Should embodied cognitive science go radical? A hint from music

Pierre Saint-Germier
University of Geneva

1 Introduction

That human cognition is in some sense embodied has now become more or less accepted by a large number cognitive psychologists and cognitive neuroscientists. This does not mean that everyone agrees about the sense in which cognition is embodied. Shapiro distinguishes three importantly different meanings of "embodied cognition" (2011), Wilson counts up to six (2002). In any case, everyone agrees that some understandings of embodiment are rather weak and almost uncontroversial. Others are much stronger. One of the strongest views one can find in the literature is what has come to be called "Radical Embodied Cognition" (Chemero 2009):

> Radical embodied cognition amounts to two positive claims and one negative claim.
>
> **Radical embodied cognition, claim 1** Representational and computational views of embodied cognition are mistaken.
>
> **Radical embodied cognition, claim 2** Embodied Cognition is to be explained via a particular set of tools *T*, which includes dynamical systems theory.
>
> **Radical embodied cognition, claim 3** The explanatory tools in set *T* do not posit mental representations. (Chemero 2009, p. 29)

In this paper, we will focus mostly on the issue of mental representations. A distinctive aspect of Radical Embodied Cognition is its eliminative stance toward mental representations. This eliminativism is a distinctive element in the package, since not all proponent of Embodied Cognition are ready to throw mental representations away (although some other Radicals[1] are willing to so). A popular, rather Moderate, view of embodiment is that our mental representations are grounded in sensory-motor representations (REF). The general issue we would like to tackle in this paper is whether Embodied Cognition should go radical and forget about mental representations. But before presenting more precisely the focus of the paper, we will have to make a little methodological detours

2 Verbal disputes in the Philosophy of Cognitive Science. A diagnosis and a remedy

Reboul (2011) offers a diagnosis of the debate over REC and a critique of this particular view. The diagnosis is that much of the debate seems to be merely verbal. Chemero promotes eliminativism about mental representation, in accordance with claim 3. But Chemero defends his eliminativism by defining "mental representation" in such a way that it does not require decouplability (2009, p. 66). It is then not too difficult to claim that such mental representations are idle mediations between the

[1]See for example Hutto and Myin (2012).

brain, the body and environment and are best dispensed with. But should we really accept to define "mental representation" in this way? Chemero might insist that defining "mental representation" in such a way that it requires decouplability is just begging the question against eliminativism. So it looks like the disagreement between the realist and the eliminativist about mental representation centrally hinges on how we define words rather than on substantive empirical facts.

One possible explanation for this situation is that each party in the debate wants to decide the question by an analysis of the *intension* of the basic concepts involved in the debate, namely COGNITION and MENTAL REPRESENTATION. The remedy Reboul proposes is that we focus instead on the *extension* of those concepts for which both parties agree. This is the case for COGNITION , if not for MENTAL REPRESENTATION. So rather than quibble over the intension of COGNITION, we should examine phenomena which clearly fall under the extension of the concept, and ask ourselves, for example, if decouplable mental representations are required to best explain them. With such rules in place, disagreements about realists and eliminativists about mental representation should be explainable and resolvable on the basis of empirical reasons rather than definitional choices. And as Reboul insists, although some behavioral capacities which are presumably cognitive may be explained without decouplable mental representation (collaborative hunting among Tai chimpanzes for example), some other uncontroversially cognitive phenomena can hardly be explained without them, such as engaging with fiction or performing thought experiments. Even though Chemero claims that the tools of dynamical systems theory can describe systems capable of decouplable representations and explain some "representation-hungry" capacities[2], it is still not clear how these tools can provide interesting

[2] Chemero's favorite examples are Van Rooj, Bongers and Haselager's "non-representational approach to imagined action" (2002), i.e. imagining using a given stick to reach distant objects, and the study by Stephen, Dixon and Isenhower of gear-systems problem solving with the tools of dynamical systems theory (2009).

explanations of fiction and thought experiments. The conclusion is that the program of radical embodied cognition is at best limited in scope. This might not be bad news if taken as a modest claim. If taken as a general claim about the nature human cognition, it is as yet still unwarranted.

