DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

Towards a computational account of art cognition: unifying perception, visual art, and music through Bayesian inference

Jorge Aurelio Menendez; CoMPLEX, University College London; London, UK

Abstract

Being an exclusively human ability, art cognition occupies a unique position to inform our knowledge of human cognition. To do so, however, it must be understood in computational terms generalizable to other cognitive processes. Here, such an account of art cognition is sought. Specifically, I argue that Bayesian inference drives art cognition, by showing how it may underlie music and visual art perception. The resulting account provides the sketches of an economical and unified theory of art cognition consistent with contemporary theories of perception.

1. Introduction

How and why we experience art has long been an elusive question in cognitive science. Art seems to appear universally across human cultures in many forms [1], yet we lack a firm understanding of how and why we treat certain stimuli as art. Why do certain harmonic modulations feel so exciting? What makes observing a landscape feel so different from looking at a painting of it? Such questions remain outstanding in contemporary cognitive science. Moreover, in addition to being interesting in their own right, they also hold potential to shed light on more general properties of the human brain – the only brain we know that creates and enjoys art.

For answers to these questions to be satisfying and useful, however, they must be articulated in a language that can situate them within the broader context of human cognition. Modern cognitive science presumes cognition is a product of computation, so art cognition ought to be explained in such terms as well. Moreover, the language of formal computation can describe activity at the level of neurons and the level of behavior, allowing us to extrapolate from the latter to the former and vice versa. [2]

The question thus becomes, what kinds of computations give rise to the experience that is art? The purpose of this paper is to suggest an answer. Importantly, rather than focusing on the sensory processes necessary to see or hear an artwork (e.g. color perception, pitch perception), I focus on how the outputs of such processes may be exploited to experience some stimulus as art. This process is what I will call *art perception* (e.g. listening to music), as opposed to *sensory perception* (hearing sounds). Experiencing any artwork requires both: *sensory perception* provides a percept of the piece that is then processed in a certain way such that it produces the experience. The computations involved in these latter stages – *art perception*[1] – are the focus of this investigation.

A natural place to begin, then, is sensory perception itself. Much work has focused on uncovering the computational principles governing sensory perception, so can we extrapolate these principles to explain art perception? I will suggest that the answer is yes. Specifically, I will look at a particular hypothesis known as the *Bayesian brain hypothesis* [3] that makes specific claims about the computations underlying sensory perception and the rest of cognition. I will argue that it has the potential to also explain art perception. I will propose that the same mathematical formalism – Bayesian inference – could underlie both sensory perception and art perception.

I begin by discussing the Bayesian brain hypothesis and its central ideas. In §3, I show that it can be applied to explain music perception, tying in the mathematics of Bayesian inference with the ideas of late musicologist Leonard Meyer. I go on to discuss how it could also explain our appreciation of visual art in §4 and finish by discussing what the import the Bayesian brain hypothesis can bring to art cognition and what future work is necessary. I emphasize here at the outset that my approach is purely theoretical.

2. The Bayesian brain hypothesis

In recent years, much attention has been given to the hypothesis that the brain is built to compute probabilities. Within perceptual science, empirical motivation for this idea comes from evidence showing that perceptual measurements in the brain randomly fluctuate and underdetermine their causes, such that our perceptual system must have a means for inferring the structure of distal stimuli from a noisy and indeterminate proximal signal. [3]

One such way of doing this is *Bayesian inference*, a mathematical formalism for evaluating and updating the probability of an event in the face of new relevant information. It is based on a formula for computing the probability of a hypothesis H, conditional on some data D – '*the probability of H, given D*'. It is called *Bayes' rule*:

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)} \tag{1}$$

Provided some set of hypotheses (e.g. presence/absence of edges) and relevant observations (activation of retinal ganglion cells), this formula yields the probability distribution of hypotheses that takes into account the observed data – the *posterior probability distribution* $p(H|D)$. In other words, it provides a way for evaluating which of the hypotheses under consideration are more or less likely under the observed data. The same formula can be used to update these probability evaluations in the face of new incoming data. Importantly, it can be shown that it does so optimally, in the sense that it always converges to the true probability distribution [4] and it generalizes better than any other method. [5] But the true power of this formalism is that anything can be substituted for H and D, provided that the corresponding *prior* $p(H)$ and *likelihood* $p(D|H)$ functions are known.[2] As such, Bayesian inference has been used to successfully model various cognitive

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

abilities [3, 6, 7]. Beyond a formula, the Bayesian brain hypothesis is a framework for thinking about the computations the brain performs to do what it does, the claim being that it computes conditional probabilities in accordance with Bayes' rule.[3]

In addition to the wealth of empirical results showing that humans behave in a Bayesian manner [6] (i.e. in accordance with what would be predicted by Bayes' rule), strong theoretical reasons exist for why the brain should carry out such computations. Computing precise predictions under uncertainty is essential for the survival of any organism. Any decision requires predicting possible outcomes and evaluating which are more or less likely. Safely navigating an environment requires updating expectations about where dangers may lie. And, at a lower level, biological perceptual systems are so noisy that they demand a probabilistic inference of what is causing the sensory signal, since several causes may be consistent with it. [3] Because of the uncertainty inherent in each of these problems, probabilistic computation is particularly well suited for solving them. Bayesian inference simply provides the optimal method for computing the required probabilities.

These inherent properties of all organisms that move and perceive have led some to suggest that computing predictions is in fact the primary purpose of the brain. Neuroscientist Rodolfo Llins has argued that brains evolved specifically to predict the outcomes of motor movements, proposing that the higher reasoning of the modern human brain is an evolutionary derivative of this primitive function. [8] And Friston's Free Energy Principle claims that all of neural computation hinges on minimization of prediction error. [9] In this same vein, I will argue that prediction underlies art perception. The Bayesian brain hypothesis provides a framework for thinking about how this could be so.

It is an important caveat, however, that Bayes' rule tells us nothing about how neurons carry out Bayesian inference. Indeed, this is one of the questions currently at the forefront of the Bayesian approach [7] and several promising proposals exist. I will discuss one possibility called predictive coding, but there are many others (e.g. [10]). The point here is to clarify that, rather than a discussion of how the *brain* gives rise to art perception, this paper is chiefly a discussion of how *computation* gives rise to art perception. I will propose the culprit is a particular kind of computation – Bayesian inference. How the *brain* carries out Bayesian inference is a separate problem that is in fact being dealt with extensively today. [7] I will only hint at some possible answers to this secondary question.

