Mammals and Music among Others: Crossmodal Perception and Musical Expressiveness

W. P. Seeley

Bates College

Word Count:
2897 (text)
3764 (w/ notes and references)

Author's Note: The following is a draft copy of a paper submitted for consideration for The Philosophy of Perception and Aesthetics Conference, University of Antwerp, December, 2012. Please do not cite without permission of the author. Please address all correspondence to W. P. Seeley, Department of Philosophy, Bates College, Lewiston, ME, 04240, USA (wseeley@bates.edu).
Abstract:
The paradox of musical expressiveness can be interpreted as a question about the nature of artistic communication in music. We express an emotion when we publically display it in our behavior. Behavioral expressiveness is thereby one side of a communicative exchange whose function is, at least in part, to help integrate and coordinate the activities of individuals in social contexts -- it is a subconscious or subliminal means for an agent to reveal emotions indicative of their current beliefs, desires, and action tendencies to others. Musical works are inanimate artifacts. They don't have beliefs. They don't have desires. They don't have emotions. They don't have behaviors. Nor do listeners ordinarily believe they do. Therefore the exchange between music and a listener is not adequately structured to support an expressive communicative event. Yet, nonetheless, we regularly recognize music as expressive of emotions and are sometimes even aroused to experience those emotions. Thus the paradox. In this paper I propose an integrated crossmodal model for our engagement with expressive music derived from a biased competition model for selective attention and a diagnostic recognition framework for perception and sketch its potential impact on the discussion of issues associated with the paradox of musical expressiveness.
The paradox of musical expressiveness for pure instrumental music can be interpreted as a question about the nature of artistic communication in music.¹ We, as human agents, express an emotion when we publically display it in our behavior, when our postures, gestures, and actions are expressive of our current, emotion laden, mental states. Behavioral expressiveness is a communicative event, or better, one side of a communicative exchange whose function is, at least in part, to help integrate and coordinate the activities of individuals in social contexts -- it is a subconscious or subliminal means for an agent to reveal emotions indicative of their current beliefs, desires, and action tendencies to others (Robinson, 2007, Smyth 1984). Musical works are inanimate artifacts. They don't have beliefs. They don't have desires. They don't have emotions. They don't have behaviors. Nor do listeners ordinarily perceptually categorize musical works as animate agents with beliefs, desires, or action tendencies. Therefore the cognitive exchange between music and the listener is not adequately structured to support an expressive communicative event. Yet, nonetheless, we regularly experience music as expressive, and are sometimes even aroused to the emotions we thereby encounter, just as we do and are in ordinary social interactions with other mammals like ourselves. Thus the paradox.

A range of solutions have been proposed to this putative problem. The most obvious is to reorient the communicative exchange. We could think of pure music as itself a type of behavior, as a behavioral expression of the emotional states of some agent, either the composer, a performer interpreting the work, or a fictional persona. Certainly pure music expresses the emotions of the composer or the performer some of the time. However, successful composers and performers are professionals with technical expertise in their craft. This expertise includes an explicit analytic sense of what it takes to put together and perform a successful piece -- a skill that enables composers and performers to deploy expressive strategies independent of their own emotional engagement with the music. So, even if a capacity to "feel the music" plays a significant role in their productive practices, it is unrealistic to imagine that they are regularly expressing their own occurrent emotions. It is, of course, possible that the intention of a composer or performer is that listeners approach a work as if it were expressive in this way, as if it were expressive of the emotions of some imagined fictional creative persona. Surely this is the case in some contexts, but, again, likely not all. For instance, the abstract compositions of Stravinsky and Bird Songs of the Mesozoic are awash in expressiveness but they seem to lack any center of focus or thematic continuity that would denote a unified fictional persona.² Furthermore, even if this solution is generally plausible in some contexts, it belies a bigger problem. How is music a behavioral expression of the emotion of some real or fictional persona? How is it an extension of the human bodily capacity to display an emotion? What are the perceptual cues that could lead to such an interpretation? What are listeners responding to in expressive pure music.

My primary interest in this paper is the relationship between contour theory and embodied appraisal accounts of musical expressiveness (Kivy, 2002; Prinz, 2004; Robinson, 2005, 2007). Contour theory provides an account of the formal and compositional features, the perceptual cues, that composers

1. Other types of music are more explicitly associated with characters and narrative content sufficient to account for their expressive qualities, e.g. the adolescent romantic laments of popular rock and roll musicians or Prokofiev's Peter and the Wolf.