This paper largely endorses Reboul's diagnosis. Its focus, however, moves to music cognition. Music cognition seems to be, *prima facie* at least, an easy case for Radical Embodied Cognition[3]. Our relation to music, and to rhythm in particular seems to involve very intimately our bodies. Our bodies spontaneously synchronize to hip grooves (at least for some of us). And it does not seem that mental representations play a crucial role there. It is all a matter of coupling between the physical signal and our body.

Although Emobodied Cognition has its supporters in the field, it is not clear however that all of them would endorse the Radical version of the view offered by Chemero. For example, Maes, Leman, Palmer and Wanderland (2014) in one of the most detailed account of action-based effects on music perception rely on the theoretical framework of event coding theory (Hommel et al.) which is unambiguously representational[4]. However at the end of the paper they show a growing sympathy for Radical Embodied Cognitive Science, without stressing the fact that a full commitment to this approach would require the elimination of mental representations from their framework, and consequently the abandon of Event Cody Theory.

[3] Chemero as a matter of fact co-authored a paper on jazz improvisation which uses the tool of dynamical systems theory to study the synchronization of bodily movements during improvisations (Walton et al. 2015).

[4] The main claim of event coding theory is that "perceptual contents and action plans are coded in a common representational medium" (Hommel et al. 2001, p. 849)

Following Reboul's suggestion, one might start looking for interesting phenomena which clearly all within the extension of the concept COGNITION. In order to put further pressure on Chemero's eliminativism about representationalism, one might consider "representation-hungry" cognitive capacities. Notational audiation (the capacity to mentally hear an auditive image of a musical piece while reading its score),[5] the phenomenon of earworms (musical melodies spontaneously popping into ours minds without external stimuli)[6] or the capacity of composers to create new music on a table, with just pencil and paper[7]. This is not the strategy we will follow here. We will consider phenomena which do not seem *prima facie* "representation hungry" and show that they raise a difficulty for the eliminativism of Radical Embodied Music Cognition. The argument will depart in another important way of from the kind of argument Reboul uses against Chemero, but more on that later. It is time for some music.

3 The perception of rhythmic structure

In this section we consider how an Embodied Cognition (which may or may not be Radical at this point) can explain the perception of rhythm structure. The best way to do so is to see in general terms what it takes to explain musical structure, in general terms, before to put the focus on rhythm.

Here is how Lerdahl and Jackendoff, two pionners in the field, characterize the phenomena of musical structure :

> One commonly speaks of musical structure for which there is no direct correlate in the score or in the sound waves produced in performance. One speaks of music as segmented into units of all sizes, of patterns of strong and weak beats, of

[5] See Brodsky et al. 2008 for a study which provides evidence for the view that notational audiation involves both phonatory and motor processing.

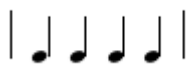
[6] See, for example, Beaman and Williams 2010.

[7] Although this phenomenon is still understudied, see Féron and Donin 2012.

> thematic relationships, of pitches as ornamental or structurally important, of tension and repose, and so forth. Insofar as one wishes to ascribe some sort of "reality" to these kinds of structure, one must ultimately treat them as mental products imposed on or inferred from the physical signal. In our view, the central task of music theory should be to explicate this mentally produced organization. Seen in this way, music theory takes a place among traditional areas of cognitive psychology such as theories of vision and language. (1983, p. 2)

On this view, which is quite standard from the standpoint of classical cognitive science, one is said to understand a piece of music only if one is able to assign a structure to this bit of music. Since this structure is not always manifest in the physical signal itself, this assignment of structure has to be attributed to the mind.

This assignment of structure to the music we hear is something we do quite spontaneously. I shall here give a brief example borrowed from Philip Johnson-Laird (1991, p. 90) to give an idea. If one counts "1 2 3 4" in a regular way so as to establish a meter and then clap the following rhythm at the same tempo:



listeners will judge all four notes to be of equal duration. So far, so good. Now if one counts "1 2 3 4" and then claps

the listeners will judge the last note to be shorter than its predecessor, which is remarkable since all the notes have the same physical duration in both rhythms.

The standard explanation for the example under discussion can be outlined as follows: The listener perceives both rhythms as having a meter of four beats to the measure, as suggested by the experimenter counting "1 2 3 4". So they assume that there will be another note, on the first beat of the next measure. Consequently, in the

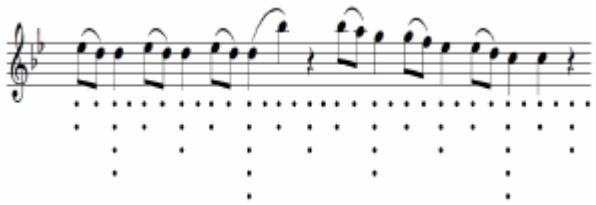
second example, the listener construes the last note as an eighth note, inducing the judgment that the last note is shorter than its predecessors. For if this note were construed as a quarter note, it would contradict the anticipation of a note on the first beat of the next measure. Construing the fourth note as an eighth note permits to avoid this conflict. In the first example, however, the last note can be interpreted as a quarter note, like its predecessors, consistently with the anticipation. Hence the judgement that all the notes have the same duration.

Lerdahl and Jackendoff, in their 1983 book, give explanations of the same sort, although more sophisticated, and of more sophisticated phenomena. They provide a theory of metrical structure along the following lines. The basic ingredient of the metrical structure is the beat. Beats are idealized events of infinitesimal duration. The beats have to be equally spaced. However, a series of beats forms a metrical structure only if there is a distinction between strong and weak beats. For beats to be strong or weak there must exist a metrical hierarchy, that is two or more levels of beats. In 4/4 meter, for example, the first and third beats are felt to be stronger than the second and fourth beats and so are beats at the next larger level. The first beat is felt to be stronger than the third beat and so appears at the next level of metric organization, and so on.

So the basic question the classical cognitive scientist has to answer is how metrical structures, so defined, are to be assigned to bits of musical surface. To answer this question Lerdahl and Jackendoff distinguish two kinds of rules, well-formedness rules and preference rules. Rules of the first kind define the set of possible metric structure in the idiom under investigation. For classical tonal music, our two authors give four basic rules, two of which are claimed to be universal, the other two being peculiar to the tonal idiom. A universal WFR is for example the rule that every beat at a given level be also a beat at the smaller levels present at that point in the piece.

The rule that in each metrical level, strong beats are spaced either two or three beats apart, unlike the preceding, is characteristic of classical tonal music.

The preference rules then address the problem of relating these possible structures to a given musical surface. For it can happen that various possible metrical structures be compatible with a given musical surface. The task of the preference rules is then to give the structure that corresponds to musical intuitions in the idiom under consideration. For example, where two or more groups or parts of groups can be construed as parallel, they preferably receive parallel metrical structure. Lerdahl and Jackendoff's tour de force is to have provided a system of rules capable of explaining the perception of very complex pieces of music, among which some of the masterpieces of European classical music. Here is an example of a assignment of metrical structure to the beginning of Mozart's Symphony in G minor:



The relative metrical strength of a beat is represented by the numer of dots under the corresponding note.

It has to be noted that this theoretical approach is in a sense psychologically incomplete, something the authors perfectly acknowledge. Their purpose was to set out a theory of the *competence* of a hearer familiar with a particular idiom. Which means that they did not inquire into the nature and details of the mechanisms at work in rhythm perception, among other aspects of music perception. They only described what the mechanisms should arrive at. Thus, given a musical surface, an account the mechanisms of rhythm perception should detail the effective procedure

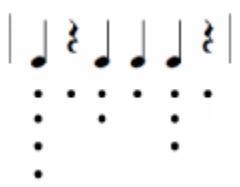
by which the mind constructs in real time a mental representation of the rhythm structure from the acoustic input.

So much for the classical cognitivist part of the story. As you have probably remarked, the body is absent from the plot. Actually the body is absent both from the *explanans* and the *explanandum*. The target phenomenon is described in terms of mental operations and the *explanans* is constituted by rules. So there are at least two ways for the embodied cognition theorist to challenge the cognitivist explanation: reject the initial description of the phenomenon in sole terms of mental inference or reject the explanation itself. In the first way, you say something like : "You forget a crucial part in the phenomenon to be explained, namely the fact that it is the active response of an embodied organism to a physical stimulus, not just a mental inference. Obviously, the body is involved. Look at phenomena like tapping, for example. Sensory-motor interactions clearly are involved in rhythm perception. So you need to change the explanandum".

