Musical Meter, Social Cognition, and Musical Expression:
An Inquiry in Cognitive Aesthetics

Justin London (Carleton College)

Invitation to a march (with apologies to Arthur Laurents)

Consider a march—Sousa’s *Washington Post March*, to be precise:

![Washington Post March Sheet Music](image)

After the stirring introduction, the main strain begins, and all parts, from the lead cornet to the second trombone, beg us to move along with them. Indeed, it is hard not to tap your toe to the music. Almost all music invites movement like this, of course, but if a piece has a steady beat and is lively and quick, it is much more likely to bring about this reaction. Moreover, such music invites not just our individual engagement, but the collective, coordinated engagement of the entire audience. Marches in general do this particularly well. When we listen to *Washington Post March*, whether in a concert hall or at a parade, we are hearing in time (and often, as a result, moving in time), together.

William H. McNeill, an historian who has written on Toynbee, technology and military history, and on the rise of Western civilization, also wrote a fascinating book on a rather different topic. *Keeping Together in Time*, inspired in part by McNeill’s own experiences on the parade grounds of World War II, is a study of dance and drill in human history. In it, he develops his theory of “Muscular Bonding”—the collective enactment of a shared rhythm. McNeill takes special note of the social aspect of such coordinated action, as it:

constitutes an indefinitely expansible basis for social cohesion among any and every group that keeps together in time, moving big muscles together and chanting, singing, or shouting rhythmically.¹
Muscular bonding, in addition, has particular emotional and affective properties. It is a diffused emotional response, not based on the encounter with a particular stimulus. Instead, moving in time can “make us feel good about ourselves . . . and vaguely pleased with the world at large.” This explains, perhaps, how ubiquitous our predilection is for moving together in time:

American football crowds, South African demonstrators, patriotic parades, and religious rituals of every description all draw on the emotional affect of rhythmic movements and gestures.

Current research in social cognition, especially work in joint attention and emotional contagion, broadly supports McNeill’s introspective reports regarding the affective qualities of collective movement. As a musician and music theorist interested in the psychology of rhythm, I would go further to posit that overt movement is not a precondition of the sort of collective affect McNeill describes (although it is often present when we tap our toes and sway our heads). Simply hearing, attending, and internalizing rhythms without overt movement can have a similar effect. This is what music does to an audience; it gives us a visceral sense of a shared, collective emotional and aesthetic experience. With that in mind, I want to examine two questions. First, exactly how does music engender our shared, collective aesthetic experience of it (and specifically in ways that our experiences of painting, sculpture, and perhaps even drama do not)? Second, why is it that some music fails to do so?

In the pages that follow, I will discuss meter as a special case of entrainment, that is, our general capacity for rhythmic synchronization and sensory-motor coordination. Metrical entrainment provides an answer (or at least part of an answer) to the question “how does music get us to move together in time?” Metric ambiguity, on the other hand, would seem to be the very antithesis of entrainment. Here I will define it as a partial inhibition of our capacity for entrainment. Next, I will give a detailed case study of metrical ambiguity as inhibited entrainment, the opening of the second tableau of Stravinsky’s Svadebka (Les Noces). The broader implications of metric ambiguity will then be discussed, drawing on research in joint action, the study of how the presence of another person or persons in the environment changes the way we perceive and react to objects in our perceptual field. Along with joint action, the social aspects of our emotional responses will be considered, especially in relation to emotional “contagion,” whereby an emotional response is broadly and rapidly shared among members of a group. Metrical ambiguity serves to attenuate or impede both joint action and our shared emotional response to the music. Two related arguments will conclude this essay. First, that music’s expressive properties are bound up with all these aspects of our collective rhythmic engagement with the music in a social context. Second, what might be regarded as the aesthetic failure of much 20th-century music may well be rooted in its inability to stimulate the kind of collective social experience that has been the lifeblood of music throughout its history.
**Meter and metric ambiguity**

Musical meter is commonly defined as a regular pattern of accented and unaccented beats, typically in cycles of two or three. To play or hear metrically requires that we can generate such patterns; in dynamic systems terms, it means that we are able to produce a self-sustaining pattern of oscillation. Entrainment (also referred to sometimes as attunement) involves the coordination of rhythmic activity between two or more oscillatory rhythmic systems. These systems could be two coupled mechanical processes, a mechanical process and a biological process, or two or more biological processes. Technically, entrainment obtains when there is any phase locking between two (or more) self-sustaining oscillatory/periodic systems. Phase locking is achieved when the actions of the two oscillators are coupled in some fashion, and their oscillation periods afford it. The classic test for entrainment involves perturbing the phase or period of one oscillator and observing the reaction of the other; in a bona fide case of entrainment, mutual adaptation will occur.

