Bidirectionally Re-Engineering the Braindance

According to the ideomotor principle, sounds are recognised by a model of their generation. So the cognitive processes that occur when listening to Billie Holliday sing should be the same processes that would lead to singing just like her. Unfortunately, that is not the result. It is a probabilistic model at a certain level of abstraction, it grasps the gist of the singing by making an approximate model of its generation, and most people will not have the fine degree of motor control necessary to sound anything like Billie Holliday. We might then think that only Billie Holliday can properly listen to Billie Holliday, since her ideomotor representations will exactly correspond to the generation of the sound. But exact models are not necessarily the best models. The usefulness of a model - its epistemic or pragmatic value - is relative to the use that it is put to. There is no optimal perspective when it comes to modeling complex phenomena such as this. Someone who is completely incompetent at singing may grasp perceptual details or conceptual implications that a great singer could not. Wiese argues that ideomotor representations may constitute examples of the kind of mediating representation that HiTEM calls for, since they are 'gist-like' motor diagrams that are neither properly perceptual nor yet conceptual.[1] Moreover, we might say that they are models whose epistemic or pragmatic values are open to limitless redescription.

Since SMA and pre-SMA responses are less pronounced for unfamiliar or random sound sequences than familiar or rhythmic sequences, we should expect that the more that music tends towards complex unpredictable structures and noise, the less it will elicit ideomotor diagrams. However, several factors may go against this general bias. Firstly, unfamiliarity is relative to prior experience so 'experts' in the generation of noise should have stronger responses, and listeners accustomed to such auditory sequences will also have a greater depth of resolution in their ideomotor image. Secondly, attentional dynamic will alter the amount of processing dedicated to the signal, so if 'noise' is attended to *as* music then we should expect SMA and pre-SMA responses. Thirdly, if the sequence is treated as a salient source of uncertainty then computational resources will be recruited to minimize prediction error. Where unpredictable sensory signals cannot be explained away by low-level hypotheses the signal will ascend the hierarchy demanding more ever more abstract modes of analysis by synthesis as it travels upwards. It will be grasped both at the highest level of generality possible (maximise the entropy of the proposal) and with the highest degree of accuracy pertinent to the task-relevant segregation of events.

Lima et al. effectively put forward something like a 'muscular bonding hypothesis' but at the subpersonal level of neuronal activation patterns. That is, the common coding binding perception and action, and the positive affects accompanying the expressive entrainment of the body to the acoustic signal, is also a means of wider social commonality, coordination, and convergence. However, we should be wary of conflating uncertainty reduction at the social level with the individual level, since there are major differences in their dynamics. Before engaging with this problem lets look at how producing and listening to music can be understood in terms of uncertainty reduction and inference to hidden cause.

In many cases of perception, if the inference to hidden cause is uncertain, active inference can be called on to minimise surprisal. However, this is often not possible with sound, and not helpful with music. If there is a persistent auditory signal then we can perform active inference by varying our position with regard to it, but many sound object-events are too fleeting or far away, and varying our position does little to reduce uncertainty. We cannot pick up a sound object and move it around in the same way as other objects, though we can manipulate the causal source and the resulting sensory signal, and this is just what musicians do. Generating musical sound requires the predictive control of the body and (often) its interaction with some material locus of sound generation, both of which are inferred causal structures beyond the Markov blanket - e.g. the kazoo must be coordinated with many parts of the body including lungs, diaphragm, vocal chords, lips and tongue - so the operative dynamics of all these components must be hypothesised. Those predictively controlled interactions enable the model of the inferred causal structure (i.e. the action opportunities

[1] Wiese, W. (2017). Predictive Processing and the Phenomenology of Time Consciousness - A Hierarchical Extension of Rick Grush's Trajectory Estimation Model. In T. Metzinger & W. Wiese (Eds.). *Philosophy and Predictive Processing.* MIND Group. p.15

afforded by the kazoo) to be enriched by a series of interwoven perceptual inferences and active inferences, this is what Clark calls the 'motor-informational weave'.[2]