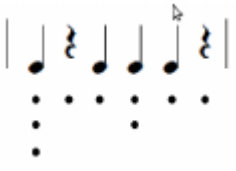
I won't explore this strategy here, however. For I think this is not the strongest way to introduce an Embodied Cognition approach. It is, in contrast, a much stronger way of challenging the cognitivist to take the explanandum as she characterizes it and show that her explanation won't work unless it refers to bodily states or processes. So the embodied cognition theorist, in order to make a strong case, must show how the body is involved in the apparently very mental act of assigning a metrical structure to an auditory input.

Now there is evidence that the body is involved. In a series of studies, Jessica Silver-Philips and Laurel J. Trainor (2005, 2006, 2007) focus on ambiguous rhythmic patterns. Consider the following rhythmic pattern:

Suppose this pattern is played with no physical accent. There are two ways for the hearer to understand this pattern as a metrical structure: in duple form or in triple form. In the duple case, the pattern is divided in three group of two beats:



In the triple case, the pattern is divided in two groups of three beats:



The hypothesis under investigation is that bodily movements can help to disambiguate the rhythm. Various experiences have been designed to test it. A first one (2005) was done with 7 month-old infants. The infants were divided into two groups. Those of the first group were bounced up and down while held in the arms of the experimenter on every second beat. Those of the third group were bounced on every third beat. Thus, both groups heard the same stimulus, but had different movement experiences. After training, infants preferences were tested for an accented version of the auditory stimulus in duple form versus an accented version in duple form pattern. Infants controlled how long they listened to each version of the rhythm in a head turn preference procedure. Now infants chose to listen longer to the auditory test stimulus with accented beats that matched the beats on which they were bounced. Thus their bouncing determined whether infants later preferred the auditory rhythm pattern congruent with duple or triple form. This effect was then showed to be independent from visual information, that is the effect is still observed when the infants are blindfolded. It was also shown that the effect depends on the

personal movement of the infants. For the effect disappears if the infants don't move but simply watch the experimenter move every second or third beat.

This effect was then (2006) generalized to adults, using the same experiment, except that instead of being bounced, the adults bounced themselves by bending their knees, imitating the experimenter. At test, they identified as similar an auditory version of the rhythm pattern with accented strong beats that matched their previous bouncing experience in comparison with a version in which the accents of did not match. Similarly, the effect was shown not to depend on visual information but that movement of the body is critical.

So it looks like we have just what we need. The task which the classical cognitivist picture assigns to the mind, i.e. assigning metrical structure, is actually performed by bodily movements. This is probably at least a good *prima facie* case for endorsing Embodied Cognition[8]. Is it a sufficient reason to throw away mental representations and embrace Chemero's Radicalism?

4 Grounded mental representations

In order to adjudicate the matter, we need to go back to the general structure of the dialectical situation. The Radical wants to eliminate mental representation. For that she needs to show that all the work that is done by mental representations can be done without them. The Moderate wants to keep mental representations. For that she needs to show that some property which can only belong to mental representation is instantiated. The property used by Reboul in her argument against Chemero's Radicalism was a very strong form of decouplability. Since none of the stories Chemero is able to tell with his dynamical tools can explain the phenomena that

[8]At least Leman and Maes cites these results as evidence in favour of the Embodied approach in his review of the field (Leman and Maes 2015, p. 85).

strongly decouplable representations do explain, mental representations should not be eliminated.

In the present case, decouplability is of no use to resist Radicalism. On other hand, a full theoretical explanation of the effects demonstrated by Silver-Philips and Trainor in terms of dynamical systems has not been offered yet, to our best knowledge. So the Radical has not won yet. However, it quite plausible that a dynamical story can be told in this case.

There is however another property of mental representations one might appeal to in order to resist the elimination of mental representations. This property is intensionality – with an "s". A mental representation is *extensional* iff its reference remains stable when one its it parts is replaced by co-referring mental representaton. A mental representation is *intensional* iff it is not extensional. Suppose I have known Stefani Germanotta when we were both teenagers, before she came to be know as Lady Gaga. Suppose I go to one of her shows and see her on the stage. If there is one sense in which it is true to say that I do not see Stefani Germanotta on stage but Lady Gaga, then the mental representation I have of Lada Gaga is intensional. If on the other hand one insists that the content of this mental representation is exhausted by its referent, namely this one individual which happened to be know under the name of "Stefani" and then later under the name of "Lady Gaga", then the mental representation is extensional. It follows that the mental representations which are posited in order to make sense of the ambiguity of the unaccented musical stimulus studied by Silver-Philips and Trainor have to be intensional, for we have to be able to make a difference between hearing it "as in duple form" and hearing "as in triple form" although both representations have the same referent.

The remaining question is whether the story the eliminativist is in a position to tell is able to account for these intensional differences in our relation to the physical

stimuli. Chemero's theory of mental representation is unfortunately silent on the topic of intensionality. This theory shows how some properties usually attributed to mental representations can be accounted for in terms of dynamical systems theory, i.e. more specifically in terms of adaptive oscillators. Chemero shows for example that adaptive oscillators can represent objects which are absent from the physical environment, and so mimic some elementary sort of decouplable mental representations (2009, p. 58-60). None of the examples he discusses show clearly how an adaptive oscillator can achieve intensionality in the required way. But it is not implausible that they can do so. Let us grant him as much. Does it follow that we do not need mental representations to explain the phenomenon?

One could adopt the exact opposite conclusion. If a dynamical system is able to instantiate intensionality, it follows that the dynamical system is after all representational. But this kind of reply seems to give rise to the same methodological problems we tried to avoid in the first place. Is intensionality a mark of representationality? If the Moderate says “yes” the Radical will say that she begs the question against her. If the Radical says “no”, the Moderate will say the same. We are back to square one.

So let us follow, once again, Reboul's advice. Let us avoid verbal issues and investigate the phenomena in an open-minded, empirical way. Is there any more to be said of the disambiguation of rhythmic patterns by bodily movements?

In a subsequent paper, Philips-Silver and Trainor (2008) tried to refine their result. For while they had shewn a cross-modal interaction between movement and auditory perception, they did not indicate which aspects of the movement are critical to the effect. An issue was whether passive movement was sufficient to yield the effect with adult subjects. An experiment with a seesaw-like bed showed that passive movement was indeed sufficient. The question then was: is any particular part of the body responsible for the effect? Further experiments with the seesaw were done. In the first

one, the setting was such that the movement of the seesaw affected only the head. In the second one, it affected only the legs and feet. It appeared that passive movements of the leg congruent with a duple, respectively a triple, rhythm did not induce the subsequent perception of the corresponding rhythm, but passive movements of the head did. The authors suggested then that the vestibular system may play a key role in the effect.

This hypothesis was strengthened in another study by Trainor and Gao (2009) which shows that the same effect can also be accomplished by direct galvanic stimulation of the vestibular system. The galvanically induced sensation, without any actual movement, that the head moved from side to side on either every second or on every third beat of the ambiguous auditory rhythm pattern strongly biased whether adults perceived it as being in duple or in triple time.

In the light of these new facts, the situation dramatically changes. It turns out that the bodily movements are causally, but not constitutively, tied to the disambiguation which happens in the brain. The processes we have to mention in order to account for the phenomena are intra-cranial processes, which are intentionally directed toward a target and possess the property of intensionality. They looks pretty much like the neural substrates of good old mental representation. The important respect in which they differ from classical cognitivist mental representations is that they are intimately connected to sensory-motor representations plausibly located in the vestibular system.

Of course, the Radical can always re-describe the facts in terms of dynamical systems involving the body, but those further facts indicate that only a small part of that system is causally relevant to the observed effect. Whether the Radical accepts to call these events “mental representations” is only a verbal issue. But what the facts point to is not a Radical view of embodiment according to which mental

representations are dissolved into couplings involving the brain the body and the environment. Rather, they provide evidence for a more Moderate understanding of embodiment where mental representations are *grounded* in sensory-motor representations.