When entrainment occurs, one can often distinguish between the *driving* oscillator and a *driven* oscillator. In addition, there may be more than one driving or more than one driven oscillator—a single periodic rhythm may trigger a set of oscillators that characterize a complex dynamic system or a population of individual oscillators. Entrainment is most often described in terms of a 1:1 relationship between the driving and driven oscillators, but it can involve other ratios, too (e.g., 1:2, 1:3, etc.).

Entrainment is obviously involved when we observe coordination of a motor behavior with an external, periodic stimulus (e.g., tapping along to the beat of a metronome). However, as Mari Jones and her colleagues have shown, entrainment also involves our attention apart from (but, of course, closely related to) our motor behavior. Indeed, coordinated motor behavior presumes some coordination of attention, both for monitoring the external rhythm and for self-monitoring one’s synchronized action. Thus entrainment not only guides how we interact with the world, but also our perception of it; we perceive the objects of our entrainment differently, with heightened attention to both those objects and to ourselves. This may explain in part McNeill’s observation that keeping together in time “makes us feel good about ourselves . . . and vaguely pleased with the world at large.”

Human rhythmic entrainment occurs within a particular temporal range, from one tenth of a second, or 100ms, to about two seconds, though more complex sequences can be longer, up to about 5-6 seconds. Within this range, there are several sub-ranges, which produce differences in our entrainment characteristics (100-400ms, 400ms to 1 second, and 1-2 seconds). As a result, our rhythmic perception and performance is quantitatively and qualitatively different in these ranges. For coordination of human action/interaction, entrainment most crucially involves coordination of the higher-level component periodicities of a hierarchically-organized set of rhythms (those in the range of 1–3 seconds). These periodicities provide the temporal frameworks for more complex
motor/action sequences (one has to know, for example, “when to begin” in the production and coordination of a more complex gesture or action). Indeed, in many instances it is clear that faster/lower-level actions are not consciously timed, though the events that initiate and/or terminate these action sequences are timed. Although entrainment may be observed on various levels, it is the synchronization on higher levels that gives rise to coordinated social behaviors.

It must be emphasized that rhythmic perception and entrainment are not passive aspects of our perception, but are bound up with action and movement. As various studies have shown, when we listen to a rhythmic stimulus, both the auditory and the movement-related areas of the brain, especially the cerebellum, basal ganglia, and supplementary motor areas show coordinated activation. As Katie Overy and Istvan Molnar-Szakacs have noted, “music is perceived not only as an auditory signal, but also as intentional, hierarchically organized sequences of expressive motor acts behind the signal; and that the human mirror neuron system allows for corepresentation and sharing of a musical experience between agent and listener.” As we shall see, this co-representation is crucial to joint action and to the affective responses that arise when we experience music together.

Meter is entrainment in the special context of performing or listening to music. As entrainment more generally guides our coordinated actions and perceptions, meter serves as the temporal framework that guides our listening and performance. For the audience, the rhythms produced by the musician(s) serve as the driving oscillators, and its toe-tapping responses are the driven oscillations made manifest. For a group of performers, entrainment is more supple and complex, as it involves different musical and social roles within the ensemble. Even solo performers create feedback between their behaviors and the sounds they produce. Meter is subject to the same perceptual and cognitive limits as is entrainment more generally. It is hierarchically structured, with different performance and perceptual attributes accruing to different structural levels (i.e., subdivisions, beats, and measures). Musical entrainment crucially involves the coordination of entire metric cycles between performer(s) and listeners(s) by aligning their sense of the downbeats as well as the structure of the overall beat pattern. For listeners, entrainment is a hallmark of an engaged experience of the music, as entrainment allows one to “move with the music” either virtually or actually, covertly or overtly.

In both musical and more general contexts, entrainment enables temporal coordination. Of necessity, keeping together in time entails two processes: (a) producing (or imagining) movement—a gait, a gesture, a musical phrase; and (b) coordinating those movements. For musical performers this involves keeping their metric entrainments in phase (i.e., having the same sense of the downbeat) so that their complex motor actions (e.g., producing a complete rhythmic figure) may be produced in synchronization with others. For listeners, keeping their sense of meter in phase enables them to accurately predicted how and when the music will unfold in time.