2. I saw a spot on the TV news magazine program Nightline recently in which Grand Master Flash reported that he was in awe of Stravinsky as a young man because the music was so complexly layered… and so scary!
and performers use to express emotions in music: aspects of the musical structure of expressive works resemble the dynamic structure of expressive behaviors and movements, including in some cases the dynamic contour of the subjective feeling of the emotions expressed. So the central claim in contour theory is that musical expressiveness depends on a recognizable resemblance between the feelings associated with emotions, the behaviors and movements used to express them, and the compositional structure of musical works, e.g. the interplay of tension and resolution in the rhythmic, harmonic, and tonal structure of contemporary blues compositions and classical sonatas. This is a cognitivist view. Contour theorists argue that we perceptually recognize the emotions expressed in music, but we do not thereby feel them. We may feel moved by the compositional complexity or beauty of the music. But it is never the case on this account that sad or joyous music moves us to sadness or enthusiastic joy.

Embodied appraisal theories provide a model for the psychological mechanisms that underwrite the experience of musical expressiveness: the dynamic contour of expressive music is sufficient to automatically and directly induce a range of autonomic responses in listeners, an interplay of tension and release, that resemble the unfolding dynamic contours of the experience of different types of emotions. Listeners recognize the expressive qualities of the music in much the same way they become aware of their emotional responses to ordinary environmental stimuli, through a process of cognitive monitoring that modulates and reinforces associated autonomic responses. The claim here is that quick and dirty automatic appraisals of the behavioral significance of stimulus in ordinary contexts are realized in autonomic responses that encode the value of that stimulus to our behavioral goals, and that we consciously experience these autonomic responses as emotions. In musical contexts these same patterns of autonomic response suffice to enable perceivers to categorize, and thereby recognize, works as expressive of a particular types of emotions. It is an open question whether the experience of these recognizably stereotyped patterns of low level autonomic responses are sufficient to arouse a full blown emotion in musical contexts. Nonetheless, given that they have a readymade mechanism in place to explain the arousal of emotions in response to expressive music, they potentially have more explanatory power than contour theory.

Contour theory and embodied appraisal theories are not the only other games in town. However, the approach here is promising. My intuition is that once we have a better sense of the perceptual cues and psychological mechanisms that mediate the expressive communicative exchange between musical works and listeners the paradox of musical expressiveness will seem less daunting. This is not to say that psychological facts about our engagement with music will dissolve the paradox. Nor is it to say that they will make the issue less interesting, e.g. a banal artifact of perceptual processing. Rather, my argument is that we can fruitfully treat questions about musical expressiveness as questions about the ways listeners acquire, represent, manipulate, and use information encoded in the musical works -- as questions about the range of computational processes that underwrite behavioral responses to expressive music. This kind of an exploration can help to determine which musical features contribute to the expressive content of a work, help provide a more adequate account our standard responses to the expressive qualities of pure music, and thereby help winnow the range of plausible theoretical solutions to the paradox. In the second half of this paper I sketch a model for musical expressiveness derived from a biased competition model for selective attention and a diagnostic recognition framework for music perception (Desimone and Duncan, 1995; Pessoa, Kastner, and Ungerleider, 2002; Duncan and Barrett, 2007; and Schyns, 1998). The central claim of this model is that a crossmodal attentional network which integrates auditory, sensorimotor, and affective information underwrites our engagement with artworks in a range of media and transparently explains the capacity of musical works to communicate expressive qualities (see Carroll
and Seeley, in press; Seeley, 2012; and Carroll, Moore, and Seeley, 2012). I argue that this model collapses the distinction between cognitivist theories like contour theory and embodied appraisal accounts of musical expressiveness and briefly discuss the implications of the approach for some issues in the philosophy of music.

