move your body or head to the external, musical stimulus.

If you will recall our rave, listeners moved in sync to whatever beat the DJ provided, ensuring that the movement of their body matched the activity in the world (the beat of the music). As Rylan headbangs along with the beat of *Du Hast*, she is actively responding to the energy of the music and (unknowingly) to the movement of vestibular fluid in her ears via coordinated activity between the brainstem-centered affect system and motor effector systems. This minimizes prediction error through matching the movement of fluid in her ears with the expected correlated movement of her body, maintaining an energized emotional state. Henry, however, standing still in staunch rebellion is remaining in a state of unresolved prediction error, maintaining his agitation.

In addition to communication with the vestibular nerve, the reticular formation in the brainstem is in a structurally advantageous location to transmit this auditory phenomenon of music toward regions of visceral and autonomic processing, including the insular cortex, contributing to the experience of goosebumps, or frisson, in peak emotional experience of music.

Nuanced Emotions

While the insular cortex of PP accounts of emotion tracks and integrates specifically interoceptive processes, the brainstem-centered and diencephalon-centered systems of

QTE perform integration of input from a variety of other affect systems.[7] The reward circuit within the diencephalon-centered system is particularly activated in response to peak emotion in music, and is associated with a goose-bumps sensation called 'chills' or 'frisson' (Salimpoor et al. 2011). As part of the peripheral physiological arousal effector system, frisson may occur as a result of the fulfillment, or release, of built up tension within a musical piece. I posit that this occurs because the world (the music) finally matches the expectation of the listener, and the sudden minimization of exteroceptive prediction error, results in the physiological response of goosebumps.

In the middle of *Du Hast*, all of the back-up instrumentation including the heavy beat of the guitar and drums disappears, leaving only the lead singer, a light synthesizer, and a sudden wealth of prediction error for a period of about fifteen seconds. When the beat 'drops' (to use the terminology of our rave example), Rylan and the other Rammstein fans begin again to jump and headbang along to the beat. Rylan experiences chills in reaction to this sudden minimization of exteroceptive prediction error and resetting of homeostatic arousal states at the peak emotional moment of the song. Henry does not have this reaction, perhaps because, when the background instrumentation dropped out, he expected the song to be ending. In this case, the sudden reintroduction of the heavy beat instruments gives Henry an unexpected, and unappreciated surprise, increasing the deviation from his physiological homeostasis.

Henry's agitation, over time, is enhanced through not only an increased amount of prediction error but also the high rate of prediction error as his body continues to deviate from its levels of homeostatic norms, even after peak moments in the music. This persistent prediction error, over time, not only causes him to feel a general negatively valenced arousal state but also contributes to Henry's nuanced emotional experience of unease. In contrast, Rylan's persistent minimization of prediction error, as well as return to an expected homeostatic range after peak moments, causes her to feel not just a positively valenced arousal state, but contributes to her nuanced emotional experience of exuberance.

Complex Social Emotions

Music may affect immediate survival functions through homeostatic autonomic activity and activation of the dopaminergic system (the reward circuit), as outlined in part three concerning the brainstem and diencephalon-centered systems, but it may also

7. Including the brainstem, hippocampus, orbitofrontal cortex, as well as the amygdala, the anterior cingulate cortex, and the anterior insular cortex (Koelsch et al. 2015).

contribute to long term survival functions by enhancing the effect of attachment related emotions through group music making or group music listening.

The habitual predictions of the hippocampus-centered and OFC-centered systems contributes to higher level habitual predictions which are learned early in life. These habitual predictions are associated with the socio-cultural norms to which individuals are exposed early in life, resulting in firmly-established models of such phenomena as the language and music they hear, and social environments in which they interact. If higher-level expectations are met *at the same time* as the lower-level physiological homeostatic expectations are met, then prediction error among these quartets of systems are being maximally minimized, leading to a sense of group belonging.

Rylan has been listening to music like Rammstein since she was in grade school, and she has well-developed habitual predictions surrounding heavy metal music, metal concerts, and metal fans. In addition, the positive effect of their music on her physiological homeostasis and corresponding emotional experiences has been well established. Because prediction error is being minimized at both higher-level conceptual processing, as well as lower-level physiological processing, at this Rammstein concert, Rylan feels as if she is *right where she belongs*.