We are now ready to define metric ambiguity. From the perspective of meter-as-entrainment, metric ambiguity requires a temporal structure that is still metric, while simultaneously inhibiting entrainment to some extent. Such temporal structures characteristically incorporate some aspects that can be regarded as regular or periodic
(with the proviso that these rhythms occur within the range of human rhythmic perception and cognition noted above). Given that the beat level is the necessary substrate for any metric structure, metric ambiguity cannot occur in the absence of a steady beat or tactus. Indeed, if we do have counter-stresses that threaten to undermine our sense of the beat or its location, the result is syncopation, or a wholesale shift of the beat. Likewise, metrical ambiguity does not involve the relative stability of sub-tactus levels of meter, as these levels often come and go and may shift in organization (as triplets momentarily replace duplets, or running eights are further divided into sixteenths, and so forth). Thus metric ambiguity involves problems of organization on higher levels and is often described in terms of accentual conflict, which undermines the established length of the measure, the location of the downbeat, or both. And if “metric entrainment” characteristically involves entire metrical cycles, then it follows that inhibiting such entrainment can lead to the inhibition of larger-scale periodicities while maintaining the continuity of lower levels of structure. This is precisely what happens in the opening measures of the second tableau of Stravinsky’s Svadebka.

**Stravinsky’s hiccups**

Svadebka or Les Noces (“The Wedding”) is the tale of a Russian peasant wedding --or such a wedding as Stravinsky may have imagined it--portrayed in four tableaux. The second tableau takes place at the house of the groom, an important station within the traditional Russian peasant wedding ritual. At the start of this tableau, the groom’s hair is being combed in advance of his wedding (a similar ritual took place with the bride in the first tableau). Example 2 provides a melodic reduction of the opening choral melody:

As Margarita Mazo suggests, the opening of this scene is built upon a popevka, a melodic-rhythmic motive commonly found in Russian folk music. This popevka figure consists in a series of five eighth notes followed by an eighth rest, creating a gap which marks the end of the figure. The means of marking a group boundary used here is common to both music and language. Closure within this unit is reinforced by the pitch pattern: D-D-D-C-D, in which the last element returns to the main pitch. Each time the D-D-D-C pattern begins, the initial D is marked by a crisp four-stroke roll on the snare drum. Rhythmic boundaries in this passage are also marked by texture (soloists with chorus, as opposed to chorus alone), dynamic accent, and the rhyme scheme:
Pree-chee-stah-ya *mat*,
Kha-dee, kha-dee k nam u *khat*,
Sva-khe pa-mah-*gat*,
Ku-dri ras-chee-*sat*.

Translation:

Pure mother,
Come, come to our hut,
To help the matchmaker
To comb out the curls,
Khvetis’s curls, Pamfil’evich’s light brown curls.\(^{15}\)

The words underscored in the translation correspond to those “interjections” of the soloists and chorus, and this textural thickening occurs when the groom’s name is called out by his buddies. Thus the beginnings and endings of the *popevka* figure are marked in multiple parameters, with strong associations between metric position within each parameter (e.g., the drum roll always occurs on a downbeat). The *popevka* figure projects a sense of 3/4 without any anacrases. In other words, it is perceived as a dactylic rhythmic figure that is accentually in phase with a ternary meter.

After its first presentation, the *popevka* motive immediately repeats, or at least seems to, with the first four notes the same: D-D-D-C with its onset marked by the drum roll. At that moment, therefore, we might infer that a simple triple meter will ensue, but the return to D is delayed by a leap up to F and then a descent back down, creating the extended figure: D-D-D-C-F-E-D. The return to D is only a weak indication that this is the end of the figure, but the next note has a dynamic accent and the drum roll, giving it a strong sense of motivic onset. This assists us in hearing the next five notes as a unit: D-C-F-E-D, followed by an eighth note rest, and we hear the unit as a repeat of the end of the previous figure (note the segments bracketed in Ex. 2). We hear this organization even though the repetition of the initial D obscures the motivic boundary. The process might be characterized as a metrical “hiccup.” This is a direct musical analogue to a common type of stutter and “repair” in speech, in which one repeats part of a phrase (e.g., “I’m taking aunt Nellie to the new hair (um, um) . . . to the new hair salon.”). The repair is either the iteration of a mistake, or more plausibly, the re-iteration of a deviant utterance to show that it was not a mistake. Since the second segment of our melody offers an alternative approach to the end of the phrase, as the final pitch D is now approached from above rather than below, the ensuing repair requires backing up and repeating some elements that appeared prior to the trouble spot, in this instance, the leap up to the F. In speech, this process is known as “anticipatory retracing.”\(^{17}\) The “repair” motive terminates with an eighth-note rest. As a result, the lengths of the first three rhythmic groups consist in 6, 7, and 6 eighth notes, respectively. More precisely, however, the lengths might be expressed as 6, 6+1, and 6 eighth notes, because the
middle unit is expanded by way of an “infix” (as opposed to a prefix or suffix). Thus, although the second figure bobbles the sense of 3/4, the motivic repair restores it.