Can a musical work, an auditory stimulus, sound like the appearance of an expressive bodily gesture, a visual stimulus? What could underwrite such a resemblance? Psychologists of both dance (Krumhansl, 1997; Krumhansl and Schenck, 1997) and music (Vines, Krumhansl, Wanderly, and Levitin, 2006; Chapados and Levitin, 2008) have applied similar methodologies to explore this question. In this range of studies participants were divided into auditory only (AO), visual only (VO), and audio-visual (AV) groups. The AO group listened to a sound recording of either the music the dance was choreographed to (the target of the dance study was Balanchine's ballet for Mozart's Divertimento No. 15) or the performance of the piece of pure music (Stravinsky's Three Pieces for Clarinet Solo). The VO group watched a soundless video of either the target dance performance or the performance of a piece of pure music. The AV group watched the same video with the associated sound accompaniment. Participants were asked to make continuous judgments of the strength of either felt tension or emotion using a foot pedal or a manual slider. Continuous measures of the ebb and flow of tension and emotion were strongly correlated across all three experimental conditions in each of the studies (regardless of whether a foot pedal or a manual slider was used). Further, AV responses could be modeled and predicted as a function of the additive combination of AO and VO measures. Skin conductance, an autonomic response canonically associated with embodied appraisals and the feeling of an emotion, was strongly correlated with continuous measures of perceived tension in all three conditions (Chapados and Levitin, 2008).3 Finally, verbal reports of the range and strength of the emotions expressed in the three stimuli were also strongly correlated (Krumhansl and Schenck, 1997).4 The results of these studies demonstrate that there is a genuine sense in which musical passages can resemble the expressive qualities of bodily gestures, even in the abstract contexts of contemporary dance and pure music. Further, correlations between the continuous measures of perceived tension and emotion suggest, consistent with contour theory and embodied accounts, that what underwrites this resemblance is a shared dynamic contour, a stereotyped pattern of tension and release that is present in both the movements and the music.

I suggest that the ebb and flow of arousal in our autonomic responses to perceived tension and release in a musical passage is a diagnostic cue to its expressive qualities. Diagnostic cues can be defined as minimal sets of perceptual cues that are sufficient to enable an organism to perceptually recognize the shapes, identities, and affordances of objects, events, and actions, e.g. the efficient use of line and gesture in caricature and editorial cartoons. In ordinary contexts diagnostic cues represent an adaptive strategy, a means to enhance the computational efficiency of limited capacity perceptual systems in a replete and noisy environment. Minimal sets of diagnostic cues are continuously matched to general declarative knowledge of the structure and function of object, event, and action types. The match between diagnostic

---

3. Krumhansl (1997) reports that verbal reports of the experience of different types of musically expressive properties in music are strongly correlated with different stereotyped patterns of autonomic responses.

4. See Krumhansl and Schenck (1997). Verbal reports were based on a post experiment questionnaire asking participants to rate the strength of the emotions they felt on a scale of 0-8. The emotions queried were: afraid, amused, angry, anxious, contemptuous, contented, disgusted, embarrassed, happy, interested, relieved, sad, and surprised. One interesting thing to note is that the general correlation demonstrated between perceived tension and the qualitative strength of emotional responses is so stale that the former is often used alone in these kinds of studies an index of the expressive quality of music.
cues and categorical knowledge often suffices for object recognition and can be used to both direct attention and orient our bodies to features of the local environment salient to our current behavioral goals. Biased competition models of selective attention provide a mechanisms to realize a diagnostic recognition framework for perception. On these accounts fronto-parietal and cortico-thalamic networks bias the firing rates of populations of neurons within, and facilitate crossmodal integration among, unimodal perceptual processing systems. These top-down feedback projections enhance the firing rates of populations of neurons that would encode the presence of expected features, inhibit the perception of distracting information, and prime the skeletomuscular system to anticipated changes in the local environment related to our own actions and the behaviors of others. These processes enhance the computational efficiency of visual recognition and facilitate the coupling of perception and action by priming perceptual systems to the presence of behaviorally salient features at expected locations.