5 Conclusion

Reboul has suggested that Radical Embodied Cognition faces important difficulties as a general view of cognition, insofar as it promotes an elimination of mental representation. Her point relied on the consideration of "representation-hungry" capacities (for fiction and thought experiments). But following the same methodological rules to avoid verbal issues, we have shown that even in the case of music perception, which seems, *prima facie*, more friendly to Radical Embodied Cognition, the eliminative stance faces important difficulties. An alternative, representationalist and hence Moderate view of Embodied Cognition (since elimination seems to be the measure of Radicalness) looks consequently more promising. "If one oversteps the bounds of moderation, said a famous stoic philosopher, the greatest pleasures ceases to please". It would be unfortunate if Embodied Cognitive Science were to deprive us from the pleasures of music.

References

Beamann P. and Williams T. (2010), Earworms ('stuck song syndrome'): Toward a natural history of intrusive thoughts, *British Journal of Psychology*, 101, 637-653.

Brodksy W., Kessler Y., Rubinstein B. S., Ginsborg J., Henik A. (2008), The mental representation of music notation: notational audiation, *Journal of Experimental Psychology: Human Perception and Performance*, 34 (2), 427-45.

Chemero, A. (2009), *Radical Embodied Cognitive Science*, Cambridge MA, The MIT Press.

Féron, F.-X., and Donin, N. (2012), Tracking the composer's cognition in the course of a creative process: Stefano Gervasoni and the beginning of *Gramigna* », *Musicae Scientiae*, 16 (3), 262-285.

Hommel, B., Müsseler J., Aschersleben G., Prinz W. (2001), The Theory of Event Coding (TEC): a framework for perception and action planning, Behavioral and Brain Science, 24 (5), 878-937.

Hutto D., and Myin, Erik (2012) *Radicalizing Enactivism: Basic Minds Without Content*, Cambridge MA, the MIT Press.

Johnson-Laird, P. (1991), Rhythm and Meter. A Theory at the Computational Level, *Psychomusicology*, 10:88-106.

Leman M., P. J. Maes (2015), Music perception and Embodied Cognition, In: Shapiro L. ed. Routledge Handbook of Embodied Cognition, London and New York, Routledge, 81-89.

Lerdahl F. and R. Jackendoff (1983) A Generative Theory of Tonal Music, Cambridge MA, The MIT Press.

Maes, P. J., M. Leman, C. Palmer, M. M. Wanderley (2014), Action-based effects on music perception, *Frontiers of Psychology*, 3 (4):1008

Reboul, A. (2011) Radical embodied cognition vs. "Classical" embodied neuroscience. *The Journal of East China Normal University*, 6.

Shapiro, L. (2010) *Embodied Cognition*, London and New York, Routledge.

Phillips-Silver J. and L. Trainor (2005) Feeling the Beat: Movement Influences Infant Rhythm Perception, *Science*, 308, 1430.

Phillips-Silver J .and L. J. Trainor (2007) Hearing what the body feels: Auditory encoding and rhythmic movement, *Cognition*, 105, 533-546.

Phillips-Silver J. and L. J. Trainor (2008) Vestibular influence on auditory metrical interpretation, *Brain and Cognition*, 67, 94-102.

Trainor L. J. and Gao X. (2009) The primal role of the vestibular system in determining musical rhythm, *Cortex*, 45, 35-43.

Stephen, D., J. Dixon and R. Isenhower (2009) Dynamics of representational change: Entropy, action, cognition. *Journal of Experimental Psychology: Human Perception and Performance*, 35 (6), 1811–1832.

van Rooij, I., R. Bongers and W. Haselager (2002) A non-representational approach to imagined action. *Cognitive Science*, 26, 345-375.

Walton, A., M. J. Richardson, P. Langland-Hassan, A. Chemero, A. Washburn (2015), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society.*

Wilson, M. (2002) Six views of embodied cognition, Psychonomic Bulletin & Review, 2002; 9(4), 625-636.