It is also worth noting that the hiccup occurs in the context of an almost-constant stream of rapid eighth notes. At the marked tempo of ♩<Arial Unicode MS 16 pt.> = 120, the eighth-note-based notes (and syllables) are too fast to be heard as beats. Beats emerge, instead, at the quarter-note level. At this tempo, and even more so were the tempo a bit faster, we experience an interesting perceptual effect. The rapid quarter-note pulse is almost, but not quite, too fast to be heard as a beat (recall the 400 ms threshold noted above). In metronomic terms, the beat- vs. beat division boundary occurs around 150 BPM. In addition, we strongly prefer periodicities in the range of approximately 600ms, or 100 BPM, the area of “maximal pulse salience” as described by Parncutt. But Stravinsky fails to provide periodicity at anywhere near this range, for as these pulses are grouped in threes, the next metric level has a period of around 43 BPM (1380ms IOI). As a result, even without the hiccups, our sense of triple meter is perceptually somewhat unstable.

The opening motive is again repeated, this time exactly (D-D-D-C-D). Rather than ending with an eighth rest, however, Stravinsky gives us the entry of the soloists, forte dynamics, new pitches, and a new motive (designated by the parallel thirds and forte dynamic in Ex. 2). These features interrupt the chorus’s repeat of the opening figure. It may well be that the changes just described are meant to underscore musically the important references to the Groom’s name, first his given name (Khvetis) and then his patronymic (Pamfil’evich). Following this, the chorus tries to articulate the popevka again but is interrupted once more. The passage seems more settled after the second interruption when the seven-note extended figure and subsequent "repeat and repair" of measures 2-3 are repeated. And the section closes with a final repeat of the opening five-note figure (D-D-D-C-D). A new section begins in the next measure (again, the eighth-note rest is absent).

The grouping structure of the entire passage is diagrammed in Table 1.

<table>
<thead>
<tr>
<th>Length</th>
<th>6</th>
<th>(6+1)</th>
<th>6</th>
<th>(6-1)</th>
<th>6!</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Basic motive</td>
<td>Extension</td>
<td>Repair</td>
<td>Basic motive, truncated</td>
<td>Interuption</td>
</tr>
<tr>
<td>Length</td>
<td>(6-1)</td>
<td>6!</td>
<td>(6+1)</td>
<td>(6-1)</td>
<td>(6-1)</td>
</tr>
<tr>
<td>Description</td>
<td>Basic motive truncated</td>
<td>Interuption</td>
<td>Extension</td>
<td>Repair truncated</td>
<td>Basic motive truncated</td>
</tr>
</tbody>
</table>

The average length of the rhythmic unit is six eighth notes, but I am not arguing for statistical listening (i.e., that we infer a sense of 3/4 from the emerging average length of the popevka presentations). Clearly, these segments do not fit into 2/4 or 4/4 meter, for no grouping structure supports such duple architecture. Nor can we discern a stable non-isochronous meter (e.g., a pattern of 6-7-6-7, etc.). Still, although 3/4 might be operative—as the popevka strongly implies—it, too, lacks stability at the level of the bar. Thus, the groupings of 5 and 7 really are hiccups; they create what might be called measures manqués. As our sense of the location of the downbeat is constantly thwarted,
we have an attenuated appreciation of the higher-level metrical organization. We can thus coordinate our entrainment/virtual movement with the D-D-D-C-D motive only weakly, for we simply cannot be sure where it will start. As David Huron suggests:

Segments of his [Stravinsky’s] music exhibit a systematic organization whose purpose is to actively subvert the perception of meter. Although his music remains highly rhythmic, some of Stravinsky’s most distinctive passages are methodically "contrametric" in structure.19

In her article on “Metric Irregularity in Les Noces,” Gretchen Horlacher argues against hearing shifts like those described above in relation to a “background periodicity,” as suggested by Pieter van den Toorn.20 According to van den Toorn’s analysis, the listener maintains a stable 3/4 meter while the surface alignment shifts in and out of phase with it. This is a necessary precondition for van den Toorn’s notion of metrical displacement, whereby, in lieu of the morphological development of a motive, a motive is developed by putting it in a variety of metric contexts, most especially by shifting the phase relation between the motive and its metric frame.21 Displacement presupposes a stable and fixed background meter.

Hornlacher rejects the notion of a stable metric background, and I agree. The background meter here (and elsewhere) is not robust enough to be maintained by the listener while these putative displacements take place. I, too, prefer a more radical rather than conservative mode of hearing the meter in this passage. However, I disagree with Horlacher’s claim that “the pitch D can be heard in all three metric contexts”—that is, as a downbeat, as an offbeat, and as a weak beat.22 Of course, this is true within the popevka motive itself, as its first three Ds are a downbeat, an offbeat, and then a weak beat. Rather than tell the story of pitch-class D, I would prefer to relate the tale of the popevka motive in toto. Moreover, when we hear a D supported by the drum roll sound, we know it is a downbeat, a D which usually marks the start of a series of Ds. Because of those associations we can hear the beginning of the popevka motive in a relatively fixed metric fashion and thus instantly grasp the metric placement of the notes that follow. Conversely, the D that ends the “stretch” figure in the first hiccup does not have this contextual/timbral support. While a conservative metric hearing would put this D on the downbeat of a perceived measure, we immediately know something is off as it lacks the support of the snare drum. The next D, accented and supported with the drum roll, confirms that a motivic game is afoot. Thus we have a more supple rhythmic figure, one that ranges from the relatively fixed metrical posture it assumes at its onset to the significantly more pliant approach to meter it exhibits as it unfolds and ends.

I also disagree with Horlacher’s larger claim that “by virtue of the regularity of their patterning, the instances of ostensible metric irregularity together produce a relatively stable perceived meter.” She goes on to say “I am arguing here for a contextually-determined cyclic meter that may be inherently irregular.”23 My disagreement turns on the notion of “cyclic.” If the measure level is unstable, one cannot have a stable metric cycle, and indeed, one then does not have “meter” in the usual sense of the word. What I think Horlacher is trying to suggest is that we have a stable
association between the D-D-D-C sub-motive and the onset of a unit that *ought* to be three beats (or six pulses) long. But this “ought” is continually thwarted. We are in agreement, however, on the upshot of Stravinsky’s motivic machinations: in this tableau, a game is being played with the *popevka* figure and its relation to its implied meter.

To summarize, in the opening measure of this tableau, a nearly continuous level of eighth-note pulses is established and maintained. These pulses are too fast to be heard as beats. The central motive, the *popevka*, which is comprised of five notes and a rest, is first articulated, and it strongly implies ternary meter. But this figure goes on to be expanded or truncated in subsequent presentations. As a result, it becomes metrically unmoored. These expansions and truncations affect the beat and downbeat location relative to the constant stream of eighth-note pulses. Although the beat period is stable—we are never tempted to hear a tactus on the dotted quarter-note level, for example—the beat shifts in a discontinuous fashion. As a result of these expansions and truncations, the measure level is unstable with respect to both period and phase. It is not clear if the meter really is 3/4, 5/8, or 7/8 (although 3/4 seems most likely). Metrically, the result is that, although we know what to do and what to expect (i.e., the *popevka* in 3/4), we cannot know exactly when to expect it.

**Social cognition, joint action, and emotional experience**

Thus far we have been considering meter, entrainment, metric ambiguity, and Stravinsky’s hiccups in a conventional music-analytical context, that of an abstract or ideal listening situation, which presumes solitary listeners, each individually engaged with the music. In such a context, any entrainment that occurs is between the musical rhythms and the rhythmic response(s) of the listener; the social setting and the presence of other listeners are generally thought to be irrelevant, if they are considered at all. Yet more often than not, we still listen to music with others, whether we are listening in our cars, in our kitchens, at a club, or in a concert hall.

In other words, the musical experience is a social activity, and thus music cognition typically occurs in a social setting; it is a species of social cognition. *Joint action* is the study of perception and behavior in a social setting. Studies of joint action acknowledge that seeing or hearing something alone is often very different than seeing something in the presence of another person. Tom Cochrane notes that joint action occurs when two or more perceivers:

integrate their attention to something, both in terms of when and what they attend to, as well as the sorts of properties of that object they are interested in. . . . [J]oint attention can imbue the object with additional social meaning, as something that signifies our relation to each other, as something that affords a co-operative action or as something that arouses a social emotion in us.\(^\text{24}\)

Thus if you and I happen to be sitting on a park bench watching some playful birds, there is more going on than our simultaneous perceptions of the birds and their actions. Our
individual bits of bird watching are now enmeshed in a social context. As John Campbell notes:

Just as the object you see can be a constituent of your experience, so too it can be a constituent of your experience that the other person is, with you, jointly attending to the object. This is not to say that in a case of joint attention, the other person will be an object of your attention. On the contrary, it is only the object that you are attending to. It is rather that, when there is another with whom you are jointly attending to the thing, the existence of that other person enters into the individuation of your experience.²⁵

Naomi Eilin notes that the added social dimension of joint action affects our perceptions, emotions, and actions: “Each subject is aware, in some sense, of the object as an object that is present to both subjects. There is, in this respect, a ‘meeting of the minds’ between both subjects, such that the fact that both are attending to the same object is open or mutually manifest.”²⁶ This is what Cochrane means by such constructions as the “integration of attention” and “signifying our relation to each other”--the latter by way of recognition of our mutual (or non-mutual) reactions to the object(s) of our joint attention.

Listening to music in a social setting is a paradigmatic example of joint action. It is not simply that in such settings you and I are both listening to the same music, at the same time, and in the same place. Joint action also includes my awareness of your presence, and your awareness of my presence, and our mutual awareness that the music is the object of both of our attentions. This in turn (and most crucially) leads to our mutual awareness of each other's reactions to the music. Moreover, unlike bird watching, which may or may not involve joint attention, listening to music is an activity that involves joint action by design; pieces of music, like other forms of art, are intended to function as the objects of joint attention. Hence composers and performers presume that their music will be heard in social settings and that our responses to the music will occur in those settings.

Our (joint) attention and response to music is not only cognitive and kinesthetic; it is also emotional. As Antonio Damasio notes, all perceptions involve emotional arousal and emotional valence. From an evolutionary/functional perspective, emotions are part of the suite of physical and mental responses to our changing environment.²⁷ Of course the expressive properties of music are aimed at particular kinds of arousal and valence for the listener; we note that the music moves us as individual listeners, and of course it also does so when we are listening alone. Emotions, however, are strongly social in nature and play out differently in a social setting. Indeed, in many instances, a social setting is prerequisite if emotional or affective responses are to occur at all. For example, Robert Provine shows that laughter rarely occurs outside of a social context and that some emotions (e.g., shame, embarrassment, and guilt) are expressly social in their nature and ethology.²⁸ Even our primary emotions (happiness, sadness, fear, anger, surprise, and disgust) are strongly affected by the social settings in which they occur. Jonathan Turner and Jan Stets, in their work on the sociology of emotion, characterize emotional responses to language as involving the following components:
(1) the biological activation of key body systems; (2) socially constructed cultural definitions and constraints on what emotions should be experienced and expressed in a situation; (3) the application of linguistic labels provided by culture to internal sensations; (4) the overt expression of emotions through facial, voice, and paralinguistic moves, and (5) perceptions and appraisals of situational objects or events.”

Most theories of emotion in a social context have focused on linguistic exchanges (i.e., conversations) and emotions as manifested in facial gestures. Thus factors such as gender and social status among interlocutors, encultured habits of facial display, linguistic vocabularies for the individuation of emotional states, and so forth, have been the subjects of recent research. But this same framework may readily be applied to a range of contexts in the area of music listening.

Among the various sociological theories of emotion discussed by Turner and Stets, Randall Collins’s Interaction Ritual Theory is the most relevant for considerations of emotion and joint action in musical contexts. Turner and Stets note that “for Collins and those who have followed his lead, the entire [social] encounter is conceptualized as a ritual. Stereotyped sequences of talk and body gesture . . . are rituals with a small 'r,' whereas interaction ritual theory sees the rituals (with a capital 'R') as the dynamic interplay of all of the elements of the encounter. . . .” Those elements involve:

1. The physical co-presence of individuals;
2. their mutual awareness of each other;
3. a common focus of attention;
4. a common emotional mood;
5. the rhythmic coordination and synchronization of their verbal and non-verbal gestures;
6. the symbolic representation of their common focus and mood with objects, words, and ideas; and
7. a sense of moral righteousness about these symbols marking their group membership among those engaged in the rituals of social encounter.

Items 1-3 above are necessary conditions for joint attention, 4-5 are the products of joint attention (more on these in a moment), and 6-7 are the cognitive results and realization of the experience—how the experience generates social capital, as it were. One can readily adapt this framework to our experience of music in a range of social settings, so long as these settings embody the criteria reflected in the following scenario. The members of an audience:
London--12

- jointly attend to the music;
- Their rhythmic entrainment to the music and to each other engenders a common affective response;
- This leads, in turn, to a shared representation and appraisal of the musical object (i.e., the musical work, its performance, or both);
- And it also leads to a shared sense of the aesthetic and cultural value of the overall musical experience.

To be sure, joint attention may occur in the absence of rhythmic entrainment, as when the music simply does not project a regular pulse or meter. In such cases there may be shared affective responses, representations, and appraisals. Yet when entrainment is present, it can both amplify the affective response as well as give rise to responses that would otherwise be absent. As Günther Knoblich and Natalie Sebanz note, entrainment of rhythmic behaviors in both musical and non-musical social settings seems to be unconscious and automatic, and perceiving another’s movements activates one’s own action system for that same movement (i.e., engaging the so-called “mirror” neuron systems). 33 This in turn leads not just to parallel actions, but to parallel affective responses as well. In their characterization of emotional contagion, Elaine Hatfield, John Cacioppo, and Richard Rapson note that mechanisms of emotional contagion especially include mimicry and feedback: “Subjective emotional experiences are affected, moment to moment, by the activation and/or feedback from such mimicry.” 34 Even the simplest of coordinated rhythmic actions can do this. Michael Hove and Jane Risen, in an empirical study of temporally coordinated joint action, asked participants and an experimental confidant to tap along simultaneously to a visual metronome (i.e., a bouncing icon on a computer screen). 35 They found that when tapping in synchrony with the confidant, participants' ratings of partiality or affiliation for the experimenter (who was otherwise a total stranger) were higher than in either of the two following circumstances: (a) a situation in which both participants and the confidant tapped, but not in synchrony; or (b) a situation in which the participants tapped but the confidant did not.

Our shared affective responses are those that are appropriate to both the expressive properties of the music and the social setting in which it is heard. A stirringly sad performance of a song in a jazz club and a stirringly sad performance of a piano sonata in a darkened concert hall may give rise to very similar feelings but rather different social manifestations. Yet in both contexts there generally are overt signs of one’s engagement with the music. These are visible and audible to the people sitting nearby. They may range from slight movements (or being “frozen in one’s seat”), to patterns of breathing, to breaking into tears. The power of music is that, even in the absence of overt motion, its rhythmic structure allows us to “inwardly second” (to use Jerrold Levinson’s phrase) what we hear and, in a social context, presume that others are doing the same. 36
Musical expression and keeping together in time

What then are the aesthetic consequences of entrainment (or the lack thereof)? We may presume that composers want listeners to grasp fully all of the aesthetic properties of their music, including, most especially, its expressive properties. We may also presume that composers imagine their works will be heard in their normative contexts—that is, in social settings ranging from church and court to the concert hall and the parlor, as noted above. Thus composers, like all human beings, have an implicit/intuitive sense of social cognition and joint action, and this entails the dynamic synchronization of the audience with their music. When this occurs, the expressive properties of the music—the music’s emotional life and movement, as enacted by the performer or performers—are grasped by each listener and confirmed and amplified by their shared parallel responses. As the audience moves together, they respond to the music more fully and in a more nuanced fashion. In many cases, the “representation and appraisal” of the music’s affective character may only be possible by way of joint action—for how, outside of a social context, can one make sense of a performer’s swagger in executing a particular rhythmic figure, or why one cringes in response to a melodic detail?

What happens, then, to an audience listening to the metric hiccups in *Svadebka*? How does this rhythmic context both create an expressive effect as well as the context in which that effect can be grasped, felt, and judged? Recall van den Toorn’s characterization of these hiccups as motivic displacement against a fixed metric framework. Van den Toorn’s analysis is motivated not just by the music, but also by a desire to confront Adorno’s characterization of Stravinsky’s rhythmic practice. For Adorno, according to van den Toorn, believed that:

[T]he need for strictly held meters is made to imply mechanization and impersonality, while mechanization leads to "anti-humanism" and a "collective" voice that stands in opposition to the individual. . . . The [literal] repetition of Stravinsky’s themes and motives is made to imply a similar identification with the "murderous’ collective," as Adorno expressed it, with "agents of destruction."  

Van den Toorn wants to acknowledge Stravinsky’s strictly held meter but deny that it leads to the anti-humanism and negative identification proposed by Adorno. In this, van den Toorn both gets it right and gets it wrong. On the one hand, our attraction to and innate propensity for entrainment—as would be engendered by a strict meter—goes beyond mechanization and the modern collective; it is a deeply rooted aspect of our social evolution and cognition. Our capacity for collective coordinated movement is something that sets us apart from our evolutionary ancestors, something that makes us uniquely human. And of course, in a wide range of musical styles from the concert hall to the dance club, we enjoy and value music that engages our capacity for entrainment in interesting ways, whether by modulating the depth of our entrainment, or by forcing us to change the way in which we entrain, whether via metric modulation or by various forms of metrical “dissonance.” As result, we move together in interesting and changing ways.
In contexts like *Svadebka*, however, metric ambiguity, even with the blows and shocks of the irregular accents, inhibits the formation of the collective movement—we cannot move in lock step to this passage. Indeed, what is enacted before us is a social situation of conversational stumbles and interruptions. In part, this is a portrayal of the hurly-burly that is taking place at the house of the groom, with the comings and goings of families and friends. It may also be the enactment of a ritual that is somewhat unsure of itself, projecting an ambivalence about the impending nuptials, a recognition of what is both gained and lost by the bride and groom and their families.

So how do we (the audience) respond to Stravinsky’s hiccups? We cannot move along with them, though we can follow closely enough behind. And I also think that, owing to the very fast beats that result from the tempo Stravinsky has chosen, we have a “cognitive compulsion” to keep after it, for the sense of the basic beat is thoroughly irresistible. As a result, a sense of frustration arises here, for the popevka figure is so enticingly assertive of 3/4 that we keep hoping Stravinsky will really adopt this meter. But he doesn’t. At some point, we realize that this listening strategy (whether conscious or not) just isn’t going to work. Once we realize this, we may then approach the metric hiccups in a different manner. As van den Toorn also notes:

> [E]xperiences of metrical conflict in Stravinsky’s music need not be the debilitating, traumatic ones scorned by Adorno. As internalized meters are brought to the surface of consciousness in the most disruptive cases of displacement, a heightened sense of attention, engagement and suspense may ensue.\(^{40}\)

The metric hiccups do sharpen our attention, just as they surprise us. Even if our automatic, perceptual response is to keep hearing the popevka motive as marking the onset of 3/4, on a higher level we recognize its metric unreliability. A game is being played here, and when we grasp this, we might, as van den Toorn suggests, take pleasure and delight in following along with these metric incongruities.

At least that would be so were we to consider our individual, quasi-analytical responses to *Svadebka*. But when we hear this in the concert hall, or at other times with others, the result may be a bit different. Beyond delight, we may feel a touch of unease, for, if we cannot smoothly follow along as individuals, we are even less able to do so as a group. Our rhythmic and hence our affective responses to this passage are attenuated. This may be regarded as a masterful compositional stroke by Stravinsky, for it allows the tension and uncertainty of the wedding scene to overflow from the stage and onto the audience. As a result, the audience is more integrated into the action. But the cost (which may admittedly be modest in this case) is a weakening of joint action and an affective uncertainty, which may also create a sense of distance between the listener and the work. Perhaps this is what lies behind Nicolai Myaskovsky’s characterization of *Svadebka*: “I hardly understand a thing although here and there some feeling comes through, it’s all very dim. . . . There’s something prickly and ascetic about it.”\(^{41}\)
Rhythmic entrainment and the joint action it engenders fulfill two functions. The first is aesthetic. As we are entrained and engaged in social cognition, our affective responses to the music are enhanced and sharpened. Our affective responses are a key component—and for many, the key component of our aesthetic experience of music. And indeed, in many post-formalist approaches to aesthetics, the function of an artwork is to afford aesthetic experience through the appreciation of its unity, complexity, and intensity and their resultant expressive properties. Yet in much of the art music of the 20th century, from Babbitt to Boulez, the absence of any regular pulse or beat forestalls the formation of any entrainment whatsoever. Indeed, it may well be for this reason, and not owing to the complexity of pitch structure often found in 20th-century music, that many concert-goers often characterize such works as “not music at all.” For these listeners, such music fails to fulfill the primary function of affording an aesthetic experience, because, whatever formal and affective properties such works might have, they are dead in the rhythmic water.

The second function of entrainment and joint action is social in that it enhances the sense of group cohesion among audience members. As William McNeil noted in the quotation with which we began this study, “rhythmic movements have an emotional affect, and keeping together in time “makes us feel good about ourselves . . . and vaguely pleased with the world at large.” The ability of music to enhance group cohesion and engender positive affect underlies its ubiquity in important social contexts across all human cultures and historical eras. Even if the marches and drills are long and difficult, or if the piece to which we are listening reflects bitter anger (Jimi Hendrix’s “Star Spangled Banner”) or tragic sadness (Mahler’s Kindertotenlieder), we still feel that such experiences are worthwhile, and this value arises not in our grasp of the negative emotions inherent in the experiences themselves, but as a result of the social bonding such experiences afford. This is what gives rise to fellow-feelings, our shared sense of the moral righteousness regarding both ourselves and the object of our joint attention, to which, as we have seen, Randall Collins alludes.

When we move together, we are most able to imagine ourselves as another person—to act out what Ted Cohen has characterized as a “metaphor of personal identification.” Such metaphors, according to Cohen:

are entrées to human understanding, to the appreciation of one another. They demand to be grasped. Grasping them is part of one’s commitment to being human, for being human requires knowing what it is to be human, and that requires the intimate recognition of other human beings.

That intimate recognition may be between ourselves and the composer, ourselves and the performer(s), or ourselves and other listeners. When we ground our listening attitude in terms of rhythmic entrainment, in terms of social cognition and joint action, we can extend our recognitions across time and space to other listeners whom we imagine to be
moving with the music as we do. Listening together, with the affirmation of fellow-feeling that it provides, is not only important in helping us understand what it is like to be another person. Listening together also helps us understand ourselves, who we could have been, or who we might become.
Endnotes


8 On the perceptual differences of rhythms in these ranges see London, *Hearing in Time*, 27-47.


11 An excellent study of the group dynamics in performance may be found in Martin Clayton, *Time in Indian music: Rhythm, Metre, and Form in North Indian Rāg Performance* (New York: Oxford University Press, 2000).

12 A passage without any clear tactus would not be metrically ambiguous, but metrically vague, as ambiguity presumes that some sense of meter—indeed, more than one sense of meter—is present. On the distinction between metric ambiguity and metric vagueness see *Hearing in Time*, pp. 79-86.


I thank my colleague Professor Laura Goering for her transliteration and translation of the Russian.


Horlacher, “Metric Irregularity in *Les Noces*,” 302-03.