Henry, however, has not grown up in an environment surrounded by metal music and has developed none of the firmly established models of the social environment surrounding metal music. Neither his higher-level conceptual expectations, nor his lower-level physiological expectations are being met, causing Henry to feel *as if he distinctly doesn't belong* at this Rammstein concert and around these Rammstein fans. Even if it is the case that Henry logically expects a certain social environment at this concert, and has developed a higher-level conceptual model of a metal concert environment, because this higher-level prediction error is not being met *at the same time* as lower-level homeostatic prediction error, he does not have all of the necessary ingredients for a feeling of belonging.

Part 5

Conclusion

Throughout this paper, I have demonstrated that the Quartet Theory of Emotion fills many of the holes left in predictive processing framework, specifically regarding the PP account of emotion. The QTE provides a neurofunctional model detailing which neural and bodily systems are carrying out active inference over which combination of neural and physiological homeostatic processes associated with varying types of emotional experience. This provides a detailed mechanism for active interoceptive inference to contribute to the formation of basic arousal states, nuanced emotional states, and complex social emotions.

Objections

The conscientious reader might draw the objection that although QTE fills in explanatory gaps within predictive processing, what does the PP account of emotion add to QTE? Might the quartet theory itself be a sufficient explanation of emotional experience without appealing to PP? This is ultimately an empirical question with ramifications for the overarching framework of predictive processing itself. If it is found that QTE covers all aspects of emotional experience, with the involvement of perhaps some other predictive process, then it will contribute to identifying at least one boundary on the limits of explanatory power of predictive processing.

However, if QTE does implement predictive processing in each of its sub-systems, then this provides further evidence toward PP as an overarching framework for the embodied brain. Indeed, the QTE as it stands already seems to incorporate the model- and world-updating features inherent in the active inference aspect of PP via the activity of the affect and effector systems respectively. These systems in QTE maintain lower-level physiological homeostasis as well as higher-level conceptual expectations involved in forming the varying levels of emotional experience—which is exactly what a full-fledged predictive processing account of emotions aims to describe. It is not only parsimonious to integrate PP and QTE approaches to explain emotional experience but also appealing to the PP framework provide a seamless integration of the personal-level phenomenological experience with the sub-personal level of the underlying neural and bodily processes, as demonstrated through the running vignette of Rylan and Henry and the personal and sub-personal differences in their emotional experiences while attending a Rammstein concert.

In this paper, I have discussed how the predictive processing framework expands upon traditional, bottom-up, two-factor theories of emotion as passive physiological evidence-building and subsequent cognitive appraisal (James, 1884; Lange, 1885; Schachter & Singer 1962) by incorporating active inference based on predictive models of the causes of external *and* internal stimuli (Seth & Critchley 2013; Seth & Friston 2016). Thus, emotional content from music is evoked as a result of active exteroceptive inference related to the physical musical stimuli (based on statistical regularities of the current musical event and past experience) as well as active interoceptive inference regarding the listener's current autonomic, and physiological states. I have demonstrated that this is a productive route toward explaining the differing feelings of our concert-goers, Rylan and Henry. In addition, I have proposed that this general predictive processing framework is implemented through a 'quartet' of neurofunctional mechanisms (Koelsch et al 2015), which are dynamically implicated in the neural and physiological processes underlying general and music-related emotional experience.

Conceptualising emotion as active inference over both external and internal processes, implemented and maintained through a dynamically interacting subset of neural pathways offered by QTE, provides both a mechanism by which music evokes emotion and results in the subjective difference in the reported emotional experience of music between individuals, based on varying levels of neural interactions which arise from different prior experiences informing both their exteroceptive and interoceptive predictive models.

Avenues for Future Research

A more complete theory of the neurofunctional and physiological processes involved in emotional experience, such as that offered by the integration of the quartet theory of emotion and predictive processing, can help to solidify the boundaries, or demonstrate the continuity of emotional and cognitive processes. In addition, by expanding the QTE+PP model to incorporate distinctly social activities such as group music making, we can enhance our understanding of how social interactions affect both emotion and cognition in an increasingly second-person neuroscience. That is, a neuroscience that takes into account the inherently social and interactionist nature of the mental and behavioral activity of its participants, rather than merely testing how participants *simulate* or *theorize* regarding the intentions and activity of other minds.[8] Emphasis on interoceptive processes, and the feelings of belonging in group musical experience can further develop

8. A second-person neuroscience assumes greater influence of the Interaction Theory of Mind, understanding that others are experienced as a subject through mutual emotional and embodied engagement. In contrast, current first-person and third-person approaches to neuroscience assume that individuals take other minds as objects, either simulating the intentions and actions of other minds through a mentalizing or mirror neuron network (Simulation Theory of Mind), or undergoing a complex series of inferences (Theory-Theory of Mind). The experimental paradigms attached to these first- and third-person accounts leave no room to demonstrate the importance of interaction in social cognition (see Schilbach et al. 2013 for further discussion).

research into the concept of the self and the extended self, and whether selfhood lies within our body or brain or emerges through interaction in our social environment.