The critical aspect of this model for present purposes is connectivity between affective attentional networks and perceptual systems. Affective perceptual processing has traditionally been associated with automatic processes mediated by direct connectivity between the sensory periphery and the amygdala. This so-called low road pathway is hypothesized to help direct attention to affectively salient behavioral stimuli independent of cognitive contributions from higher cortical processing. However, amygdala receives direct input from cortical processing areas associated with abstract pattern recognition in visual and auditory recognition, and top down influences in visual and auditory processing are integrated into low road processing via direct connectivity between areas of both the visual and auditory thalamus and amygdala. Amygdala is also reciprocally connected with orbitofrontal cortex (OFC), ventromedial prefrontal cortex (vmPFC), and anterior cingulate cortex (ACC). This cortico-fugal attentional circuit is associated with evaluations of the behavioral significance of (OFC), and gut emotional reactions (visceromotor control) to (vmPFC/ACC), behaviorally significant stimuli. Current evidence suggests that these attentional resources are necessary for affective perceptual responses. Covertly attending to neutral stimuli in a high attentional load task eliminates amygdala responses to foveated emotional faces at the fixation point, to affectively charged stimuli that participants are staring at in the center of their gaze (Pessoa et al, 2002). The conjunction of these results entails that cognitive information concerning the identity, behavioral utility, and affective significance of a stimulus directly influences the firing rates of populations of neurons in the amygdala responsible for our quick and dirty initial embodied appraisals of emotionally charged stimuli. Further, and perhaps more importantly for this discussion, feedback projections from OFC and amygdala directly influence the encoding of emotionally salient behavioral features in auditory and visual sensory areas. My proposal is that these integrated crossmodal attentional circuits constitute a set of general purpose perceptual strategies for generating and maintaining unified perceptual representations of the semantic, behavioral, and emotional significance of perceived stimuli -- perceptual representations that facilitate our capacity recognizing the expressive contours embedded in the rhythmic, harmonic, and tonal structure of works of pure instrumental music.

There is a sense in which this model collapses the distinction between cognitivist and embodied appraisal accounts of musical expressiveness. On the one hand, the unfolding temporal dynamics of a musical work is a complex perceptual cue that is diagnostic for its expressive content. In this regard listeners perceptually recognize the expressive content of the work in its rhythmic, harmonic, and tonal contour independent of the explicit arousal of an emotion. However, on the other hand, our capacity to recognize these expressive features in the music depends upon the top-down influence of affective attentional processes in perception -- processes responsible for both calculating the behavioral significance of a stimulus in ordinary contexts and our gut reactions to them. The net result of these latter...
processes is an embodied appraisal, an automatic (but not necessarily direct) somato-visceral response that is constitutive of the valenced affective profile we assign to the stimulus. The trick is that the process doesn't stop here. These processes feedback into the broad set of crossmodally integrated attentional circuits identified above and trigger higher level cortical processes that further modulate our perceptual and emotional responses, influencing how we hear the music and how it feels to hear it that way. Therefore there is a sense in which our cognitive and embodied responses to valenced emotional stimuli are inextricably intermeshed, so that our cognitive capacity to perceptually recognize expressive qualities in the structural dynamics of a musical work cannot be peeled apart from the contour of our affective response to it.

Where does this leave the discussion of musical expressiveness? First, the role diagnostic features play in perceptual recognition suggests that there is nothing odd about the communicative exchange that underwrites a listener's experience of the expressive quality of a piece of music. Expressive contours are part of the formal/compositional toolkit that composers and performers employ to communicate the emotional content of the music. These contours are robust structural features of the music. We ordinarily encounter, hear, and experience them in, and as belonging to, the music, as opposed to any associated real or fictional persona. Of course we do sometimes experience music as expressive of the emotions of a person or persona. An advantage of the model is that it can accommodate these cases. The diagnosticity of a perceptual cue is partially determined by the context in which we perceive it. Narrative cues and knowledge about a particular composer or piece can alter the diagnosticity of expressive cues in the music, rendering them expressive of the emotional states of a particular person or fictional character. Second, recognizing musical expressiveness does not necessarily entail the arousal of any emotion in the perceiver, nor does it entail an explicit conscious experience of the autonomic responses associated with embodied appraisals necessarily play a role in the experience of musical expressiveness. The latter are perceptual processes which encode affective information that influences perceptual recognition regardless of whether we consequently explicitly focus our attention on them. Nonetheless, via the integration of affective, auditory, and somato-visceral attentional circuits, the model has the structure to accommodate cases where arousal is coupled with the experience of musical expressiveness, and can explain how the arousal of an emotion can contribute to a listener's capacity to recognize the expressive quality of music. Finally, our emotional responses to music are not instances of imagination or make-believe per se, at least not ordinarily. They are genuine cognitive and emotional responses to the expressive contour of the music. However, the reciprocal cortico-fugal attentional circuits that drive the model have the cognitive structure to account for cases where explicit imagination does influence the experience of musical expressiveness, e.g. where a listener's response to a piece of narrative or abstract music is a directly influenced by their imagined engagement with a fictional character, event, or world.
References:


