Crossing the Theshold Paradox: Modelling Creative Cognition in the Global Workspace

Geraint A. Wiggins
Centre for Digital Music
School of Electronic Engineering and Computer Science
Queen Mary University of London
Mile End Road, London E1 4NS, UK
geraint.wiggins@eecs.qmul.ac.uk

Abstract

I present a hypothetical global model of everyday creative cognition located within Baars' Global Workspace Theory, based on theories of predictive cognition and specific work on statistical modelling of music perception. The key idea is a proposal for regulating access to the Global Workspace, overcoming what Baars calls the Threshold Paradox. This idea is motivated as a general mechanism for managing the world, and an argument is given as to its evolutionary value. I then show how this general mechanism produces effects which are indistinguishable from spontaneous creative inspiration, best illustrated by Wallas' (1926) "Aha!" moment. I argue that W. A. Mozart's introspective account of compositional experience closely matches the proposed process, and refer to a computational system which will form the basis of an implementation of the ideas, for musical composition.

Introduction

Computational Creativity is mired, practically speaking, in the problem of evaluation. Artefacts created by computer cannot be judged by the computer's aesthetic, for that is obscure, and evaluating them in terms human aesthetics has been shown to be unreliable due to negative preconceptions (Moffat and Kelly, 2006). One solution to this might be to compensate for that bias statistically, given the necessary models. Another is to avoid the issue of artefact evaluation altogether, and focus on process, and on building systems that apply it. Colton (2009) catchily entitles this point "Paradigms Lost", making the point that AI sometimes overtheorises, and paints itself into a corner by the application of problem solving methods to a domain in the abstract, instead of getting on and building something that concretely explores it: the subtext may be that this tendency arises from rigour envy. Colton raises a point that benefits from emphasis: he finishes the section with "the production of beautiful, interesting and valuable artefacts", and this occludes the key point in the final sentence: "the need to embrace *entire* intelligent tasks" (my italics).

The vexed question is "how?" Modelling an *entire, novel* creative process, evaluation, reflection, and all, in the abstract leads us back to the initial problem: the only way to judge it from outside is in terms of its outputs (Ritchie, 2001).

The rest of the paper is structured as follows. First, I explain the theoretical methods used, and make and important distinction between what I shall call *inspiration* and *creative reasoning*; the current proposal addresses only the first of these. Next, I describe some of the extended background to the thinking presented here, and in terms of the surrounding and supporting cognitive theory, including an apparent inherent paradox identified by its creator. Then I present the evolutionary argument for the theoretical stance taken here, and derive the (simple) principles on which my proposal is based from it. Next, turning to implementation, I summarise earlier modelling work, explain its connection with the current proposal, and describe what is necessary to extend it into the model proposed here.

The technical contributions of the paper are a variant notion of AI Agent, based on prediction from sense data, rather than on sensing, and a mechanism for deployment of that agent in a particular kind of reasoning system. The key philosophical contribution is the fact that, once this mechanism is deployed, the kind of creativity that is addressed here, inspiration, is explained *within* the basic reasoning, and needs no further explanation.

Methodology

To overcome the methodological problem introduced above, my approach is to attempt to replicate an existing creative process. The only existing creative process ready available for inspection is that of humans; these have the built-in advantage, mostly, of being able to explain what (they thought) they did, and elegant paradigms exist to empirically deconstruct that majority of aspects of human behaviour of which introspective reports are unreliable. I therefore aim to apply cognitive modelling theory and technology to human creative process, and then to evaluate the success of the enterprise with respect not only to the outputs of the computational systems produced, but to compare the various aspects of their operation with human creators. While this approach solves only part of the general problem of computational creativity, it is an area where refutable hypotheses can be made, and so demonstrable progress in a research programme (Lakatos, 1970) may take place.

For this attempt to succeed in a scientific sense, before one even considers the artefacts that the replicant creative system may produce, the theory and its associated computational

system must conform to at least the following constraints, to be said to model *human creative cognition*.

1. **Falsifiability** The system must not behave in ways which are arguably or demonstrably different from human creators while it is operating. Since we cannot, currently, know how human creators create, this is the strongest falsifiability constraint that can be applied.
2. **Evolutionary context** There must be an account of the evolutionary advantage conferred by the mechanisms proposed, a corresponding order of development, and an analysis of their appearance in successive species over evolutionary time. This account cannot be verifiable, but the lack of one leaves the biological development of the proposed solution unavailable to scientific scrutiny.
3. **Learning capability** The system must be capable of learning its creative domain. Learning should be appropriate to the domain: for example, in music, perceptual aspects should be *implicit*—that is, teaching or supervision should not be required; however, in some domains, such as mathematics, minimal supervision is evidently unavoidable, because of the need to know the meaning of symbols, to give semantics to what is being learned[1].
4. **Production capability** The system must be able to produce artefacts that are demonstrably within its creative domain, whether or not they are of quality comparable with a human creator's output. While the judgement of whether an artefact is or is not a particular kind of thing is subjective, it is not as difficult as the subjectivity of quality. For the purposes of experiment, restricted domains with clear tests must be set up, using appropriate theory from the corresponding human-creative domain.
5. **Reflection** The system must be capable of reflecting on its behaviour, modifying it, and explaining it—where necessary via indirect indicators such as those used for understanding the behaviour of humans.

In this paper, I present a hypothetical, but partly implemented, computational model of a particular kind of human creativity, and suggest that it conforms to criteria 3–4, and partly to criterion 2, though further research is required to provide more evidence against criterion 1. Criterion 5, Reflection, is conferred by location of the model within Baars' (1988) Global Workspace Theory, whose focus is consciousness; so it falls beyond the scope of the present proposal.

Background

Creativity: Inspiration and Reasoning

Wiggins (2012) introduces a distinction between two kinds of creativity: on one hand, *inspiration* and, on the other, *creative reasoning*. Respectively, these terms are intended to distinguish what appears spontaneously in consciousness—the "Aha!" moment that Wallas (1926) suggests follows "incubation"—from what is produced by the deliberate application of creative method. The spectrum between the two allows us to make distinctions between conscious creation in the deliberate planning of a formalist composer, the semi-spontaneous but cooperative and partly planned creation of the jazz improviser in a trio, and entirely spontaneous singing in the shower. Note that a non-polar position on this spectrum necessarily entails a *combination* of explicit technique and implicit imagination: there is not a smooth transition in kind between the two, but rather a *mixture* containing some of each in varying proportion.

Having made this point, I reserve creative reasoning for future work, not least because it entails that we address consciousness, which is difficult, but also because Baars' theory already provides a framework in which it may be considered, *given* a mechanism for inspiration. This is not to dismiss the deliberate end of the scale, nor to suggest that it does not exist, but merely to focus the current work on a separable aspect of the complex.

Global Workspace Theory

Bernard Baars (1988) introduces a theory of conscious cognition called the Global Workspace Theory. There is not space to describe this wide-ranging and elegant theory here, so I summarise the relevant important points. The theory posits a framework within which consciousness can take place, based around a multi-agent architecture (Minsky, 1985) communicating via something like an AI blackboard system (Corkill, 1991), but with particular constraints, which I outline below. The approach taken is to avoid Chalmers' "hard" question of "what is conscious?" (Chalmers, 1996) and instead ask "what is it conscious of, and how?" This is especially appropriate in cases such as the current paper, where consciousness is not the central issue, but presentation of information to it is.

Baars casts the non-conscious mind as a large collection of expert generators (not unlike the multiple experts in Minsky's *Society of Mind*, 1985), performing tasks by applying algorithms to data in massive parallel, *compete for access* to a Global Workspace via which (and only via which) information may be exchanged; crucially, information must cross a notional threshold of "importance" before it is allowed access. The Global Workspace is always visible to all generators, and contains the information of which the organism is conscious at any given time. However, it is capable of containing only one "thing" at a time, though the scope of what that "thing" might be is variable. The Global Workspace is highly contextualised, and meaning contained therein is context sensitive and structured; contexts can contain goals, desires, etc., of the kind familiar from broader AI. Aside from further discussion of the "threshold" idea, below, this is all that is needed to understand the purpose of the competition mechanism proposed here. Baars mentions the possibility of creativity within this framework in passing, implicitly equating entry of a generator's output into consciousness with the "Aha!" moment (Wallas, 1926). However, he does not develop this idea further beyond noting that a process of refinement may be implemented as cycling of information into the Workspace and out again; that process

[1] To ask the system to learn the semantics of the symbols to which it is exposed from context is not, in principle, unreasonable, as there is every evidence that humans do so. However, to require the system to do so when the scientific research focus is creativity seems unnecessarily difficult.

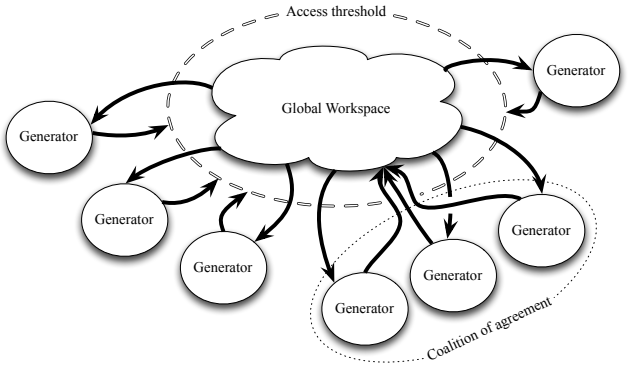

Figure 1: Illustration of Baars' Threshold Paradox. Generators generate, but need a means of recruiting support for their outputs. Individuals cannot break in; they must recruit coalitions, as shown. The only way to do so is via the Global Workspace, but before they can do so, they need the support they are trying to recruit, and therein lies the paradox.

may be equivalent to my creative reasoning. To the best of my knowledge, however, creativity in the Global Workspace has not been addressed elsewhere in the related literature.

In the later developments of the theory, Baars proposes that information integration may take place in stages, via something that one might (but he does not) call local workspaces, that integrate information step by step in a sequence, rather than all in one go as it arrives in the Global Workspace. This information integration approach has been extended by Tononi and Edelman (1998), who propose information-theoretic measures of information integration as a measure of consciousness of an information-processing mechanism. Baars has embraced the information-theoretic stance, too, and the three authors have jointly proposed to begin implementing a conscious machine (Edelman, Gally, and Baars, 2011) based on their ideas. The current work may contribute to this endeavour, though probably at a level more abstract from neurophysiology than these authors intend.

The Threshold Paradox

Baars (1988, pp. 98–99) addresses what he acknowledges is a problem for his theory. He proposes a threshold for input access to the Global Workspace, crossing of which is thought of in terms of recruiting sufficient generators to produce information that is somehow coordinated, or synchronised between them: it must be metaphorically "loud" enough to be "audible" in the Workspace. However, in terms of the Global Workspace alone, there is no means of doing this: generators can only be coordinated (whatever that means) via the Global Workspace, and so the generators are faced with the beginning artist's dilemma: you have to be famous to show your work, but you have to show your work to become famous. This form of the Workspace is illustrated in Figure 1. Baars presents two possible solutions to the paradox, which is the motivation of the current paper, but both are presented somewhat half-heartedly, leaving a gap in the theory. Here, I present a possible solution, in terms of the evolutionary argument required by my criterion 2, above.

Perception, Anticipation, and Evolution

Reaction vs. Anticipation

I now present a mechanism for managing the competition between generators in Baars' system. This mechanism may be implemented either directly or indirectly (that is, by means of some other effect)—the difference is immaterial at the current theoretical level. The key distinctions are a) between the information content and entropy (defined below) of various stimuli; and b) between organisms that react and organisms that anticipate. The design of this mechanism is motivated by evolutionary thinking: that is, by consideration of the evolutionary advantage conferred by the resulting behaviours, in humans and other animals. Thus, the evolutionary argument presented here is part of the design, not merely an example.

Russell and Norvig (1995), in their well-known AI text book, define an AI agent (of which an AI *creative* agent is presumably an instance) as a program or robot with a behaviour cycle that consists of perceiving the world and then acting on the perceptions. It seems not unreasonable to present this as a model of lower organisms, such as insects, which seem to do nothing more than react to environmental conditions, coping poorly when their evolved reactive program is interrupted. However, to model higher cognitive development, one can propose a more predictive system, in which an organism is predicting continually, from a learned model of previous sensory data, what is likely to come next, and *comparing* this with current sensory input. Doing so gives a simple but effective mechanism for spotting what is unusual, what, therefore, constitutes a potential new opportunity or threat, and what deserves cognitive resource, or *attention*. In the simplest case, the anticipatory agent can in principle avoid a threat *before* it becomes apparent, while the reactive one has to experience the threat in order to respond.

The consequence of sequence: managing uncertainty with expectation

The most important feature of an autonomous agent is not, as sometimes supposed in AI, that it is able to identify or categorise a situation from available data. What gives it the edge is that it can, in some sense, *imagine* what is to come next, and react, or perhaps *pre*act, in advance. Of course, the word "imagine" is loaded, and suggests the involvement of consciousness and even volition; I use it here deliberately to draw attention to the point that consciousness need not be implicated in this process, which can be described in completely mechanistic terms, of *prediction* alone.

In order to predict usefully in a changing world, it is necessary for an organism to learn. It must be able to learn not just categorisations (to understand what something is), but also associations (to associate co-occurrence of events with reward or threat), and, crucially here, sequence.

However, a simple statistical learning mechanism is not subtle enough (Huron, 2006). Since evolutionary success entails that an organism breeds, a mechanism which allows

that organism to learn only from potentially fatal consequences does not suffice: if the organism dies as the result of an experience, it does not benefit from the experience (or, at least, not for long). An effective strategy here lies at a meta-level with respect to a learned body of experience: if an organism is aware that it is in circumstances that it cannot predict reliably, it can behave more cautiously, its metabolism can be aroused to prepare for flight, and it can devote more attention than normal to its surroundings; thus, the effective strategy is also *a*ffective. Huron convincingly argues that this process is exapted to produce part of the aesthetic effect of music; however, for the purposes of the current section, the mere *ad*aptation suffices: self-evidently, there is a mechanism that allows uncertainty to affect behaviour in humans and other animals, and that mechanism does not rely on explicit reasoning: indeed, the converse is the case: we feel nervous in uncertain situations, and the feeling serves to make us wonder why, as well as to heighten our attention to appropriate sensory inputs and to prepare for flight. This mechanism, and the associated affective response, is not the same as fear, but can lead there *in extremis*.

Finally, any kind of learning of this nature is inadequate unless it includes *generalisation*. It is necessary to be able to generalise from both co-occurrence and sequence that similar consequences arise from similar events, encounters, etc. Without this, mere tension cannot lead to the fear that is appropriate at the sight of the bared fangs of a previously-unexperienced large animal. This accords with proposals such as that of Gärdenfors (2000), that perceptual learning systems are motivated by the need to understand similarities and differences between perceived entities in the world, and to place observations at the appropriate point between previously experienced referents.

Prediction, Prioritisation and Selection

Given a model of the world, suitably subcategorised into types, situations, etc., one can imagine a set of generators using the model with recent and current perceptual inputs matched against precursors of sequential associations, making predictions, on a basis that is stochastic, and conditioned by the model. Making such predictions quickly, one at a time, would be valuable, but, given the nature of brains, slow, multiple predictions, in parallel, are a more likely candidate for evolutionary success, and the more the better—as in Baars' proposal. But this begs a question: arbitrarily many predictions occurring simultaneously will be an impossible, incomprehensible babble, so how will useful candidates for prediction be selected? Baars' solution is the problematic threshold, described above.

Another shortcoming of the Global Workspace Theory is unclarity about precisely what the notion of generators "recruiting" one another means. The *effect* is something like an additive weight: the more generators that are "recruited", the greater the impact of their output. In my proposal, we will avoid answering this question, by approximating the effect of the recruitment, rather more simply. I return to this below; in the argument that follows, I will use the analogy of sound volume to refer to this property: "loud" predictions come from many generators, "quiet" ones do not.

My proposal here is based on statistical, frequentist notions of learning, and so my reasoning is couched in terms of statistical models; however, I do not think that the reasoning is in principle exclusive to such models, and it should not be supposed that the proposal is *restricted* in this way. In this view of the system surrounding the Global Workspace, there are many independent subsystems, which are making multiple predictions by biased sampling from a predictive statistical model of (assumedly) reasonable quality. It is also appropriate to assume imperfect models: each of these generating subsystems will have a fragmentary, partial view of its world and its predictions, as to model everything all the time in massive-parallel would be prohibitively expensive. It follows from the use of frequentist models that the more expected occurrences are the more likely ones to be predicted: the commonest predictions will be the most expected ones. This means there are relatively "loud" groups of contributions, reinforcing each other. Conversely, extremely unlikely predictions will be proposed by only a very small number of generators, and as such will never be "audible".

In a model of prediction and action based solely on this frequentist principle, an organism will tend do the commonest thing, even when inappropriate, and therefore will be doomed to failure: it will not "imagine" unlikely and surprising situations, and will not therefore prepare itself against necessary eventualities. To see this, consider a territorial animal, on patrol, and let it be a high enough species to learn its reactions. Today, our animal senses the things it usually senses, and the vast majority of things in the world today are the same as they were the last time it passed this way. One tiny difference is a scent that it does not recognise, that it has not experienced before. Since this difference is small in comparison to the rest of the data in the world, and it has not been experienced before, in purely frequentist terms, it will be ignored: it is unlikely, and it has no known consequences and determines little or no probability mass.

In Baar's theory, the pure frequentist approach, where the most likely outcome is chosen, corresponds with multiple generators in coalition generating that outcome. The likelihood of each generator predicting an outcome is proportional to the "volume" of that outcome across the set of generators. Therefore, we can neatly draw a veil over the mechanistic gap left by Baars' idea of coalition formation, and simply use the likelihood of the outcome, p, to model its outcome.

In reality, though, we know well that to carry on as normal will not be the reaction of an animal in these circumstances: it will experience Huron's proposed affective response, described above. Therefore, it is necessary to hypothesise a mechanism to cause that response. In our current simple context of abstracted statistical modelling, the obvious choice for such a mechanism is the notion of *entropy*, as formalised by Shannon (1948). MacKay (2003) makes a distinction between *information content*, h, which is defined as an estimate of the number of bits required to describe an event, e, given a context, c, or its *unexpectedness*:

$$h(e \mid c) = -\log_2 p(e \mid c),$$

and *entropy*, H, which is defined as an estimate of the *uncertainty* inherent in the distribution of the set of events $\mathcal{E}$

from which that e might be selected, given the context, c:

$$H(c) = \sum_{e \in \mathcal{E}} p(e \mid c)\, h(e \mid c) = -\sum_{e \in \mathcal{E}} p(e \mid c)\, \log_2 p(e \mid c).$$

H is maximised when all outcomes are equally likely, and minimised when a single outcome is certain. Both h and H are useful to our hypothetical animal.

First, consider h_t, the unexpectedness of a partial model of the actual on-going experience in a particular state, t. If the experience is likely (in particular, if it is *readily predictable* from what has gone before), it is not unexpected, and therefore h_t is low; if it is unlikely, it is unexpected, and so h_t is high. An experience such as encountering a *new* scent is maximally unlikely, in frequentist terms. To model this, I propose that individual generators are sensitive to their own h_t value, and decrease their notional "volume" when it is low. Thus, the likelihood of models of the experience in which the new scent is included being heard in the theatre is positively related (possibly in a non-trivial way) to its unexpectedness. I call this the *recognition-h* case. It may explain why unexpected things are noticed.

Now, consider, h_{t+1}, the unexpectedness of a predicted situation. It is maximally unlikely that a *prediction* will be made including a scent that has not been encountered before, and, as above, we would therefore expect h_{t+1} to be very high, causing alarm. Excess of such predictions, or repeated occurrence of a single one, would lead to a state of constant anxiety[2]. I call this the *prediction-h* case. It may explain why surprising predictions are more likely to draw attention than prosaic ones.

Of course, in a simplistic frequentist account, predictions introducing new percepts or concepts cannot arise, because they entail the creation of new symbols. This is why it is necessary to include generalisation and/or interpolation in the theory (see above). Gärdenfors (2000) presents a theory that explicates the symbolic representations more commonly used in statistical AI modelling in terms of an underlying, sometimes continuous, geometrical layer, and, at least at perceptual levels, places cognitive semantics at the centre of mind. In particular, an outline mechanism is supplied whereby previously unencountered stimuli may be assigned first non-symbolic, and then symbolic, representations. It is important to understand that the semantics in these theories are internal to the organism experiencing them, and have no *definition* in terms of the external word; rather they have external associations, which can serve to allow intersubjective meaning, but they themselves are ineffable.

The problem of over-active prediction-h is mitigated by the mechanism supplied above, in which prediction is probabilistic and (broadly) additive across predictors, modelled by p. There are two opposing forces here, one of which changes inversely relative to the other, and because they are co-occurrent, their effects should (broadly) multiply. Therefore, the overall outcome audible in the global workspace can be estimated by multiplying the probability, p of an event (which estimates the likely number of generators predicting it) by h (which estimates the volume at which they are predicting). The resulting likelihood is illustrated by the unit-free diagram in Figure 2. This creates a bias away from predictions which are either very likely or very unexpected, reducing the power of the very unlikely or the very obvious to attract attention. This may explain why unlikely possibilities do not prevent action by overwhelming the acting organism with choice.

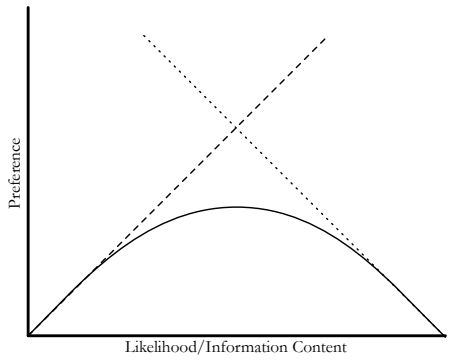

Figure 2: Illustration of the interaction between likelihood and unexpectedness. The overall likelihood (solid) is formed by the multiplication of two monotonic functions: the unexpectedness of a generated item (dashed) and the number of generators likely to agree on it, according to its likelihood (dotted).

It is important to see the difference between recognition-h and prediction-h in the context of the Global Workspace. I propose that generators may generate structures of either kind, and that the two will be in competition for the resource of attention. Thus, clear and present danger or benefit will outweigh predicted likelihoods, because the distribution of *potential* predictions is over a much wider range of possibilities than that over *actual* perceptions, and therefore, comparatively, probability mass is spread more thinly. Conversely, for example, likely but unexpected predicted benefits can outweigh less seriously dangerous present circumstances—thus, prioritising an unusual positive opportunity can be mechanistically explained as an emergent behaviour.

Given that there are now two kinds of generator (or generator output), I must propose a means of distinguishing between them, though this is not a key focus of the current argument. Without such a means, consciousness would be unable to distinguish between the perceived world and the predicted one[3].

Sensing Certainty

Shannon's H is interesting here in a different way. As explained above, H is the expected value of the information

[2]Indeed, some humans who suffer from anxiety, in the clinical sense, report intrusive, repetitive thoughts predicting problems or worries of one sort or another, the anxiety being aroused by fear of what *might* happen. Their situation would be explicable in terms of a breakdown of this mechanism.

[3]Coupled with a deficit in suppression of less likely outcomes, as above, this situation might lead to some root symptoms of schizophrenia: hallucinations, delusions and cognitive disorganisation.

content of a given distribution, so it is different in kind from h, which deals with individual situations, actual or predicted. It is best characterised as the uncertainty inherent in a distribution, and, indeed, a uniform distribution always gives the maximum entropy for a given alphabet size. Unlike h, H really only has meaning in the predictive context: once one knows which possibility of a range is the right one, only information content is really relevant. However, in the predictive context, a predicted outcome of which one is certain is much more useful than one of which one is unconfident: H measures this difference.

I propose, therefore, that, in the predictive generators, higher H also predicts lower volume, so that less certain generated outputs are de-emphasised. This, then, I call *prediction-H*. It may explain how it is possible to *feel certain* about intuitions (as opposed to be convinced of reasoned argument). It also prevents the Global Workspace from being flooded out with predicted information that is not strongly supported, allowing the important material to shine through. A particularly interesting point is this: should a generator make an unlikely prediction, that has sufficient prediction-h to be "audible", in the absence of other explanations, that prediction will have low prediction-H, and so will not be suppressed by this final mechanism. Increasing the range of possibilities over which the distribution holds, even if they are unlikely, increases prediction-H and thus decreases certainty. Under this régime an organism that has less experience is more likely to admit unlikely predictions to consciousness; this might be taken to account for the tendency, for example, of children to be more affected by imagined fears than adults.

No straightforward diagram can be drawn of the effect of prediction-H on the overall likelihood of a generator taking over the Global Workspace, because the numbers depend heavily on the multidimensional distributions from which the various Hs are calculated.

This leaves us with a "volume" value for each generator, T, which is estimated by the following, for either kind of h, above:

$$T = \frac{p \times h}{H}.$$

I propose that, at any given moment, this "volume" value is used in deciding which of the range of possible inputs, derived from matching sensory input to statistical models in memory, enters the Global Workspace. This is illustrated in Figure 3.

Generation, Creativity and Intuition

In the previous section, I outlined a simple, comparative mechanism by which statistically likely and information-theoretically rich structures can emerge from a multi-agent system furnished with high-quality models of a domain of knowledge. With such a mechanism, the Threshold Paradox disappears. I should also note that it is possible that such a mechanism is one of Baars' own proposals; however, if so, it is not clearly specified as such. The remaining question is then: how does this mechanism for choosing access to consciousness help to simulate creativity?

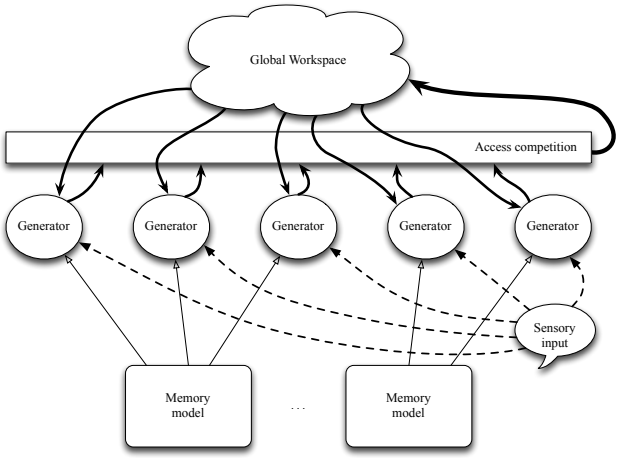

Figure 3: Schematic diagram of my proposal for the Global Workspace. In this version, there is no need for a threshold of access. Instead, the generators compete, one against another, and probability and information content determine the winner.

Perhaps surprisingly, an answer may be found in the writing of Wolfgang Amadeus Mozart (quoted by Holmes, 2009, pp. 317–8):

> When I am, as it were, completely myself, entirely alone, and of good cheer – say traveling in a carriage, or walking after a good meal, or during the night when I cannot sleep; it is on such occasions that my ideas flow best and most abundantly. Whence and how they come, I know not; nor can I force them. Those ideas that please me I retain in memory, and am accustomed, as I have been told, to hum them to myself.

One might paraphrase the opening sentence here as "when I am not being bothered, and when I have no worries and no particular goals", which in turn means "when I have no distractions" or "when I have no information-rich input to consciousness from outside or within". This accords with a situation when the Global Workspace is occupied only by weakly-informative ephemera, and when generators are receiving little or no external stimulus.

Recall now my earlier proposal that effective animals will base their actions not only on received stimuli, but on the results of the comparison of received stimuli with predictions about the current state of the world made from the previous state(s). Suppose that those generators continue to generate, even when there is very little informative input. Given the appropriate knowledge and tendencies in a particular individual (music—and, by all accounts, scatology—in Mozart), generators will begin to freewheel, within the same statistical framework as above, but lacking the statistical prior of a particular stimulus. The outputs, one might expect, will be rather more diffuse and perhaps less highly rated than when directly stimulated, but this is not a brake on their progress towards the Global Workspace, because *there is little or no competition*. At this point, the diagram in Figure 2 becomes recognisable as the Wundt curve, as it defines a

sweet spot of balance between dullness and over-complexity in information-theoretic terms.

Mozart, above, describes a particular kind of musical approach, where one essentially enters a quiescent state, in order deliberately to allow Baars' generators to freewheel; I find that the same method works for writing. But this is only one case of many. For example, the mechanism above also accounts for why hearing a musical phrase, or even a non-musical pitch sequence, may give rise to new musical phrase: the percept conditions the generators in a particular way, and so affects the likely outcomes, which are generated all the time. The ones with the right statistical properties make it into the Global Workspace, and so can be further elaborated.

Note that this mechanism can apply to any statistical model available to the generators, so it need not be restricted to music (as it is in the system components summarised in the next section). In principle, the same idea can work with any model from which statistical likelihoods can be computed. This means, for example, that it can account for the generation of sentences, and therefore possibly internal speech. If internal speech is equated with essential thought, as commonly, then the current approach can account for general creative thought and for the emergence of particular thoughts into consciousness as intuition. It can also, via prediction-H, account for the (sometimes inappropriate) feeling of certainty associated with thoughts and intuitions.

Thus, I suggest that "Threshold Paradox" as a name for this issue needs to be reinterpreted. The paradox is not in the nature of the threshold, but in the formulation of the Global Workspace as requiring one. The current theory reformulates entry to the Workspace as purely competitive, without a particular boundary, so, one might say, the paradox arose from the assumption that the Threshold exists.

What is more, in the present theory, there is no longer any need to search for an explanation of creativity as a distinct phenomenon. In my approach, non-conscious creativity is happening all the time as a result of on-going anticipation in all sensory (and other) modalities. When conditions are right, this essential survival mechanism is not so much *exapted* for creativity, but gives rise to creativity as a side effect.

Towards a Creative System

To ground this theory in a technical base, I now summarise research that has already been conducted towards building a system of the kind proposed here, in the domain of musical creativity. Pearce and Wiggins (2006)[4] describe a statistical model of musical learning, based on, but extending, statistical language learning methods. Wiggins (2011) has shown that the extensions to the musical model can also benefit language models. Pearce and Wiggins (2007) showed how the model could generate entire musical melodies, though the requirements of the current proposal are less stringent, as fragmentary musical ideas are all that is required: in this case short sequences of notes that might be consciously assembled into melodies. Most importantly, the model has been used to demonstrate that high information content corresponds with increased beta-band synchrony in human listeners (Pearce et al., 2010), providing at least circumstantial evidence that cognitive resource (i.e., attention) does indeed follow information content, which would accord with that information's entry into the Global Workspace when the circumstances, as described above, are right.

Ponsford, Wiggins, and Mellish (1999), Whorley, Pearce, and Wiggins (2008), Whorley, Wiggins, and Pearce (2007) and Whorley et al. (2010) have presented more complex models for dealing with deeper aspects of music than melody.

A crucial piece of evidence for the model of creativity proposed above is embedded in the workings of Pearce's perceptual model—recalling that perception and prediction are closely linked in the view of the world presented here. There are two sub-models, both of which contain multiple predictors. The distributions output by the two sub-models are combined multiplicatively, with weightings derived their relative information entropy. The distributions output by the multiple predictors *within* each of the two sub-models are combined in the same way. Other configurations (for example, a one-stage combination of all of the distributions, instead of this two-stage combination) produce a less successful model of human behaviour. This system matches exactly against the multi-stage version of Baars' Global Work Space, described above, coupled with my proposal for a competition mechanism based on information content and entropy.

There is still substantial work to be done on this model before the simulation of creativity can be claimed. The next threshold to cross is not a paradox, but the engineering task of implementing the integrated multiple generators in the model described above, to test out the this particular approach to competitive generation in the Global Workspace.

[4] A fuller presentation of the modelling work published up to 2007 is given in Pearce's (2005) PhD thesis.

Acknowledgments

I gratefully acknowledge the contribution of the ISMS group, most particularly Roger Dean, Ollie Bown, Jamie Forth and Marcus Pearce, of Joydeep Bhattacharya, and of three anonymous referees, to the thinking presented here. Funding was provided by EPSRC Research Grant EP/H01294X/2, "Information and neural dynamics in the perception of musical structure".

References

Baars, B. J. 1988. *A cognitive theory of consciousness*. Cambridge University Press.

Chalmers, D. J. 1996. *The Conscious Mind: in search of a fundamental theory*. OUP.

Colton, S. 2009. Seven catchy phrases for computational creativity research: A position paper. In *Proceedings of the Dagstuhl Workshop on Computational Creativity*. Germany: Schloss Dagshtul.

Corkill, D. D. 1991. Blackboard systems. *AI Expert* 6(9):40–47.

Edelman, G. M.; Gally, J. A.; and Baars, B. J. 2011. Biology of consciousness. *Frontiers in Psychology* 2.

Gärdenfors, P. 2000. *Conceptual Spaces: the geometry of thought*. Cambridge, MA: MIT Press.

Holmes, E. 2009. *The Life of Mozart: Including his Correspondence*. Cambridge Library Collection. Cambridge University Press.

Huron, D. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Bradford Books. Cambridge, MA: MIT Press.

Lakatos, I. 1970. Falsification and the methodology of scientific research programmes. In Lakatos, I., and Musgrave, A., eds., *Criticism and the Growth of Knowledge*. Cambridge, UK: Cambridge University Press. 91–196.

MacKay, D. J. C. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge University Press.

Minsky, M. 1985. *The Society of Mind*. New York, NY: Simon and Schuster Inc.

Moffat, D., and Kelly, M. 2006. An investigation into people's bias against computational creativity in music compositio. In *Proceeings of the International Joint Workshop on Computational Creativity*.

Pearce, M. T., and Wiggins, G. A. 2006. Expectation in melody: The influence of context and learning. *Music Perception* 23(5):377–405.

Pearce, M. T., and Wiggins, G. A. 2007. Evaluating cognitive models of musical composition. In Cardoso, A., and Wiggins, G. A., eds., *Proceedings of the 4th International Joint Workshop on Computational Creativity*, 73–80. London: Goldsmiths, University of London.

Pearce, M. T.; Herrojo Ruiz, M.; Kapasi, S.; Wiggins, G. A.; and Bhattacharya, J. 2010. Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. *NeuroImage* 50(1):303–314.

Pearce, M. T. 2005. *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. Ph.D. Dissertation, Department of Computing, City University, London, UK.

Ponsford, D.; Wiggins, G. A.; and Mellish, C. 1999. Statistical learning of harmonic movement. *Journal of New Music Research* 28(2):150–177.

Ritchie, G. 2001. Assessing creativity. In *Proceedings of the AISB'01 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, 3–11. Brighton, UK: SSAISB.

Russell, S., and Norvig, P. 1995. *Artificial Intelligence – a modern approach*. New Jersey: Prentice Hall.

Shannon, C. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423, 623–56.

Tononi, G., and Edelman, G. M. 1998. Consciousness and complexity. *Science* 282(5395):1846–1851.

Wallas, G. 1926. *The Art of Thought*. New York: Harcourt Brace.

Whorley, R.; Wiggins, G. A.; Rhodes, C.; and Pearce, M. 2010. Development of techniques for the computational modelling of harmony. In Ventura, et al., eds., *Proceedings of the First International Conference on Computational Creativity*.

Whorley, R. P.; Pearce, M. T.; and Wiggins, G. A. 2008. Computational modelling of the cognition of harmonic movement. In *Proceedings of the 10th International Conference on Music Perception and Cognition*.

Whorley, R. P.; Wiggins, G. A.; and Pearce, M. T. 2007. Systematic evaluation and improvement of statistical models of harmony. In A. Cardoso, and G. A. Wiggins., eds., *Proceedings of the 4th International Joint Workshop on Computational Creativity*, 81–88.

Wiggins, G. A. 2011. "I let the music speak": cross-domain application of a cognitive model of musical learning. In Rebuschat, P., and Williams, J., eds., *Statistical Learning and Language Acquisition*. Amsterdam, NL: Mouton De Gruyter.

Wiggins, G. A. 2012. Defining inspiration? Modelling non-conscious creative process. In Collins, D., ed., *The Act of Musical Composition – Studies in the Creative Process*. Aldershot, UK: Ashgate.