If, according to the maximal PP account, all action is active inference, what uncertainties are being reduced by the actions of a musician? At the simplest level, the musician predicts a certain sensory signal (e.g. the next note in the piece they are playing), and uses active inference to bend the world to fit its predictions. If the note doesn't sound like prior expectations there will be an error signal, and the musician can try again. It is quite clear that learning such as this reduces long-term average surprisal by fitting a model to a domain and increasing fine motor control based on this ongoing trial and error. However, if the musician is improvising there is no explicit representation of what the next sensory state should be, so error has to be assessed on the basis of a model at a higher level of abstraction. If the improvisation occurs within the context of traditional jazz, then error is evaluated according to the more or less explicit rules of the genre, which are structured around the maintenance of the key signature and the rhythmic consistency of the piece. The set of moves permitted by free jazz is much larger, so error checking requires a more abstract model. In free improvisation the space of possibilities is maximised, so that it may sound like there are no rules structuring what is and isn't permitted, but this is because the model is at an even higher level of abstraction, and may be continually changing.

Perception is directed towards inferring hidden causes of sound signals because this is the best way of reducing uncertainty (or predictively maintaining homeostasis according to FEP), however in many cases, especially when the object in question is artifactually constructed as in the case of music, the causal structure of the sensory signal will be analysable at a number of different levels. For example, in the case of listening to a recording of a live concert, the cause of the music can be thought as the apparatus now generating the sound, the instruments and gestures employed in the original performance, the medium in which it is stored, the mind of the composer, the socio-economic conditions of its distribution as a commodity, and it's socio-historical constitution within a repertoire of recognisable musical gestures. In general, the level of causal analysis that would be pertinent to the predictive processing of music seems to be the latter. That is, in order to predict the sensory signal the best cues are likely to be internal to the semiotic structure of the music rather than in some extrinsic condition such as the causal structure of the instrument or the musician's body.

There are at least three levels pertinent to the question of how we are able to extract these cues and 'follow' music: from low-level regularities, to mid and high-level regularities. In recognising an auditory sequence as music, there is a bidirectional inferential cascade. Top-down, an event segmentation occurs, priors relevant to musical cognition kick in, and more sore specific context-sensitive and task-dependent priors are activated as the sequence unfolds. Bottom-up, the prediction error signal ascends the hierarchy with attention modulating the precision weighting as it goes. At the low level we have the rhythmic and spectral regularities that can be extracted at various spatio-temporal scales. In auditory cognition, as we've seen already this can be decomposed into several nested scales ranging from the sub-millisecond level to the consciously apprehended scale of the specious present and to wider conceptually mediated predictions. Mid-level priors concern the recognition of generic idioms, identifiable styles, and the interpretation of semiotic and semantic orientational cues. At the high end, prediction error is minimized because the auditory signal is addressed according to a prior related to sound-that-ought-to-be-taken-as-music. This may even be thought as a hyperprior, that is a prior at a high level of generality or abstraction that forms the condition for multiple priors to form on its basis, thus entrenching itself in the inferential economy. As we've already seen the ancestry of this hyperprior dates back to the Ice Age, in the communicative bifurcation between protomusic and protolanguage.

Metzinger argues that because humans have a running epistemic agent model (EAM) they have two distinctly coded layers in their phenomenal self-model; the spatial and temporal model of the non-neural body, and the inferentially encapsulated processes of the neural body, which are temporal but no longer tied to the spatial frame of the non-neural body. This is interesting because it may be that the two layers of the phenomenal self model relate to two different listening modes or experiences (écouter and entendre), and that associating with the transparency of the EAM is what underlies the tendency to listen to music as if it were a-spatial and abstracted from causal analysis. That is, when listening to music attention can either be directed exogenously, towards the causal structures involved in sound generation; or endogenously, towards the regularities and singularities internal to the musical structure. These are not

---

[2] Clark, A. (2017). How to Knit Your Own Markov Blanket: Resisting the Second Law with Metamorphic Minds. In T. Metzinger & W. Wiese (Eds.). *Philosophy and Predictive Processing.* MIND Group. p.14

mutually exclusive modes of listening however, so there is no reason besides Schaefferian dogmatism to eradicate the former in favour of the latter. We might call the condition of mistaking the perspective from the EAM as given (i.e. failing to acknowledge the processes that are generative of this phenomenal experience) as 'Scruton's ear' (an affliction that is strongly related to class and race).