References

Clark, Andy. 2013. "Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science." *Behavioral and Brain Sciences* 36: 181-253.

Clark, A. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford, United Kingdom: Oxford University Press.

Huron, D. 2006. *Sweet Anticipation*. Cambridge, MA: MIT Press.

Frijda, N. H. 1993. "The place of Appraisal in Emotion." *Cognition and Emotion* 7 (3/4): 357–387.

Gerrans, P. (Presenter). 2017, June 29. *Emotions and Subjective Perspective*. Speech presented at The Human Mind Conference, The Møller Center, Cambridge, United Kingdom.

Hohwy, J. 2013. *The Predictive Mind*. Oxford, United Kingdom: Oxford University Press.

James, W. 1884. "What Is an Emotion?" *Mind 9*: 188-205.

Juslin, P. N., & Västfjäll,D. 2008. "Emotional Responses to Music: The Need to Consider Underlying Mechanisms." *Behavioral and Brain Sciences* 31: 559–575.

Koelsch, S. 2013. "Emotion." In *Brain and Music*, 203-240. Hoboken, NJ: Wiley-Blackwell.

Koelsch, S. 2014. "Brain Correlates of Music-Evoked Emotion." *Nature Reviews Neuroscience* 15: 170–180.

Koelsch, S. et al. 2015. "The Quartet Theory of Human Emotions: An Integrative and Neurofunctional Model." *Physics of Life Review* 13: 1–27.

Lange, C. G. 1885. "The Mechanism of the Emotions." In *The Emotions*, edited by D. Dunlap, 33–92. Baltimore, MD: Williams & Wilkins.

Meyer, L. B. 1956. *Emotion and Meaning in Music*. Chicago, IL: The University of Chicago Press.

Narmour, E. 1990. *The Analysis and Cognition of Basic Melodic Structures: The Implication Realization Model*. Chicago, IL: University of Chicago Press.

Phillips-Silver, J. 2009. "On the Meaning of Movement in Music, Development, and the Brain." *Contemporary Music Review* 28 (3): 293–314.

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. 2011. "Anatomically Distinct Dopamine Release During Anticipation and Experience of Peak Emotion in Music. *Nature Neuroscience* 14: 257–262.

Schachter, S., & Singer, J. E. 1962. "Cognitive, Social, and Physiological Determinants of Emotional State." *Psychological Review* 69: 379–399.

Scherer, K. R. 1999. "On the Sequential Nature of Appraisal Processes: Indirect Evidence from a Recognition Task." *Cognition and Emotion* 13: 763-793.

Scherer, K. R. 2009. "The Dynamic Architecture of Emotion: Evidence for the Component Process Model." *Cognition and Emotion* 23 (7): 1307–1351.

Scherer, K. R., & Zentner, M. R. 2001. "Emotional Effects of Music: Production Rules." In *Music and Emotion: Theory and Research*, edited by P. N. Juslin & J. A. Sloboda, 361–392. New York: Oxford University Press.

Schilback, L., et al. 2013. "Toward a Second-Person Neuroscience." *Behavioral and Brain Sciences* 36: 393–462.

Sel, A. 2014. Predictive Codes of Interoception, Emotion, and the Self [Review *Interoceptive inference, emotion, and the embodied self*, by A. K. Seth]. *Frontiers in Psychology* 5: 1-2.

Seth, A. K. 2013. "Interoceptive Inference, Emotion, and the Embodied Self." *Trends in Cognitive Sciences* 17 (11): 565–573.

Seth, A. K., & Critchley, H. D. 2013. Extending predictive processing to the body: Emotion as interoceptive inference [Review *Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science*]. *Behavioral and Brain Sciences* 36: 47–48.

Seth, A. K., & Friston, K. J. 2016. "Active Interoceptive Inference and the Emotional Brain." *Philosophical Transactions B* 371 (1708).

Van de Cruys, S. 2017. Affective Value in the Predictive Mind. In *Philosophy and Predictive Processing*, edited by T. Metzinger & W. Weise. Frankfut am Main: MIND Group.

Vuust, P., & Frith, C. D. 2008. "Anticipation is the Key to Understanding Music and the Effects of Music on Emotion [Review *Emotional Responses to Music: The Need to Consider Underlying Mechanisms*, by P. N. Juslin & D. Vastfjall]." *Behavioral and Brain Sciences* 31: 599–600.