I will now go on to explore how the Bayesian machinery that has been applied to explain sensory perception could also explain how we appreciate music. I begin with music perception because the role of probability and prediction in this domain has been previously explored. I will show that this previous account in fact aligns elegantly with the Bayesian brain hypothesis.

3. Prediction in music

Leonard Meyer first proposed that probability plays a central role in music perception in a seminal paper on meaning in music. [11] His thesis was that, akin to other behavioral phenomena, the perception of music can be reduced to prediction: enjoying a piece of music consists of predicting what will come next and subsequently verifying whether those expectations are confirmed or violated. For example, Richard Wagner's notorious Tristan chord in *Tristan und Isolde* (1865) has the effect it has because the listener constantly expects a harmonic resolution that doesn't come until the climactic *liebestod*.

His theory thus presupposes probability judgments in music perception – judgments that guide the predictions and expectations of how the music at hand will unfold in real time. Could these computed probabilities be the output of Bayesian inference, just like those that have been proposed to underlie sensory perception? In Meyer's terminology, listening to a piece of music consists of the continual processing and updating of two elements: the *antecedent* – what we have heard – and the *consequent* – what we expect to hear. Music perception consists of predicting the consequent based on the antecedent. According to Bayes' rule, this prediction should be based on the posterior probability computed as follows:

$$p(consequent|antecedent) = \frac{p(antecedent|consequent)p(consequent)}{p(antecedent)} \quad (2)$$

In typical Bayesian terminology, the *data* is the music heard until now – the harmonic, rhythmic, and melodic progression of the music – that informs the probability distribution over *hypotheses* about what will come next – one's expectations of what chord, duration, or pitch will follow. Provided the space of possible consequents (the *hypothesis space*) and the prior and likelihood functions, Bayesian inference gives us a way of predicting the consequent from the antecedent.

So what are the hypothesis space, the prior, and the likelihood function? And, how are they determined? Meyer's answer is that they depend on the musical style(s) one has been familiarized with. In a given style of music, certain features are more or less likely to follow others – in Western music, for example, a tonic chord is likely to follow a dominant seven chord. Also, the range of values subsumed under a particular feature space may differ from style to style – the pitches used in Western tonal music are always taken from the chromatic scale, but Indian music utilizes microtones. [12] Given the surprising ability of infants and adults to learn the statistics of their auditory environment [13, 14], we should expect such statistical regularities to be internalized by listeners such that their musical predictions are sensitive to the statistics of the style(s) they have been exposed to. For example, a Westerner's posterior distribution over harmonic consequents for an antecedent consisting of a dominant seven chord should generally peak at the tonic[4] (i.e. $\arg\max_\theta p(\theta|\mathrm{V}^7) = \mathrm{I}$). Additionally, the domain of this function (the hypothesis space) should be restricted to harmonies that are used in common practice tonal Western music (e.g. should exclude cluster chords).

It is easy to see how such functions could be formalized as likelihood and prior probability density functions in the Bayesian framework. The above posterior distribution, for example, would be derived by computing $p(\mathrm{V}^7|\theta)p(\theta)$, where $p(\mathrm{I})$ and $p(\mathrm{V}^7|\mathrm{I})$ would be quite high for a Western listener, who would have been exposed to countless musical stimuli in which there were many instances of a tonic chord (I) preceded by a dominant seven chord (V^7), and many tonic chords spread throughout.[4] It is thus plausible that music perception in a given style is governed by likelihood and prior functions sensitive to the statistical regularities of that style, which are acquired and tuned through extensive exposure to it. Computing these functions to arrive at the posterior, however, is non-trivial, as the likelihood and prior functions may

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

take complex forms that can make the computation intractable. [15] A formally equivalent alternative that has been proposed to underlie other parts (if not all) of cognition is called predictive coding. [9, 16] The idea behind this implementation of Bayesian inference is that two streams of information are simultaneously active, one encoding current predictions of the system (e.g. the visual system) and the other encoding the prediction error (the divergence between the current prediction and the current empirical observation, e.g. current retinal stimulation). As the system continually makes observations and predictions, the prediction errors feed back to the prediction stream, tuning it to the statistics of the current environment being observed and thus appropriately altering the predictions to "explain away" the prediction error. [16] This computation is in fact formally equivalent to Bayesian inference, where the top-down predictions serve as empirical priors in inferring the posterior distribution over what observations are to be expected. [9] Predictive coding models have successfully explained several phenomena in vision, including extra-classical receptive field effects in the retina [17] and binocular rivalry [18], among others (for a review, see [16]). Furthermore, neurophysiological evidence exists indicating the presence of two neural processing streams encoding predictions and prediction error, respectively. [19, 20]

Putting aside technicalities of how Bayesian inference would be implemented in this context, the Bayesian account of music perception I am illustrating explains several phenomena seen in listeners across cultures and musical styles. The idea that a given listener's musical predictions are dependent on the musical styles she is familiar with can account for why a listener exposed to only, say, Western music will struggle to hear any logic in microtonal Indian music. This account also correctly predicts that novices and experts should differ in their perception of music [21, 22, 23, 24] (although, see [25]), since more exposure to a musical style should theoretically lead to a posterior distribution more sensitive to subtle musical elements in it (e.g. modulations, motivic manipulations). These observations reflect the fact that music perception is to a great extent driven by the integration of perceptual information with prior knowledge, a computation that is neatly formalized by Bayesian inference. What may be music to some may be noise to others, and this is determined by previous experience that shapes the prior and likelihood distributions. Noise occurs when these functions can't work with the provided antecedents and consequents; music arises when they output a rich palette of predictions by exploiting viable consequents and informative antecedents.

Of course, providing the perceptual material that originates this palette is the job of the composer. Doing this well requires an intimate understanding of the statistical regularities present in the musical style being composed in – working within these regularities is precisely what it means to write a piece *in* that style. The key point here is that a skilled composer can use these regularities to exploit listeners' musical intuitions to manipulate their expectations. By remaining within the confines of the style and veering away from it from time to time in a carefully balanced way, a composer can maintain the informativeness of the antecedent so as to simultaneously guide and violate the listener's expectations. Some examples of this (borrowed from [11]) are provided in table 1 below, embellished with the Bayesian terminology I introduced earlier.

Table 1. How composers manipulate and violate listeners' expectations

Compositional technique	*Explanation in terms of Bayesian inference*	*Musical example*
Delaying the expected consequent	The consequent assigned highest probability does not occur next, but it remains the most likely as the music continues until finally arriving at that consequent	*Tristan und Isolde* (1865) Richard Wagner: Tristan chord that doesn't resolve until the very end of the opera
Ambiguous antecedent	The posterior probability distribution over possible consequents is relatively flat – many consequents are predicted with equal probability	*Mazurka Op. 24 No. 2* (1836) Frederic Chopin: opening measures ambiguous as to C major or G major
Unambiguous antecedent followed by unexpected consequent	Observed consequent is assigned (by the posterior distribution) a very low probability relative to other possible consequents	*Symphony No. 3 in Eb Major, Op. 55* (1806) Ludwig van Beethoven: C# in opening Eb major triad theme

I have argued that Bayesian inference could underlie music perception and has potential to explain several music perception phenomena. But do we have any empirical evidence that Bayesian inference is really happening in these cases? The idea that predictions are central to music perception has intuitive appeal for any musician, and much scholarly work has supported it. But concrete evidence demonstrating that these predictions are computed in a Bayesian manner will require Bayesian models of music perception to be implemented and corroborated with behavioral results. Bayesian models have been applied to several areas of music cognition [26], but not to musical predictions. Two main challenges arise here. On the theoretical front, we need to formally define possible prior and likelihood functions, or hierarchical predictive coding models. One way of tackling this is by uncovering the statistical regularities of particular styles of music (e.g. [27]), regularities that should be reflected in the prior and likelihood functions of a listener indoctrinated in that style. Or, musical grammars may be proposed that translate to particular prior and likelihood distributions. Alternatively, hierarchical Bayesian models (e.g. predictive coding models) can be trained on musical examples of a given musical style and subsequently analyzed to examine the musical predictions they make.

On the experimental front, we need a way of measuring listeners' actual predictions to corroborate and inform our models. Directly determining a listener's prediction in a given instance may be impossible, but there are many ways of testing the agreement between the musical predictions of a model and of a human. [28, 29, 30] For example, it has been shown that musical expectations are predictive of priming effects in reaction time experi-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

ments. [28, 30] Thus, if a Bayesian model predicts that chord *x* is more likely than chord *y* to succeed some presented chord progression, then a priming experiment with humans should show greater priming effects for *x* than for *y* following the presentation of that chord progression. Alternatively, listener predictions may be inferred from other measurable variables such as musical tension [26, 31] or emotional response (i.e. from a violation or dramatic satisfaction – Wagner's *liebestod*, for example – of a prediction), which can be measured with galvanic skin response. [23, 24]

A significant prediction that falls out of several instantiations of the Bayesian theory of music perception that I am presenting is that any given human should be able to learn any musical system. If we consider the predictive coding proposal claiming that our musical predictions are tuned to the statistics of the musical styles we are familiar with, it should be theoretically possible to train anyone to be able to enjoy and appreciate any type of music (as long as it contains the necessary elements for expectations and predictions to be made). On the other hand, the proposal that the likelihood and prior functions are determined by some kind of learned or innate musical grammar does not entail this prediction at all. This claim can be tested by taking artificial language approaches frequently employed in linguistics. [32]

Another empirical approach is to continue the search for musical universals. [33] One of the features that makes the theory I am sketching here so attractive is that it should apply across all musical styles and cultures (and all visual art forms, see below). Hence, musical elements crucial to allowing the formation of predictions and expectations should exist in all musics in the world. One proposed universal that supports this idea is the use and repetition of motivic units, which can happen at multiple musical levels. [34] Relatedly, all musical systems seem to be isorhythmic: once a rhythmic pattern is established, it is continued throughout the piece (e.g. time signatures in Western music). [34] It is also worth noting that any listener of any culture will always believe that the music he is familiar with sounds the most *natural*. [33] This easily falls out of my proposal that a listener's music perception machinery is tuned to the statistics of the musics they have been exposed to.

Focusing in on Western music, a significant observation that may reflect the central role of prediction in music perception is the historical development of Western composition. Western compositional innovation has followed a trend seemingly directly related to the facilitation and manipulation of musical prediction. Early polyphonic music followed strict rules that facilitated prediction; in the medieval *organum*, for example, tritones between overlapping voices were always avoided [35], ensuring that the consequent would not contain one (thus making it more predictable). Conversely, Beethoven and Brahms (600-700 years later) embraced dramatic harmonic modulations to unpredictable harmonic areas, frequently employing tritones to do so (e.g. Beethoven late string quartets). In the latter half of the 19th century Richard Wagner's operas flaunted his ability to keep listeners' predictions unsatisfied for long periods of time (e.g. table 1, 1st row). Decades later, in the 1920s Arnold Schoenberg invented the twelve–tone technique that influenced many later 20th century composers; this technique emphasized the emancipation of music from tonality, thus discarding tonal centers and keys[4], making it virtually impossible to make any melodic or harmonic predictions (e.g. *String Quartet Op. 28*, Anton Webern (1937-8)). Digging deeper into 20th century music, we find that many compositions that are meant to push the boundaries of what constitutes music explicitly remove the elements of music that allow prediction (e.g. *4:33*, John Cage (1952); *Pome symphonique*, György Ligeti (1962)). The trend is clear: from facilitating prediction to manipulating it to removing it. While this pattern may not necessarily imply the causal link between prediction and music perception that I am proposing (correlation is not causation!), it does suggest some relationship between them. The Bayesian account I have presented accounts for this relationship, while also explaining specific historical events such as the mixed reception of Beethoven's 3rd Symphony and the chaotic premiere of Stravinsky's Rite of Spring [35]: the listeners' prior and likelihood distributions were not yet tuned to such innovative compositional techniques (like the Westerner listening to Indian music).

To summarize, this section argues that music perception could arise from prediction via Bayesian inference. Bayesian inference has the power to explain individual differences across cultures, historical eras, and levels of expertise by formalizing the integration of perceptual information with prior knowledge. Specifically, prior knowledge of a musical style is formalized by prior and likelihood distributions that reflect the statistics of that style. This account also welcomes mechanistic proposals for how this prior knowledge could be acquired and integrated with incoming perceptual information (e.g. predictive coding). Importantly, I have suggested we aim future empirical and theoretical work towards revealing the formal properties of these prior and likelihood distributions by constructing and testing Bayesian models. This will yield the kind of rigorous evidence needed to verify the theory I have presented.

4. Visual art as Bayesian inference?

If we look closely, we find that the historical development of visual art follows a trend similar to that of music, away from the predictable. Renaissance and baroque art, for example, focused on realistic depictions of religious figures and events – a subject matter that is in a certain sense predictable, concerning actual physical objects and forms depicted as they are in reality (fig. 1, see below). Going forward 200 years, impressionists continued painting familiar scenery but now depicting it in a more abstract manner (fig. 2). Rather than presenting objects as they look in real life, impressionists aimed to create an 'impression' of the scene by depicting constant shifts of light. [36] The later cubism (fig. 3) and abstract art (figs. 4-5) styles of the 20th century depart greatly from predictability on a canvas, depicting forms that exist only in one's imagination, only vaguely evoking elements of physical reality. As with music, 20th century visual art is often explicitly concerned with challenging the audience's interpretive ability by deviating wildly from one's predictions (e.g. Dadaism, fig. 6). Again, this evidence does not necessarily imply that visual art perception relies on predictions. But the parallel historical development does suggest at least that music and visual art perception may arise from similar processes.

In fact, there is a lot more in common between music and visual art. In many Western languages, the same vocabulary is used to describe the effect of an artwork, whether it is visual or auditory. Indeed, both musicians and visual artists strive for the common goal of creating something 'beautiful', whatever that word

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

may entail. The empirical results tell a similar story: Ishizu & Zeki (2011) found that the medial orbitofrontal cortex activates proportionally to judgments of beauty in both auditory and visual domains, suggesting that the neural substrate of beauty in music overlaps with that in visual art. This result resonates strongly with the hypothesis I am putting forth here: that art perception across modalities is governed by the same probabilistic computation. Indeed, an extension of my hypothesis could be that aesthetic beauty itself is a manifestation of probabilistic computation.

What this discussion points to is the question of whether visual art perception could be the outcome of Bayesian inference, as I proposed was the case with music perception. One interesting and quite comprehensive proposal of the cognitive processes underlying visual art perception is the Leder et al (2004) model of aesthetic appreciation. The authors propose that the process of experiencing a piece of visual art consists of arriving at a satisfying interpretation of it. As demonstrated by the Dada art movement of the early 20th century, visual art emphatically relies on context: putting a toilet on a pedestal in an art museum elevates its status to more than a functional object (fig. 6). Such a context challenges the viewer to interpret the perceived object, since the artist must have deliberately placed the toilet there for some reason. According to Leder et al, inferring these reasons – whether it relates to putting common-place objects on a pedestal or using certain colors rather than others, etc. – is the information processing problem that gives rise to experiencing and appreciating visual art.

Importantly, the solution to this problem is never a one–shot process. Rather, the viewer formulates and cycles through different hypotheses until finding the one that produces the most satisfying interpretation. Leder et al call this the Cognitive Mastering–Evaluation feedback loop. Evidence for such a mechanism comes from arguments and experimental findings supporting the claim that the process of obtaining a satisfying understanding of an artwork is a rewarding one. For example, one of Ramachandran & Hirstein's (1999) laws of aesthetic experience states that the challenge of distinguishing figure from ground, or object recognition, can produce a pleasurable effect (e.g. that "aha!" moment when you see the Dalmation in fig. 7). Elaboration effects showing that viewers assign higher aesthetic ratings to art when they are given extra information about it (e.g. a title) also support this idea.[5] [40, 41, 42]

This process of selecting and evaluating hypotheses about an artwork is particularly relevant to the present discussion because the probability distributions involved in Bayesian inference simultaneously provide solutions to both of these stages. Namely, the posterior distribution can provide a hypothesis of the most likely interpretation, and the likelihood distribution provides a means to test it. Consider Bayesian inference over different possible interpretations $I_1, I_2, ...$ of an artwork with features $f_1, f_2, f_3, ...$:

$$p(I_k|\{f_i\}) = \frac{p(\{f_i\}|I_k)p(I_k)}{p(\{f_i\})} \tag{3}$$

Provided the currently observed features of the artwork $\{f_i\}$, this formula is a means for evaluating the suitability of different interpretations. Using the resulting posterior probability distribution, an appropriate interpretation can be selected by, for example, taking the interpretation I_k at which the distribution peaks.[6] However, due to attentional limits and biases, not all features can be noted at once, so some features remain to be observed to confirm the currently selected interpretation. Supposing the brain computes the prior and likelihood functions above, the search for more features can be guided by the likelihood distribution $p(\{f_i\}|I_k)$. This function is a kind of exercise in theory of mind: if I wanted to produce an artwork concerning I_k, what features would I likely incorporate?

I illustrate with the following simplified example. Suppose you perceive sad faces in a painting. The posterior distribution will yield a high probability for I_k = '*the artist is depicting sadness*', in part because the probability returned by the likelihood $p(sadfaces|I_k)$ will be high (because, presumably, you would expect someone who wanted to depict sadness to paint sad faces). You thus select the interpretation I_k. Having selected I_k, you now consider peaks in the likelihood distribution $p(\{f_i\}|I_k)$, i.e. features likely to be in a painting in which the artist is depicting sadness. In a sense you are predicting certain features to be in the painting. Two such predictions may be the presence of the feature *darkness* (low luminance) and the feature *blue* (short wavelength light), since the probabilities $p(darkness|I_k)$ and $p(blue|I_k)$ are high. You subsequently turn your visual attention to these features to corroborate the predictions. If you find that these are not present in the painting, you compute the new posterior distribution

$$p(I_k|sadfaces, \neg darkness, \neg blue) \tag{4}$$

You now select the hypothesis most probable under this distribution and restart the process, the loop continuing until a stable solution is achieved. This iteration of inferences could constitute the computations underlying the feedback loop that Leder et al (2004) hypothesize plays a central role in visual art perception.

Furthermore, these inferences need not only consider such high-level interpretations as '*the artist is depicting sadness*'. Viewing a given artwork will entail considering and selecting multiple hypotheses at multiple levels ranging from the emotional and symbolic (e.g. visual metaphors) down to the purely visual and aesthetic (e.g. color, lighting).[7] Continuing the simplified example above, in addition to considering and evaluating the interpretation that the artist is conveying sadness, one might also interpret a contour in the painting as a frown. Once you have arrived at this hypothesis (that that contour is a frown), you will subsequently test it by searching for frown-like visual features (e.g. concavity). Should you find these, you will have confirmed your hypothesis and strengthened your understanding of the artwork, leading to a rewarding experience (much like the rewarding effect of object recognition discussed by Ramachandran & Hirstein, 1999). This explanation can also account for the peak-shift effect, whereby the emphasis of an object's essential features creates a pleasurable aesthetic effect, as in caricatures [39]: after recognizing Barack Obama in fig. 8, the exaggerated features of his face have a pleasant effect – this could arise from the dramatic confirmation of the initial hypothesis that Obama is being depicted.

Moving on to thinking about how a chain of inferences like the one I am proposing could be implemented in neural hardware, it is easy to see how the predictive coding scheme easily accommodates it. We can imagine the top-down prediction signals arranged hierarchically, where the highest levels correspond to high-level interpretations (*sadness*) and the lower levels to low-level interpretations (*frown*), all the way down to in-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

dividual line segments (concave curve) and spatial frequencies. This architecture would elegantly accommodate the Bayesian account of visual art perception in a framework that makes empirical predictions even at the neurophysiological level. Furthermore, the predictive coding framework provides a mechanistic explanation of the attentional search for features under a given interpretation/prediction signal: attention simply corresponds to appropriately changing the weights assigned to the prediction error signals (in neurophysiological terms, the gain of the neurons producing these signals). [9, 16] Back to the sad painting example: when you are searching for the feature *blue*, the brain adjusts the weights of the error signals such that the error signal corresponding to color predictions is prioritized.

Just like the Bayesian account of music perception, Bayesian visual art perception can explain individual differences across cultures and levels of expertise [44] because the prior and likelihood distributions provide a way of integrating prior knowledge into the computation. For example, the likelihood distribution reflects the statistical dependencies and regularities between features and artistic intentions that a viewer has encountered in her past experience viewing and/or creating art, and through her explicit knowledge about artistic techniques and tendencies. General statistical regularities observed in every day life should come into play as well (e.g. sad people tend to frown). [45] Expertise leads to a likelihood function more sensitive to nuanced interpretations and features of visual art. The same applies to the prior distribution and hypothesis space – experts know interpretations to look for *a priori* and will consider more possibilities.

Empirically confirming the Bayesian theory of visual art perception that I have presented here will again require two lines of work: (i) constructing and implementing computational models of visual art perception, and (ii) comparing the predictions of these models with those of humans. One of the primary difficulties with the former is that the hypothesis space in this case seems impossible to define. In the case of harmony or melody, the hypotheses are discrete and occupy a finite space. But when it comes to high-level interpretations of an artwork – e.g. '*this sculpture is about death*', '*the artist was in love when he painted this*' –, the possibilities seem endless and the space they encompass infinitely dimensional.[8] Overcoming this difficulty will either require focusing on modelling only lower-level interpretations that are easier to represent mathematically or modelling visual art perception in a carefully controlled experimental task restricting the hypothesis space to a finite set without making the interpretation process too artificial. Another possibility is building hierarchical Bayesian models that learn the statistical regularities in sets of artworks and thus 'learn' the possible high-level interpretations. These models can also be enriched by incorporating features of the early human visual system, which would feed into the probabilistic computation the kinds of signals actually available to the visual system. [46] The predictions of such hierarchical models can be tested as follows: (1) group together artworks assigned similar high-level interpretations by the model, (2) have human subjects assign their own labels to each of these groups, and (3) see if individual human responses to single artworks correspond to the group-level labels assigned by the human subjects in step (2).

Directly probing specific human interpretive predictions may be virtually impossible, as many of the interpretive steps may occur at infinitesimal timescales. Here, it will be important to design clever laboratory tasks that manipulate the features of the presented artworks without altering the art too artificially and seeing how interpretations differ across conditions. Interviews will be useful for uncovering subjects' interpretations, although more precise methods are desirable. Some possibilities here are priming methods analogous to those used to probe musical expectations [28, 30], or measuring correlates such as emotional response, again using galvanic skin response. [39]

We have seen that the Bayesian brain hypothesis (and particularly the predictive coding instantiation of it) has the potential to explain visual art perception, although empirically testing the hypothesis in this domain may prove particularly difficult. I have additionally argued that historical, sociological, linguistic, and neurobiological evidence warrants a unified theory of art cognition that explains both visual art perception and music perception with the same principles. The Bayesian account I have presented provides such a theory. Furthermore, it explains individual differences in both domains and makes empirical predictions at the neurophysiological level (predictive coding).

5. Discussion

As I stressed in the introduction, the above remains in solely theoretical territory – more work is required to refine these ideas into specific testable predictions and to verify these predictions experimentally. By the nature of the theory, this will require a computational modelling approach. It is of course desirable to develop experiments showing that music perception is governed by musical expectations and that visual art perception is governed by interpretive decisions and attention. But ultimately what will be needed to verify the Bayesian brain hypothesis in this context is evidence that these expectations and interpretive decisions are being computed in a Bayes–optimal manner. This inherently requires constructing mathematical models that define what it means to be Bayes–optimal, and testing human behavior to verify that it agrees with the models.

It is worth noting that the theory I have proposed here (particularly in its predictive coding instantiation) satisfies each of the three criteria Ramachandran & Hirstein (1999) state are necessary for a good theory of art:

> "Any theory of art (or, indeed, any aspect of human nature) has to ideally have three components. (a) The logic of art: whether there are universal rules or principles; (b) The evolutionary rationale: why did these rules evolve and why do they have the form that they do; (c) What is the brain circuitry involved?"

The Bayesian brain hypothesis proposes answers to each of these questions. Importantly, it does so in an explicit way by formally specifying the computations underlying art cognition. Such an understanding of art cognition is desirable, as the language of computation can bridge explanations across the Marrian levels [2] and provide a common terminology to relate art-related cognitive processing to the rest of cognition. Moreover, like language, art cognition is exclusively human, putting it in a unique position to inform our understanding of general computational properties of the human brain that make us different from other animals.

I now go on to summarize the answers the Bayesian theory of art I have presented gives to each of Ramachandran & Hirstein's (1999) three questions:

(a) The logic of art. The underlying universal principle of

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

Bayesian art cognition is the optimal integration of perceptual information with prior knowledge. Such prior knowledge may consist of statistical regularities of a given style (e.g. I follows V^7), explicit rules (e.g. voice-leading rules), or meta-information about the given artwork being experienced (e.g. Picasso was depressed when this work was made). Bayes' rule provides a way of integrating this knowledge with the incoming perceptual information from the artwork being viewed or heard to produce a meaningful prediction of what the work is about, or, in a temporal context (as in music), of what will come next. Bayesian inference can thus account for the fact that different people can experience the same artwork differently – a characteristic feature of art. Furthermore, several perceptual phenomena have been shown to arise from Bayesian computation [3, 6, 7], so the Bayesian framework provides a means to explain art cognition using the same principles underlying sensory perception. Given the intimate relationship between these two processing domains, this is desirable. Lastly, I have shown that these principles may underlie both music and visual art perception. As reviewed in §4, a theory of art cognition that predicts a computational overlap in music and visual art perception is warranted by several observations: parallel historical trends (from predictable to unpredictable), linguistic overlap (we use similar language to describe both), and shared neurobiological substrates. [37]

(b) The evolutionary rationale. The evolutionary origin of art is one of the most difficult and mysterious questions in art cognition, since art does not seem to directly fulfill any survival needs. The Bayesian theory of art cognition, however, inherits all the evolutionary arguments for the Bayesian brain hypothesis, which is firmly grounded in evolutionary theory. [47] Provided the multitude of ways in which predictions about the world can aid in survival, we would expect the human brain to have evolved to compute these accurately. Bayesian inference is the mathematically optimal way of doing so, so it is easy to see how natural selection would give rise to brains that perform it.[9] Should art cognition be a product of Bayesian inference, the argument for why it evolved is straightforward: the human brain evolved to perform Bayesian inference and Bayesian inference leads to art cognition. In other words, art cognition is a byproduct of the natural selection for optimal probabilistic inference.

(c) The neurobiological substrates. The Bayesian brain hypothesis makes several empirical predictions about the underlying neurobiology of art cognition. One of the most promising proposals in this respect is that of predictive coding. [9, 16] Aside from this, some more modest claims that fall out of the Bayesian theory of art cognition are that the neural circuits involved must encode probabilities or probability distributions in some way, and must have a means of manipulating them [3] (see also [10] for an interesting proposal of how this could be done under certain conditions).

The explanatory power of the Bayesian brain hypothesis is demonstrated by its ability to provide answers to each of these questions (however, the fact that it can account for so many broad phenomena may also be its greatest weakness [15]). It is thus essential that we empirically test whether Bayesian computation indeed plays such a crucial role in art cognition. If it does, art cognition would hold as another piece of evidence supporting the strongest claim of Bayesian brain hypothesis, namely that the whole brain is built to compute probabilities in a Bayes–optimal fashion. It could also provide support for other related large-scale theories of neural computation (e.g. the Free Energy Principle, [9]). It is important to note here that it could be the case that art cognition is the product of Bayesian computation but not along the lines of the inferences I have proposed here. In this vein, it is important that more theoretical work continue to develop the ideas I have outlined and propose other possible probabilistic inferences that may play a role in art cognition (see, for example, [26]).

It should be noted as well that, even if art cognition is indeed fraught with Bayesian inference, Bayesian computation may only account for one component of art cognition. The arguments I have put forth suggest that this component would be the the most crucial to music and visual art perception, but there may be others that also contribute to experiencing art. For example, associations in long-term memory by themselves can give significance to a work of art (e.g. if you grew up on a farm, Monet's *Haystacks* paintings may bring back memories from your youth).

A related problem that must be taken into account when considering the present proposal is the scientific bias inherent in its Western origins. Visual art and music can take drastically different forms from culture to culture [33], but the ideas presented here, while presented objectively, are grounded in a Western musical education and a Western conception of art. Other conceptions of art may not fit in with the proposals I have made. However, even if it turns out that the theory only holds for Western music, it will may still have interesting implications. Such a result would mean that the appreciation of Western music is an entirely distinct cognitive process from the appreciation of other types of music. In that case, the theory should make substantial predictions about the artistic differences between Western and other cultures. Such factors are important to consider and illustrate the importance of anthropological work to the study of art cognition (e.g. ethnomusicology).

Before concluding, a significant philosophical implication of the Bayesian theory of art should be mentioned. The idea that art is first and foremost a phenomenon arising from integrating perceptual information with prior knowledge raises the issue of whether music and visual art perception are instances of cognitive penetration. Cognitive penetrability is the idea that perception could be affected by higher cognition (e.g. explicitly known facts, beliefs). At first glance it seems that my account of art cognition is indeed an illustration of this. Empirically, this may be problematic, as several recent studies have refuted that cognitive penetration happens, even in canonical demonstrations of it [48, 49]. However, the present proposal need not be interpreted in a way that necessarily implies cognitive penetration. Even though the Bayesian account intimately intertwines music and visual art perception with sensory perception, the experience of art may still arise without altering the sensory percept whatsoever. Expecting a given chord to come next, for example, won't necessarily change what the next chord sounds like, it might just change how you feel when you hear the next chord. Seeing sad faces in a painting might make you expect and look for the color blue, but it won't necessarily make the colors in the painting look 'bluer'. Altogether, the Bayesian theory remains agnostic as to whether cognitive penetration occurs when experiencing art. It is important to note, however, that certain instances of it – predictive coding [16] and the Free Energy Principle [9], for example – certainly do

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

imply cognitive penetration.

6. Conclusions

I propose that art cognition fundamentally consists of integrating perceptual information with prior knowledge, and that this computation is governed by Bayesian inference. This proposal makes testable predictions at behavioral and neurobiological levels, which should be empirically verified. The Bayesian theory of art cognition holds great promise because it can parsimoniously explain perception, visual art, and music with the same computational principles. Furthermore, it offers an evolutionary rationale for the uniquely human ability to create and enjoy art.

Importantly, the current treatment demonstrates the powerful computational mechanisms that may underlie art cognition. As a result, art cognition has the potential to shed light on some of the most complex computational processes implemented by the human brain. This fact highlights the paramount importance of investigating the computational substrates of art cognition. Art cognition can serve as a powerful probe into the inner workings of human cognition, but only with a computational understanding of it.

Notes

1. The terms *music perception* and *visual art perception* are thus defined as individual instances of *art perception*.
2. The normalizing constant $p(D)$ is constant across different H's (hypotheses) being evaluated, so can usually be ignored.
3. An alternative but weaker claim is that the outputs of the brain's computations, however they are computed, are in agreement with the output of a Bayesian algorithm.
4. In Western music theory, the terms tonic and dominant seven designate the two most important harmonies in tonal music. The tonic chord defines the *key* of a piece of music: the pitch space encompassed by its principal melodic and harmonic components (deviations from the key are especially salient, e.g. "bluesy" notes in jazz). This defines the piece's tonal center – the pitch space a piece typically begins and ends in. The dominant seven chord contains the pitches adjacent to those in the tonic chord, thus creating a tension that is relaxed when a dominant seven chord is succeeded by a tonic chord (the dominant seven chord is 'drawn' to the tonic chord). This succession is termed a "full authentic cadence" and is one of the cornerstones of Western tonal music (e.g. the transition from "to" to "you" in the last phrase of "Happy birthday").
5. Such elaboration effects have been shown to affect the judgment and experience of music as well. [43]
6. This strategy, called the MAP estimate, is often computationally intractable. Other possibilities exist, such as the Maximum Likelihood (ML) estimate, which takes the value at which the likelihood distribution peaks (as opposed to the value at which the posterior distribution peaks). This possibility in fact suits the proposed mechanism well, as the likelihood function would be the only required function for the whole computation. Furthermore, the ML estimate is in fact equivalent to the MAP estimate when the prior distribution is uniform, which would be a reasonable assumption in the present context (i.e. all interpretations equally likely *a priori*).
7. Of course, the selection and evaluation of these hypotheses may happen at different timescales, with high-level interpretations on the order of seconds or minutes and low-level interpretations on the order of milliseconds.

Figure 1. *Ecstasy of Santa Teresa*, Gian Lorenzo Bernini (1947-52)

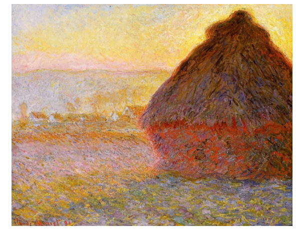

Figure 2. *Haystacks (sunset)*, Claude Monet (1890-1)

Figure 3. *Les Demoiselles d'Avignon*, Pablo Picasso (1907)

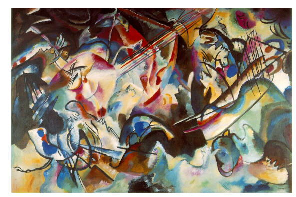

Figure 4. *Composition VI*, Wassily Kandinsky (1913)

8. Of course, such interpretations exist for music as well. However, because music unfolds temporally, an interpretation for the whole piece cannot be arrived at until it is over. In this sense, one can think of Bayesian inference as playing two roles in music perception: the role of predicting the music as it unfolds and the role of interpreting the meaning of the piece after you have heard it. These two processes may inform each other, while neither necessitates the other.
9. A significant caveat here is that the path dependence of natural selection would likely lead to effective *non-optimal* solutions to survival. [15]

References

[1] E. Dissanayake, The artification hypothesis and its relevance to cognitive science, evolutionary aesthetics, and neuroaesthetics, Cogn. Semiotics, 5 (2009).
[2] M. Carandini, From circuits to behavior: a bridge too far?, Nat. Neurosci., 15, 4 (2012).
[3] D.C. Knill & A. Pouget, The Bayesian brain: the role of uncertainty in neural coding and computation, Trends Neurosci., 27, 12 (2004).
[4] L.J. Savage, Foundations of statistics, Dover, 1972.
[5] O. Schulte, The logic of reliable and efficient inquiry, J. Philos. Logic, 28 (1999).

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

Figure 5. *No. 61 (Rust and Blue)*, Mark Rothko (1953)

Figure 6. *Fountain*, Marcel Duchamp (1917)

Figure 7. Figure adapted from Ramachandran & Hirstein (1999)

Figure 8. Adam Zyglis, 2010

[6] T.L. Griffiths, N. Chater, C. Kemp, A. Perfors & J. Tenenbaum, Probabilistic models of cognition: exploring representations and inductive biases, Trends Cogn. Sci., 14, 8 (2010).

[7] A. Pouget, J.M. Beck, W.J. Ma & P.E. Latham, Probabilistic brains: knowns and unknowns, Nat. Neurosci., 16, 9 (2013).

[8] R.R. Llinás, I of the vortex: from neurons to self, MIT Press, 2002.

[9] K. Friston, The free–energy principle a rough guide to the brain?, Trends Cogn. Sci., 13, 7 (2009).

[10] W.J. Mae, J.M. Beck. P.E. Latham & A. Pouget, Bayesian inference with probabilistic population codes, Nat. Neurosci., 9, 11 (2006).

[11] L.B. Meyer, Meaning in music and information theory, J. Aesthet. Art Critic., 15, 4 (1957).

[12] M. Mann, Some indian conceptions of music, P. Roy. Music Assoc., 38, 5 (1911).

[13] J.R. Saffran, R.N. Aslin & E.L. Newport, Statistical learning by 8–month–old infants, Science, 274, 5294 (1996).

[14] J.R. Saffran, E.K. Johnson, R.N. Aslin & E.L. Newport, Statistical learning of tone sequences by human infants and adults, Cognition, 70, 1 (1999).

[15] J.S. Bowers & C.J. Davis, Bayesian just–so stories in psychology and neuroscience, Psychol. Bull., 138, 3 (2012).

[16] A. Clark, Whatever next? Predictive brains, situated agents, and the future of cognitive science, Behav. Brain Sci., 36, 3 (2013).

[17] R.P. Rao, & D.H. Ballard, Predictive coding in the visual cortex: a functional interpretation of some extra–classical receptive–field effects, Nat. Neurosci., 2, 1 (1999).

[18] J. Hohwy, A. Roepstorff, & K. Friston, Predictive coding explains binocular rivalry: An epistemological review, Cognition, 108, 3 (2008).

[19] S.M. Sherman & R.W. Guillery, On the actions that one nerve cell can have on another: distinguishing “drivers”from “modulators”, Proc. Natl. Acad. Sci., 95, 12 (1998).

[20] D. Mumford, On the computational architecture of the neocortex. II. The role of cortico-cortical loops, Biol. Cybern., 66, 3 (1992).

[21] R.S. Wolpert, Recognition of melody, harmonic accompaniment, and instrumentation: musicians vs. nonmusicians, Music Percept., 8, 1 (1990).

[22] R.S. Wolper, Attention to key in a nondirected music listening task: musicians vs. nonmusicians, Music Percept., 18, 2 (2000).

[23] P.O. Peretti, Changes in galvanic skin response as affected by musical selection, sex, and academic discipline, J. Psychol., 89, 2 (1975).

[24] S.D. Vanderark & D. Ely, Cortisol, biochemical, and galvanic skin responses to music stimuli of different preference values by college students in biology and music, Percept. Motor Skill, 77, 1 (1993).

[25] E. Bigand & B. Poulin–Charronnat, Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training, Cognition, 100, 1 (2006).

[26] D. Temperley, Bayesian models of musical structure and cognition, Music. Sci., 8, 2 (2004).

[27] D. Ponsford, G. Wiggins & C. Mellish, Statistical learning of harmonic movement, J. New Music Res., 28, 2 (1999).

[28] J.J. Bharucha & K. Stoeckig, Reaction time and musical expectancy: priming of chords, J. Exp. Psychol. Human, 12, 4 (1986).

[29] E. Bigand & M. Pineau, Global context effects on musical expectancy. Percept. Psychophys., 59, 7 (1997).

[30] M.A. Schmuckler & M.G. Boltz, Harmonic and rhythmic influences on musical expectancy, Atten. Percept. Psycho., 56, 3 (1994).

[31] F. Lerdahl & C.L. Krumhansl, Modeling tonal tension, Music Percept., 24, 4 (2007).

[32] R.L. Gómez & L. Gerken, Infant artificial language learning and language acquisition, Trends Cogn. Sci., 4, 5 (2000).

[33] S.E. Trehub, J. Becker & I. Morley, Cross–cultural perspectives on music and musicality, Philos. T. Roy. Soc. B., 370, 1664 (2015).

[34] S. Brown & J. Jordania, Universals in the world’s musics. Psychol. Music, 41, 229 (2013).

[35] R. Taruskin & C.H. Gibbs, The Oxford History of Western Music: College Edition, New York: Oxford University Press, 2012.

[36] R. Tansev, F. Kleiner & H. de la Croix, Gardner?s Art Through the Ages, Harcourt College Pub, 1995.

[37] T. Ishizu & S. Zeki, Toward a brain-based theory of beauty, PLoS ONE, 6, 7 (2011).

[38] H. Leder, B. Belke, A. Oeberst & D. Augustin, A model of aesthetic appreciation and aesthetic judgments, Brit. J. Psychol., 95, 4 (2004).

[39] V.S. Ramachandran & W. Hirstein, The science of art: a neurological theory of aesthetic experience, J. Consciousness Stud., 6, 6-7 (1999).

[40] K. Millis, Making meaning brings pleasure: the influence of titles on aesthetic experiences, Emotion, 1, 3 (2001).

[41] P.A. Russell, Effort after meaning and the hedonic value of paintings. Brit. J. Psychol., 94, 1 (2003).

[42] H. Leder, Determinants of preference. When do we like what we know? Empirical Studies of the Arts, 19, 2 (2001).

[43] A.T. Landau, S.K. Rankin, P. Jiradejvong & C.J. Limb. (2015, August) The Prestige Effect: increasing musical appreciation through extramusical factors. Poster presented at the biannual meeting of the Society for Music Perception and Cognition. Nashville, TN.

[44] A.S. Winston & G.C. Cupchik, The evaluation of high art and pop-

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-142

ular art by naive and experienced viewers, Visual Arts Res., 18, 1, (1992).

[45] T. Watanabe, J.E. Náñez & Y. Sasaki, Perceptual learning without perception, Nature, 413, 6858 (2001).

[46] M. Carandini, J.B. Demb, V. Mante, D.J. Tolhurst, Y. Dan, B.A. Olshausen, J.L. Gallant & N.C. Rust, Do we know what the early visual system does?, J. Neurosci, 25, 46 (2005)

[47] T.L. Griffiths, N. Chater, D. Norris & A. Pouget, How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012), Psychol. Bull., 138, 3 (2012).

[48] C. Firestone & B.J. Scholl, "Top–down"effects where none should be found: The El Greco Fallacy in perception research, Psychol. Sci., 25, 1 (2014).

[49] S. Gross, T. Chaisilprungraung, E. Kaplan, J.A. Menéndez & J.I. Flombaum, Problems for the purported cognitive penetration of perceptual color experience and Macpherson?s proposed mechanism. Baltic International Yearbook of Cognition, Logic and Communication, 9, 1 (2014).

Acknowledgments

Special thanks to Dr. Mónica López-González for introducing me to the field of music cognition and allowing me to develop my ideas. Thanks also to my many professors and mentors at Johns Hopkins University, in particular Professors Jonathan Flombaum and Steven Gross, whose ideas and classes inspired many of the ideas put on paper here.

Author Biography

Jorge A. Menendez has a BA in Cognitive Science from the Johns Hopkins University (2015) and a BM in Classical Guitar Performance from the Peabody Institute (2015). He is currently pursuing an MRes in Modelling Biological Complexity at University College London, where he has active research interests in issues regarding neural computation and perception.