Most listening to music is omnivorous in the sense that it will feed on any available information to make predictions, and the principal resource is generally afforded by the cognitive scaffolding of the body. Just as children learn to count with their fingers, we follow music by moving our bodies. There are good arguments that auditory cognition is not spatio-temporal in and of itself, but rather takes place in 'pitch-time', and is immediately spatialised (at the sub-millisecond time scale) only by subservience to a multi-modal model dominated by vision.[3] Nevertheless, the spatialisation of the auditory signal proximally received at the ear is crucial to its ongoing prediction and interpretation. In the case of a musical sequence the identification of the causal locus of the sound event is less expedient for prediction than its translation into spatialising gesture at some level of abstraction. Whether or not we actually move when we hear music, the areas of the brain involved in motor control are activated, so even when Scruton listens eyes-closed to his favourite sonata he does a little braindance. There is a very compelling relationship between the activity of the musician and the activity of the brain of the listener. Of course, there are statistical correspondencies between neural activity and any external events that the cognitive system is tracking. However, with auditory information the temporal scale of the regularities of the stimulus are matched by the temporalities of neural processing. This means that the musician has her hands directly in the brain of the listener, like a spatio-temporally remote puppet-master with invisible strings. But the odd thing about this puppetry is that each time a string is pulled there is a different movement because the relation is not straightforwardly causal. The puppet autonomously recapitulates the statistical regularities of its master by predictively modeling and interpreting its movements.

The causal structure of an analogue instrument such as a bagpipes is incredibly complex, and allows for practically unlimited exploitations and explorations, however it always remains within the bounds of its timbre profile. Electroacoustic and electronic instruments allow for a much wider possibility space, effectively any sound within a certain audible range can be produced. With granular synthesis this capacity is extended down to micro-compositional elements below the threshold of audition. This means that the exploration is not so much of the causal structure of the instrument (though of course there is always a specificity to the technical interface) but of the abstract possibility space of sound. Nevertheless, every instrument is in some sense a materially constrained portal into this abstract possibility space, a '*synaesthetic exploratory probe* multimodally plugged into the sensorium'.[4] We can think of this according to the Piercean conception of epistemic convergence in the long run. That is, there is a progressive unfurling of the possibility space, an ongoing collective elaboration that grasps the absolute possibility of music. As we've already seen in other contexts, the possibility space cannot be prestated since it is constantly undergoing symmetry breaking transitions that transform the whole possibility space. The activity of making music does not grasp something that already exists, as in the case of the scientific understanding of space and time, but it grasps in the revisionary-constructive sense that it discovers and invents those possibilities.

Its interesting to think of this activity within the theoretical framework of PP or FEP in relation to the activities of animals in general. According to PP any action can be thought as both a hypothesis and its experimental testing. For example, the action of a flying squirrel could be posed as a hypothesis: 'if I jump and spread my arms maybe I can land on that tree over there and avoid this predator here'. Of course, the squirrel does not formulate this as a proposition, and even if its activity might have a normative form 'if a predator climbs my tree I ought to jump', the squirrel does not have the normative freedom of a rational agent that commits to a rule for a reason. The whole history of human musicking might be thought as an investigation of a socially instituted and normatively coordinated abstract hypothesis space. Every act of musical sound generation is in some sense a normative assertion of the kind 'music could be this'. Every sensory signal must be interpreted, so if a squirrel hears a predator it responds with a certain activity (evasive manoeuver). The interpretative degrees of freedom for a musical signal are of a different order. Every dance move can be thought as a hypothesis about what kind of action *ought to* or *could* correspond to the 'transposable aboutness' of the musical signal.

---

[3] Kubovy, M. & Van Valkenburg, D. (2001) Auditory and visual objects. *Cognition* 80 97±126.

[4] Catren, G. (2016) What is Ensoundment? In *Formulations.* Koenig Books.

That perception is tied to action through the ideomotor principle is clearly relevant to an understanding of dance. We've already seen the historical importance of dance in the communicative bifurcation between musicking and protolanguage, and in the cognitive development of children (the cross-domain transfers and cognitive fluidity induced by representational redescriptions). If we understand music to have emerged from this release from proximity where the signal is no longer interpreted according to the inferred immediate causal structure of sound production, nor according to a group of syntactic rules and precise semantic determinations (as in language), but heard according to the ancient ritualistic invocation of radical uncertainty, activating a hyperprior that grants the signal floating intentionality; then dance is precisely the expression of this xenopoietic interpretative openness, where sounds are transduced and translated into abstract ideomotor schemas. Dance is the active inference that explains away the radical uncertainty of the musical signal.

Uncertainty is about the inferred cause not the signal, and maintaining oneself in a homeostatically restricted parameter space requires understanding the conditions of possibility for remaining within that space. Since these conditions are contingent and liable to fluctuation, it is expedient for the system to reduce uncertainty regarding the conditions by moving outside of the already explored parameter space. Uncertainty reduction then demands some risk, not just the Bayesian rational 'gamble' on the inference to hidden cause, but an expansion of the parameter space (via perceptual and active inference) that is both intra-umweltic (empirical variations conditioned by the transcendental structure of the umwelt) but trans-umweltic (transcendental reflection on the conditions of possibility of experience). Exploitation and exploration have then to be thought of as referring not to already existing possibilities but as transforming the space of possibility (either or both internal states and external states) for action. Uncertainty reduction is therefore an active impetus for the cascade of symmetry breaking transitions

The PP framework provides an explanation of the problem of sound-in-itself that is very different from typical humanities arguments based on fundamental epistemic limitation (e.g. Kane or Bonnet). An explanation of the acousmatic reduction according to PP would first acknowledge that the intentional stance adopted activates a set of priors that treat the auditory sequence as in some sense self-causing, since explanation via reference to the causal structure of sound generation is bracketed. This is not an unprecedented stance. When we observe the movement of an animal we do not seek to explain this by referring to some cause other than the motivations of the animal itself. In order to explain away the sensory signal of an animal's movement we can use the 'folk psychological' language of beliefs and desires, or some scientific account of its behavioural tendencies. In the case of music there are various levels of causal analysis, such as the aesthetic choice of the composer or the internal semiotic structure of the piece. Acousmatic reduction explains away the signal by focusing on the morphological characteristics of the auditory sequence itself. Because the sound object is actively attended to, the gain on the sensory signal is increased, meaning that effective connectivity (computational processing power) is increased, and the sensory signal ascends the hierarchy without lower-level predictions explaining it away. Nevertheless, the conceptual framework provided by Schaeffer's taxonomy can be thought of as a comprehensive structure of top-down prior expectations that steers the processing of the signal.

We cannot simply dispense with prior expectations and grasp the signal in its purity. For Bonnet this means we are always trapped between sensibility and intelligibility, the empire of discourses (total audibilization) and the pure sonorous continuum, and the only way to re-establish access to the sonorous event is through embracing the savage ineffability or disruptive potentiality of the imperceptible unheard in its radical uncertainty. Bonnet rejects the possibility of sound-in-itself but asserts a deterritorialized listening that grasps the imperceptible becoming of the sonorous. As we know from Deleuze and Guattari's original formulation however, deterritorialisation leads very quickly to reterritorialisation. Deterritorialisation might be thought within the PP framework as the relinquishment of a habitual mode of identification according to pre-existing priors. But if all priors are withheld and no stable recognition distribution is formed then there is no disruption to the framework of perception and conception. Uncertainty reduction is merely shifted to another level, or rather the sonorous is grasped according to a prior that negates all other priors: the belief that an anarchic listening is possible.

PP makes the opposite claim, that sound-in-itself, or the causal structure of the world, is revealed in the revisionary-constructive process of modeling in the long run. This does not entail the negation of priors but their careful context-sensitive and task-related calibration. On top of this PP claim, which implicitly stays within the remit of intra-umweltic variations, in order to hear sound-in-itself, trans-umweltic variations or transcendental reflection is necessary. Doing that requires drawing on all the listening modes, cycling across the Ceno-Pythagorean categories, initiating cross-domain transfers and maximising cognitive fluidity. It means a complex weaving between polarites such as real and ideal, sensation and conception. If we cannot listen to sound-in-itself, it is not because it 'doesn't exist' or because we negate the concept, but because sound and music require a multi-level ontology and the integration or productive incoordination of contrasting perspectives. Catren, Negarestani, and Zalamea contend that a musical practice such as Hecker's does just this.[5]

Lastly, in considering how the PP framework impacts on an understand of the socially instituted phenomenon of music we might formulate a version of the dark room problem for music. That is, if surprisal minimisation is the goal why doesn't all music have the completely predictable, highly compressible structure of low-stimulus mulch like muzak? As we saw earlier, there are two ways of answering this. Either we reject the claim that all perception and action is directly driven by uncertainty reduction, or we reject the claim that low stimulus means low surprisal. These are not incompatible options. The best way of articulating the former option is to argue that uncertainty is relative to the scale of analysis and the relevant priors, so actions (such as listening to noise music) that increase surprisal at one scale may reduce it at another scale or in the long run. This does not go against the second option, held by maximal PP and FEP, which could be rephrased in much the same way: what counts as surprisal is relative to the set of already held prior expectations, and since we expect a certain amount of unpredictability in the structure of music, those kinds of music that are highly predictable are actually surprising.

The problem is also very similar to one we encountered with regard to biological evolution: how do we account for the increasing complexity of biological lifeforms considering the thermodynamic tendency of systems towards disorder? There the answer put forward was that there is a contragrade movement (against the flow towards increasing disorder) that occurs due to the build-up in the occupation of possible environmental niches at the lower level of complexity, and this drives an asymmetric diffusion towards increasing complexity. We can call this process anti-entropy because it does not merely resist entropy in negative terms (i.e. mitigate entropy by staving it off) but is a positive result that does not refute the law of thermodynamics but is nevertheless not described by it.[6] In musical terms, if more predictable structures of music have already been composed then composition is pushed towards the occupation of ever more complex forms. There does seem to be a shortfall of FEP in the sense that it supposes uncertainty reduction at the scale of the individual to be the same as that at the level of social systems. What FEP fails to account for are the many levels of dynamics involved in determining which uncertainties are to be reduced and how; including social norms, affective contagion, socio-economic and information asymmetries, political discourses, conceptual arguments, the management of perception by psychological manipulation, etc.

Another way of answering the dark room problem is to point to two factors in the bidirectional dynamics of PP: Firstly, due to its profuse complexity there is an inconsolability of the empirical world to prior expectations so that a lasting fit of model to domain cannot be established, whether by finding a maximally predictable place or formulating a maximally predictive model, meaning there will always be the possibility of an error signal ascending the hierarchy. Secondly, there is no lasting solution to the generative model in the other direction because hypothesis generation proliferates in predictable situations, moreover top-down expectations guide actions and these alter the environment, leading to new error signals.

---

[5] Hecker, F. (2016) *Formulations*. Koenig Books.

[6] Bailly, F. & Longo, G. (2009) Biological Organisation and Anti-Entropy, in *Journal Biological Systems*, Vol. 17, No. 1, pp. 63-96.